AN ABSTRACT OF THE THESIS OF

Adel M. Halawa for the degree of Doctor of Philosophy in Statistics

presented on February 9, 1989.

Title: Testing for Location After Transformation to Normality

Abstract approved: Redacted for Privacy

David S. Birkes

In the problem of testing the median using a random sample from a certain distribution, and if no other parametric family is suggested, the t-test is known to be the optimal procedure when this distribution is normal. If the sample appears to be non-normal, one has the choice either to consider a non-parametric approach or to try to correct for non-normality before applying the t-test.

In this thesis we investigate the effect of applying certain power transformations as an action to correct for non-normality before applying the t-test. Also we investigate the effect of applying a power transformation then trimming a certain proportion from the data on each tail as a double action to correct for non-normality. This problem is first considered by Doksum and Wong (1983), who apply the Box-Cox power transformations to positive, right-skewed data when testing for the equality of distributions of two independent samples.

In the present work we provide results for the one-sample case using two alternatives to the Box-Cox power family which are applicable to all data sets. Whenever it can be assumed that the data is a random

sample from a symmetric distribution with heavy tails, it is shown that the John-Draper family of modulus power transformations, with the transformation parameter being positive and smaller than 1 , is appropriate to correct for non-normality and the t-test based on the transformed data is asymptotically more efficient and has better power properties than the t-test based on the data in its original scale. When the data is thought to have a skewed distribution and can assume negative as well as positive values, a new family of transformations, referred to as the two-domain family, is introduced. It is shown that the t-test based on the data after applying this new transformation is also asymptotically more efficient and has better power properties than the t-test in the original scale. A simulation study shows that trimming a certain proportion on each tail of the data transformed by one of the above two transformations then applying the t-test to the trimmed samples yields a considerable gain in power compared to the t-test in the original scale.

Testing for Location After Transformation to Normality

by

Adel Mahmoud Halawa

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of
Doctor of Philosophy

Completed February, 9 1989

Commencement June, 1989

APPROVED

Redacted for Privacy

Professor of Statistics in Charge of Major

Redacted for Privacy

Chairman of Department of Statistics

Redacted for Privacy

Dean of Graduate School

Date Thesis is Presented <u>February 9, 1989</u>

Typed by the Author for <u>Adel Mahmoud Halawa</u>

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my major professor, Dr. David Birkes for his inspiration and guidance throughout this work. It was a great lesson I learned from him that, no research without problems and no solution for the problems without showing a great deal of patience. In fact I tried to quit this topic three times during different stages of the research and without the support of Dr. Birkes I would have never done it.

My deepest appreciation is to Dr. Justus Seely for his continuous support and encouragement since the first day I joined the department

Also, I would like to thank Dr. David Thomas and Dr. Kenneth Row with whom I took my first two courses at the department and whose encouragement while taking these two courses have certainly affected my performance throughout the whole period of my stay at Oregon State University.

Finally, my deepest thanks are for my mother for her great and uncountable sacrifices, for my wife Hanan, my son Sherief and my daughter Maha.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

TESTING FOR LOCATION AFTER TRANSFORMATION TO NORMALITY

1. INTRODUCTION

Let $Y_1, Y_2, \ldots, Y_n$ be a sample of independently and identically distributed random variables with distribution function F. The problem we are concerned about in this thesis is the one of testing a statistical hypothesis concerning the median of F. If F is assumed to be normal, then the t-test is the optimal testing procedure to consider, since it is the uniformly most powerful unbiased test. If F is not normal but has a symmetric distribution with a finite variance, the Central Limit Theorem will then insure that the t-statistic for testing the median (= mean) is asymptotically valid in the sense that it asymptotically gives the correct level and power. However, in this situation many researchers have shown that if F has heavier tails than a normal distribution, the t-test tends to be conservative for small samples and hence has poor power properties compared to the normal situation. Also, under the same situation it is known from the literature that, even asymptotically, the t-test is not as efficient as other robust or nonparametric procedures.

From the discussion above we conclude that the optimality of the t-test procedure is restricted to the situation where the data are approximately normally distributed. Therefore if one suspects that a set of data departs from normality, one might consider some procedure that corrects for such departure and brings the data into approximate normality. Two common procedures that can be considered in such cases are transforming the data and trimming a certain proportion from the extreme values of the data on each side. These two procedures can be combined by first transforming the data then trimming the extreme observations.

Doksum and Wong (1983) consider the problem of testing equality of the means of two independent samples from distributions that are skewed to the right. They only consider positive data sets. To each data set they apply a Box-Cox power transformation and then trim the extreme observations on each side. They investigate the efficiency and power of the t-test statistic based on such corrected data relative to the t-test in the original scale. In a technical report, Doksum and Wong (1980) report that their results are valid for testing any statement about the means of more than two populations as long as this statement is in the form of a contrast. On the other hand they state that their asymptotic results are not valid for the single-sample case.

In the present work, we go along the lines of Doksum and Wong (1983) with two alternative families of power transformations that may be used instead of the Box-Cox transformation in cases where the latter does not work well. These two families are the John-Draper family of modulus transformations, which is suitable for dealing with heavy tailed symmetric distributions, and a new family of transformations which we refer to as the two-domain family and which is suitable for dealing with nonsymmetric data by dealing differently with each tail of the distribution of the data. The main advantage of these two alternative families is that they are applicable to all data sets which may assume negative as well as positive values. Also these two families provide asymptotic results for the one-sample problem.

This thesis contains both asymptotic and simulation results. For the asymptotic results, expressions of Pitman's asymptotic relative efficiency of the transformed t-test relative to the t-test without transformation are derived under the John-Draper family for symmetric models and under the two-domain family for skewed models. These expressions include expectations that cannot be evaluated analytically, and so numerical integration techniques are used to evaluate them. It is shown that the transformed t-test using consistent estimators of the transformation parameters is asymptotically much more efficient than the t-test in the original scale. The asymptotic results concerning the two-domain family are derived using the normal maximum likelihood estimators of the transformation parameters. Also throughout the simulation study the normal maximum likelihood procedure is used to estimate the transformation parameters. We include in the thesis a proof of the consistency and asymptotic normality of these estimators under the two-domain family; similar results for the John-Draper family follow as special cases. For the simulation results under symmetric models we simulated the level and power for six test statistics. These are: the t-test in the original scale of the data, the t-test after transformation, the t-tests from data trimmed by .10 and .20 on each tail and the t-tests from data that is first transformed then trimmed by .10 and .20 on each tail. These percentages of trimming are the same as those considered by Doksum and Wong (1983). For skewed models we simulated the level and power for the first four test statistics mentioned above. The simulation results show that, under the normal model, the transformed t-test has almost the same level and power as the t-test in the original scale (the uniformly most powerful unbiased test). Under this model the trimmed tests, whether transformed or not, cannot compete with the two untrimmed tests. Also, the simulation results show that under some heavy-tailed distributions the transformed t-test is more powerful than the t-test in the original scale, which supports the asymptotic efficiency results. Under such non-normal distributions it is always the case in our simulations that the trimmed t-tests are more powerful than the two untrimmed tests. In some situations the difference in simulated powers between the trimmed tests and the transformed then trimmed tests is slight and in some other situations there is a considerable difference in favor of the transformed then trimmed tests.

The thesis, besides this introduction, contains four other chapters and six appendices.

In Chapter 2 we discuss some of the testing procedures appropriate for testing a location parameter. We also discuss four families of transformations: the Box-Cox family of power transformations and its effect in removing skewness from right-skewed positive data; the shifted Box-Cox power family and the problem of estimating the shift parameter; the John-Draper family of modulus transformations and its effect on symmetric heavy-tailed data; and a new family of transformations called the two-domain family. The chapter also includes a discussion about the effect of using a data-based estimator of the transformation parameter on subsequent analysis. The chapter is concluded with a brief review of the work of Doksum and Wong (1983).

In Chapter 3 we derive the asymptotic relative efficiency under symmetric models for the t-test in the original scale relative to the transformed t-test, using the John-Draper transformation and a consistent estimator of the transformation parameter. This includes proving that both of these two tests are Pitman regular. The definition of a Pitman regular test statistic and the derivation of Pitman's asymptotic relative efficiency under a general setting constitute the material of Appendix B. In this chapter and the next one, the transformed t-test assuming the transformation is known, is considered to be the basic test to which we relate the asymptotic results. This means that we first derive the asymptotic relative efficiency of the t-test in the original scale relative to the basic test and then we show that the asymptotic relative efficiency of the transformed t-test using a consistent estimator of the transformation parameter relative to the basic test is one. We give proofs of the main required results in the body of the chapter and proofs of intermediate results are given in Appendix C . Also the chapter includes some tables of asymptotic relative efficiency under some transformed models.

In Chapter 4 we go through the main lines of Chapter 3 with skewed instead of symmetric models and the two-domain transformation instead of the John-Draper transformation. The results of Chapter 4 are not as general as those of Chapter 3. The results of Chapter 3 are valid under the situation where there exists a transformation that can transform the data to a symmetric model. The results of Chapter 4 are restricted to the situation where there exists a transformation that can transform the data into a normal model. Arguments similar to those given in this chapter would be needed for each alternative to the normal model. On the other hand it is the transformation to normality that is of most interest when the problem is to test the mean or the median of a certain distribution. As in Chapter 3, some of the proofs are given in the body of the chapter and the rest are given in Appendix D. Also in this chapter we rely on the material of Appendix E which covers the consistency and asymptotic normality of the maximum likelihood estimators of the transformation parameters of the two-domain family.

In Chapter 5 we discuss the simulation study. The chapter includes tables of the simulated power and level under different symmetric and skewed models. A list of the programs used in simulating the power under different models is given in Appendix F, which also contains the programs that calculate the asymptotic relative efficiencies of both Chapters 3 and 4.

Appendix A contains some facts and theorems about convergence in probability and convergence in law which are frequently used in Chapters 3 and 4 with or without reference to their places in the appendix.

2. HYPOTHESIS TESTING OF LOCATION AFTER TRANSFORMATION

Let $Y_1, Y_2, \ldots, Y_n$ be a random sample with a distribution function $F_{\theta,\sigma}$ where θ denotes the median of Y and σ is a scale parameter. In the problem of testing the null hypothesis H_0: $\theta = \theta_0$ against the alternative H_1: $\theta > \theta_0$ one can either consider a nonparametric approach where no distributional assumptions are required or a parametric approach where some underlying model should be assumed. Under the parametric approach if the normality assumption can be made, then the t-test is the uniformly most powerful unbiased test for testing the above hypothesis. On the other hand, if the data suggest some sort of departure from normality, and if no other parametric model is suggested, it may be recommended to consider some procedure to correct for the non-normality of the data before calculating the t-statistic. To determine what procedure should be considered to correct for the non-normality we distinguish between two types of departures from normality. In the first type, while most of the data seem to be from a normal model, a few outlying observations exist on one or both tails. In this case trimming a certain proportion from the smallest and/or the largest observations will be the appropriate action to correct for the non-normality. In the second type, the whole data set may indicate departure from normality, such as when the distribution has heavier tails than a normal or the distribution is skewed. A power transformation may be recommended to correct for this type of departure. Also, there may be cases in which both trimming and transforming should be considered together.

In Section 2.1 we discuss the properties of the t-test and the consequences of applying the t-test to non-normal data. Also we introduce the trimmed t-test. In Section 2.2 we discuss families of power transformations, their properties and their effects on the distribution of data. The families we consider are the Box-Cox and the shifted Box-Cox families, the family of modulus power transformations (John-Draper family) and a new family of transformations which we call the two-domain family. In Section 2.3 we discuss the problem of using data-based estimators of the transformation parameters on subsequent analysis. In Section 2.4 we review the work of Doksum and Wong (1983) who were the first to consider the t-test after applying the Box-Cox power family.

2.1 Procedures for Testing a Location Parameter and their Properties

In this section we discuss the t-test and the trimmed t-test procedures for testing

$$H_0\colon \theta = \theta_0 \quad \text{against} \quad H_1\colon \theta > \theta_0 , \tag{2.1.1}$$

with emphasis on the consequences of using the t-test when the data are actually not normally distributed.

2.1.1 The t-test

The t-test statistic for testing the null hypothesis in (2.1.1) is defined as

$$t = \sqrt{n}\,(\bar{Y} - \theta_0)\,/\,\hat{\sigma}_Y \tag{2.1.2}$$

where $\bar{Y} = (1/n)\sum_{i=1}^{n} Y_i$ and $\hat{\sigma}_Y^2 = \sum_{i=1}^{n} (Y_i - \bar{Y})^2/(n-1)$. If Y is distributed as $N(\theta, \sigma_Y^2)$, then $\bar{Y}$ is distributed as $N(\theta, \sigma_Y^2/n)$, $(n-1)\hat{\sigma}_Y^2/\sigma_Y^2$ is distributed as $\chi^2(n-1)$, and $\bar{Y}$ and $\hat{\sigma}_Y^2$ are independent. It follows that t has a noncentral t-distribution with $n-1$ degrees of freedom and noncentrality parameter $\sqrt{n}\,(\theta - \theta_0)/\,\sigma_Y$ which is zero if and only if H_0 is true. Under the above setting the test procedure that rejects H_0 for values of t larger than the $(1-\alpha)100\%$ percentile of the t-distribution with $n-1$ degrees of freedom is the uniformly most powerful unbiased level-α test.

Although normality is an appealing assumption in many statistical inference problems, it is generally believed that one will never observe a random sample that is exactly normally distributed. It is more practical to assume that the data have some sort of departure from normality which may be due to the existence of some outliers in one or both tails, or due to heavy tails or skewness of the distribution of Y.

Now we consider the problem of assuming that a set of data is normally distributed when in fact it deviates from normality due to one of the above reasons. We address this problem in terms of four questions:

1. How are the level and the power of the t-test affected by non-normality of the data in large samples?

2. How does the power of the t-test compare with other tests for non-normal data in large samples?

3. How are the level and power of the t-test affected by non-normality in small samples?

4. How does the power of the t-test compare with other tests for non-normal data in small samples?

For the answer of the first of these questions we refer to Lehmann (1986, Section 5.4) and Tiku et al. (1986, p.2-3 and Chapter 4) who use the Central Limit Theorem and consistency of $\hat{\sigma}_Y^2$ to show that the t-test is asymptotically robust-valid in the sense that in the limit it gives the correct level and power no matter what the distribution of Y is, as long as the variance of Y is finite. With respect to the second question, they state that the t-test is not robust-efficient in the sense that some other test procedures like the trimmed t-test or the Wilcoxon signed-rank test are more powerful than the t-test under some non-normal models.

There are many papers in the literature that deal with the third question. For example, Tan (1982) gives a list of 55 references that deal with the distribution of the t-statistic when Y is not normal. Geary (1947) gives approximate formulas for the first four moments of the distribution of the t-statistic listed as F1 through F4 below. Let $\mu_k(Y)$ denote the k^{th} central moment of Y and let $\beta_1(Y)$ and $\beta_2(Y)$ denote the coefficient of skewness and the coefficient of kurtosis respectively. That is,

$$\beta_1(Y) = \mu_3(Y)/(\mu_2(Y))^{3/2}, \quad \beta_2(Y) = \mu_4(Y)/(\mu_2(Y))^2 .$$

Under the assumption $\theta = \theta_0$,

F1. $E(t) = -1/2\ \beta_1(Y)/n - O(n^{-1.5})$

F2. $var(t) = 1 + 1/4\ (8 + 7\beta_1(Y))/n + O(n^{-2})$

F3. $\beta_1(t) = -2\ \beta_1(Y)/\sqrt{n} - O(n^{-.5})$

F4. $\beta_2(t) = 3 + 2(3 - \beta_2(Y) + 6\beta_1(Y))/n + O(n^{-2})$.

Using these approximations, it is easy to see the following relations between the shape of the distribution of Y and the shape of the distribution of the t-statistic.

1. From F4, $\beta_2(t)$ is inversely related to $\beta_2(Y)$ and hence the heavier the tails of the distribution of Y are the lighter the tails of the distribution of t will be and vice versa. This led Yuen and Murthy (1974) to state that "It is also well known that the usual Student's t-test is conservative and hence less powerful when the underlying distribution is long tailed."

2. From F1 and F3, if $\beta_1(Y) > 0$, then $E(t)$ and $\beta_1(t)$ are both negative which means that the distribution of t will be skewed to the left if the distribution of Y is skewed to the right, and vice versa. Hence, with a rejection region of the form $\{t: t > c\}$, if the distribution of Y is skewed to the right, one would expect the t-test to be conservative, and if the distribution of Y is skewed to the left, it will be expected that the actual level of the test will always be greater than the nominal level.

3. From the order of convergence of the above approximations to the exact moments of the t-statistic under normality note that both the variance and kurtosis coefficient are $O(n^{-2})$ while $\beta_1(t)$ is $O(n^{-.5})$ and $E(t)$ is $O(n^{-1.5})$, which means that the variance and kurtosis coefficient are less affected by departure from normality.

The fourth question, about how the power of the t-test compares with other tests in small samples, can be investigated by simulation. Among others, Tiku (1980) and Doksum and Wong (1983) give some simulation results about the power of the t-test and the power of other nonparametric and robust test procedures for testing statistical hypotheses concerning some location parameter of a variety of nonnormal distributions.

2.1.2 The trimmed t-test

This test is based on a studentized version of the trimmed mean. To understand the rationale behind this test we give the following premise from Tiku (1980): "Non-normality essentially comes from the tails and once the extreme observations (representing these tails) are censored, there is hardly any difference between a normal sample and a non-normal sample." Our point of view about this premise is that it is most meaningful under the case where the majority of the data appear to be normally distributed except for some outliers in one or both tails. Trimming a few observations from each tail may correct for such type of departure from normality. However, when the majority of the data do not appear to be normally distributed, then changing the scale of Y by applying some power transformation, or by both transforming and trimming, may be more meaningful.

Let $Y_{(1)}, Y_{(2)}, \ldots, Y_{(n)}$ denote the order statistics corresponding to $Y_1, Y_2, \ldots, Y_n$ and let δ be any positive number smaller than $.5$ such that $r = n\delta$ is an integer. The δ-trimmed mean is defined as

$$\bar{Y}_{tr} = \sum_{i=r+1}^{n-r} Y_{(i)} \;/\; (n-2r) \;. \tag{2.1.3}$$

If the distribution of Y is symmetric about θ, Lehmann (1983, p.361) gives the asymptotic distribution of $\sqrt{n}\,(\bar{Y}_{tr} - \theta)$ as $N(0, \sigma_\delta^2)$ where

$$\sigma_\delta^2 = \frac{2}{(1-2\delta)^2}\left[\int_0^{\xi(1-\delta)} y^2 f(y)\,dy + \delta\xi^2(1-\delta)\right]$$

where $\xi(\delta)$ is the unique value for which $F[\xi(\delta)] = \delta$. Stigler (1973) in proving the asymptotic normality of the trimmed mean shows that uniqueness of $\xi(\delta)$ is both a necessary and sufficient condition for the distribution of $\sqrt{n}\,(\bar{Y}_{tr} - \theta)$ to be asymptotically normal. Let

$$\hat{\sigma}_\delta^2 = \frac{r(Y_{(r+1)} - \bar{Y}_{tr})^2 + \sum_{i=r+1}^{n-r}(Y_{(i)} - \bar{Y}_{tr})^2 + r(Y_{(n-r)} - \bar{Y}_{tr})^2}{(n - 2r - 1)} \tag{2.1.4}$$

Huber (1970, p.453-463) shows that under certain regularity conditions

$$\sigma_\delta / \hat{\sigma}_\delta \xrightarrow{p} \sqrt{n/(n-2r)} .$$

Hence,

$$t_{tr} = \sqrt{n - 2r}\left(\frac{\bar{Y}_{tr} - \theta_0}{\hat{\sigma}_\delta}\right) \tag{2.1.5}$$

is an asymptotically valid test statistic for testing the hypotheses given by (2.1.1) in the sense that its limiting distribution is N(0,1) when $\theta = \theta_0$.

2.2 Families of Power Transformations

There has been considerable literature on the subject of power transformations since they were introduced by Box and Cox in (1964). Power transformations are considered when there is evidence that some of the model assumptions associated with a certain data analysis procedure are violated and such violation can be removed if the random variable is expressed in a different scale. Most of the literature considers power transformations in a regression model setting. The main goal in the regression setting is to achieve linearity, constancy of variance, or some distributional assumption. However, since we are considering a single sample model, our main interest is to achieve normality. Investigating transformations to normality under regression or analysis-of-variance models may not be effective due to what Weisberg (1985, p.157) and Quesenberry and Quesenberry (1982) call the super-normality of residuals where the residuals show a normal trend even if the actual distribution of the errors is not normal. In the following we consider a definition of a power transformation and the properties it should satisfy so that the normality assumption is valid. Then, we discuss four families of power transformations each of which is valid to deal with a certain type of data.

We may define a transformation to normality of a random variable Y as a function

$$h_\lambda : Y \longrightarrow Y(\lambda) = h_\lambda(Y)$$

such that $Y(\lambda)$ is (approximately) normally distributed. The domain of h_λ is the sample space of the original variable Y and its range is the space over which we assume that the transformed variable will be normally distributed. It would be desirable to have the domain of the transformation be the whole real line in order to be applicable to all data sets and to have its range be the whole real line in order for the normality assumption to be completely valid. The parameter λ could be a vector as Box and Cox (1964), Andrews (1971) and Carroll and Ruppert (1984) mention.

2.2.1 The Box-Cox family of power transformations

Box and Cox (1964) introduce this family for positive random variables Y as

$$Y(\lambda) = \begin{cases} (Y^\lambda - 1)/\lambda & \text{if } \lambda \neq 0 \\ \ln Y & \text{if } \lambda = 0 \end{cases} . \tag{2.2.1}$$

Note that $Y(\lambda)$ is monotone increasing in Y, continuous in λ, bounded below by $-1/\lambda$ if λ is positive and bounded above by $-1/\lambda$ if λ is negative. A weakness of the Box-Cox family is that it cannot handle negative data. This drawback is eliminated by introducing a location parameter yielding the family of shifted power transformations.

2.2.2 The family of shifted power transformations

Box and Cox (1964) extend the above family when the random variable Y can assume negative values by introducing a shift parameter γ greater than the negation of the smallest value of Y, so that $Y+\gamma$ is positive for all Y and then apply the family in (2.2.1) to the shifted data.

$$Y(\gamma,\lambda) = \begin{cases} ((Y+\gamma)^{\lambda} - 1)/\lambda & \text{if } \lambda \neq 0 \\ \ln\,(Y+\gamma) & \text{if } \lambda = 0 \end{cases} . \tag{2.2.2}$$

If a set of data includes both negative and positive numbers and some transformation needs to be considered, it is always recommended in the literature to apply the shifted power transformation. This recommendation is appropriate if there is a natural lower bound for Y which can then be used as a shift parameter. If there is no such lower bound, like for example the case with differences in paired samples, and γ has to be estimated from the data, Atkinson (1983, 1985) showed that the family of shifted power transformations will not work well in general because the likelihood used in estimating the parameters is unbounded as γ approaches the minimum value of Y. This leaves the problem of how to deal with negative data unsolved.

As mentioned earlier, from a theoretical point of view it would be desirable for the range of the transformation to be the whole real line in order for the normality assumption to be valid. For the Box-Cox or the shifted Box-Cox transformations this will be true only if λ is zero; otherwise the range of the transformation is bounded as shown above and the normality assumption cannot hold. However, from the practical point of view it may be the case that over a bounded region the distribution of the transformed variable may appear to be more close to a normal distribution than the distribution of Y. The following theorem shows the effect of applying the Box-Cox transformation to a set of data that is skewed to the right.

Theorem 2.2.1 Suppose Y is a positive random variable and suppose that $0 < \lambda < 1$. Let $Z = h_\lambda(Y)$ where h_λ denotes the Box-Cox transformation. Then $\beta_1(Z) \le \beta_1(Y)$, where β_1 denotes the coefficient of skewness.

Proof Note that the inverse transformation of $z = h_\lambda(y)$ is

$$y = h_\lambda^{-1}(z) = (1 + \lambda z)^{1/\lambda}$$

$$\frac{\partial y}{\partial z} = (1 + \lambda z)^{(1/\lambda)-1}$$

$$\frac{\partial^2 y}{\partial z^2} = (1-\lambda)\ (1 + \lambda z)^{(1/\lambda)-2}.$$

Hence for $0 < \lambda < 1$, we have $\frac{\partial^2 y}{\partial z^2} > 0$. Therefore, the inverse transformation $y = h_\lambda^{-1}(z)$ is a convex function. The result follows from Theorem 2.2.1 of Van Zwet (1964). []

Hence, when the Box-Cox transformation with $0 < \lambda < 1$ is applied to a set of positive data that is skewed to the right, the transformed variable will have a distribution that is more symmetric than the distribution of the original variable. A similar description about the effect of the Box-Cox family in removing skewness from right-skewed data is given by Hoaglin, Mosteller and Tukey (1983, p.100) and by Carroll and Ruppert (1988, Section 4.2).

2.2.3 The family of modulus power (or John-Draper) transformations

This family has the form

$$Y(\lambda) = \begin{cases} \text{sign}(Y)\ ((|Y|+1)^{\lambda} - 1)/\lambda & \text{if } \lambda \neq 0 \\ \text{sign}(Y)\ \ln(|Y|+1) & \text{if } \lambda = 0 \end{cases} . \qquad (2.2.3)$$

John and Draper (1980) introduced this family for a set of difference data (in an analysis-of-variance context) when they recognized that the shifted Box-Cox transformation failed to improve the residual plots. These plots showed a symmetric distribution of residuals with longer tails than a normal pattern should show. Although in their comments John and Draper state that "The modulus transformation is clearly an alternative which may work well in circumstances in which the power transformation would be inappropriate ...", this family has never been in use since the time it was first introduced. We propose to use this family with $0 < \lambda < 1$ for data that is assumed to be symmetrically distributed with longer tails than that of a normal distribution and to use it with $\lambda > 1$ for data that is assumed to be symmetrically distributed with shorter tails than that of a normal distribution. We will not study the case $\lambda < 0$, because then the transformed variable is bounded.

Adding a centrality parameter to allow the center of symmetry to be equal to some number θ, the John-Draper family can be written as

$$Y(\theta,\lambda) = \begin{cases} \text{sign}(Y-\theta)\ ((|Y-\theta|+1)^{\lambda} - 1)/\lambda & \text{if } \lambda \neq 0 \\ \text{sign}(Y-\theta)\ \ln(|Y-\theta|+1) & \text{if } \lambda = 0 \end{cases} . \qquad (2.2.4)$$

In addition to being applicable to all data sets, we next show that this family can deal successfully with problems in the tails of symmetric distributions.

Definition A real-valued function g on an interval I is said to be antisymmetrical on I if $g(y_0+y) + g(y_0-y) = 2\, g(y_0)$ for some $y_0 \in I$ and all y_0+y and y_0-y in I .

Definition An antisymmetrical function g on I is said to be concave-convex on I if g is concave for all $y \leq y_0$ and convex for all $y \geq y_0$ and $y \in I$. the point y_0 is called the central point of g .

Theorem 2.2.2 Let h_λ denote the John-Draper transformation and suppose that $0 < \lambda < 1$. Let $Z = h_\lambda(Y)$. Then $\beta_2(Z) \leq \beta_2(Y)$ where β_2 denotes the coefficient of kurtosis.

Proof Using the inverse transformation and Theorem 2.3.2 of Van Zwet (1964), it suffices to show that Y is an antisymmetrical concave-convex function of Z with central point $Z_0 = 0$. The inverse transformation of $z = h_\lambda(y)$ is

$$y = h_\lambda^{-1}(z) = \operatorname{sign}(z)\ ((1 + \lambda|z|)^{1/\lambda} - 1) .$$

For $z > 0$,

$$h_\lambda^{-1}(0+z) = (1+\lambda z)^{1/\lambda} - 1 \quad \text{and} \quad h_\lambda^{-1}(0-z) = 1 - (1+\lambda z)^{1/\lambda}$$

Since $h_\lambda^{-1}(0) = 0$, then $h_\lambda^{-1}(0+z) + h_\lambda^{-1}(0-z) = 2\, h_\lambda^{-1}(0) = 0$. A similar argument holds for $z < 0$. Therefore Y is an antisymmetrical function of Z with central point 0 .

For $z < 0$, $\dfrac{\partial^2 h_\lambda^{-1}(z)}{\partial z^2} = -(1 - \lambda)\ (1 - \lambda z)^{1/\lambda - 2}$. If $0 < \lambda < 1$, then $\dfrac{\partial^2 h_\lambda^{-1}(z)}{\partial z^2} < 0$. Therefore Y is a concave function of Z for all $Z < 0$. Similarly for $z > 0$

$$\frac{\partial^2 h_\lambda^{-1}(z)}{\partial z^2} = (1 - \lambda)\ (1 + \lambda z)^{1/\lambda - 2}$$

which is positive if $0 < \lambda < 1$. Therefore Y is a convex function of Z for all $Z > 0$. []

Although the family of modulus power transformations solves the problem of how to deal with negative data and, by the above theorem, can be used to squeeze the tails of heavy-tailed symmetric distributions so that they become more normal, yet there are some other situations under which it is expected that the John-Draper transformation cannot be appropriate. For example it may be the case that both tails are heavy but to different degrees. In such a case we would want a transformation that squeezes the two tails differently. In some other cases we may want to leave one tail as is and change only the other tail. For such situations we introduce a new family of transformations which we call the two-domain family and which may be expected to deal properly with types of data like those described in the two cases above.

2.2.4 The two-domain family of transformations

This family is a generalization of the John-Draper family in which λ is a two-dimensional vector.

$$Y(\theta, \lambda_1, \lambda_2) = \begin{cases} (1 - (\theta - Y + 1)^{\lambda_1})/\lambda_1 & \text{for } Y \leq \theta \\ ((Y - \theta + 1)^{\lambda_2} - 1)/\lambda_2 & \text{for } Y > \theta \end{cases} \qquad (2.2.5)$$

This formula is used if $\lambda_1 \neq 0$ and $\lambda_2 \neq 0$. If $\lambda_1 = 0$, then replace $(1 - (\theta\text{-}Y+1)^{\lambda_1})/\lambda_1$ by $-\ln(\theta\text{-}Y+1)$. If $\lambda_2 = 0$, then replace $((Y-\theta+1)^{\lambda_2} - 1)/\lambda_2$ by $\ln(Y-\theta+1)$.

Note that the two-domain family is monotone increasing in Y and is continuous at $Y = \theta$. Also note that the domain of both the John-Draper and the two-domain transformations is the whole real line, so they are applicable for all data sets. Also, when λ_1 and λ_2 are both positive, the range of the transformation is the whole real line, which allows the normality assumption of the transformed variable to be valid.

The notion of using a different transformation (shape) parameter for each tail is also considered by Stukle (1988) in a different setting, where he considers different shape parameters in defining the negative and positive parts of the logit link function of a Bernoulli random variable under a generalized logistic regression model.

2.3 The Problem of Estimating the Transformation Parameter

In the analysis of any statistical inference problem based on normal-theory techniques, if it were known that the data come from a certain known transformation of a normal random variable, then certainly the analysis of the problem based on the data that are transformed back to normality will be optimal. Unfortunately, when there is some evidence that some transformation should be considered, one never knows the true value of the transformation parameter. The best that can be done is then to consider a data-based estimator of the transformation. Usually the Box-Cox maximum likelihood procedure or Hinkley's estimator are considered in this instance. But what is the effect of using such an estimate of the transformation parameter, instead of the true unknown transformation on the underlying statistical analysis? This question was raised by Hinkley (1977, p.69) who said "no published results exist concerning the effect of transformation estimation on subsequent analysis". Since then, a number of papers largely inspired by Bickel and Doksum (1981), have studied this issue for estimation in a model which is assumed to be a linear model after a power transformation. Bickel and Doksum argued that there is a high correlation between the estimate of the transformation parameter and estimates of the other model parameters. Such a correlation leads to inflation of the variances of the model estimates compared to the situation in which the transformation is known. On the other hand, Box and Cox (1982), Hinkley and Runger (1984), Carroll and Ruppert (1984) and Taylor (1986) argued that, although there may be some effect due to considering a data-based estimate of the transformation parameter, it is not as severe as pictured by Bickel and Doksum. Doksum and Wong (1983) pointed out that the reason for the argument given by Bickel and Doksum (1981) is that they neglect the Jacobian of transforming y into $h_{\lambda}(y)$ and an agreement with the other argument can be obtained if this Jacobian is considered. The t-test statistic being invariant under multiplication by constants Doksum and Wong (1983) were able to prove some asymptotic results concerning the transformed t-test as discussed in the next section.

2.4 Testing for Location when the Original Data are Transformed

Doksum and Wong (1983) consider the problem of testing the equality of distributions from two independent samples. They were able to prove that the transformed t-test using a consistent data-based estimator of the transformation parameter is asymptotically as efficient as the transformed t-test when the transformation is known. So whatever results hold for the known transformation situation are asymptotically valid for the estimated transformation situation. Using the asymptotic theory thus made available they found that there is a considerable gain in efficiency of the transformed t-test relative to the t-test in the original scale under the log-normal, log-double-exponential, Student's t, contaminated normal, gamma and exponential models. In their simulation work, Doksum and Wong found that there is indeed some gain in the simulated power when the transformed t-test is used compared to the simulated power of the t-test of the original observations, but for smaller samples the gain is not substantial. The simulated power of a transformed γ-trimmed t-test for $\gamma = .10$ and $.20$ was much higher than both the transformed and the untransformed t-tests.

Doksum and Wong (1980) state that the asymptotic argument they give is valid for the comparison of two means or more generally under an analysis-of-variance model when the hypothesis of interest is in the form of a contrast. On the other hand this argument fails for single random sample problems.

3. ASYMPTOTIC RESULTS FOR TEST STATISTICS UNDER SYMMETRIC MODELS

In Section 2.1 we showed that the t-test is conservative when applied to a set of data that has a symmetric distribution with heavy tails, and hence does not have good power properties relative to other testing procedures. Theorem 2.2.2 shows that when the John-Draper family of transformations is applied with $0 < \lambda < 1$ to a heavy-tailed set of data, it symmetrically squeezes both tails so that the distribution of the data in the transformed scale has lighter tails. In this case it would be expected that the t-test from the transformed data is less conservative than that from the original data and hence has better power properties.

In this chapter we derive the Pitman asymptotic relative efficiency of the t-test in the original scale of Y relative to the transformed t-test using the John-Draper transformation. The asymptotic results indicate that there is a considerable gain in efficiency if the transformed t-test is used for data with heavy-tailed symmetric distributions.

In Section 3.1 to simplify reference we reintroduce the John-Draper family and the different test statistics involved throughout this chapter. In Section 3.2 asymptotic properties of the t-statistic assuming the true transformation is known are derived. Section 3.3 deals with the asymptotics when the transformation is unknown. In Section 3.4 we use numerical integration to evaluate the asymptotic relative efficiency for the transformed-normal, the transformed-contaminated-normal and the transformed-Student's-t models.

3.1 Definitions and Notation

Throughout this chapter we use λ_* to denote the true value of the transformation parameter λ. We assume that λ_* is greater than zero. We use $\hat{\lambda}_n$ to denote the maximum likelihood estimator of λ, which is shown to be consistent and asymptotically normal in Appendix E. θ is used to denote the median of Y. The asymptotic results are derived for the problem of testing the hypotheses given by (2.1.1). Similar results for testing against $\theta < \theta_0$ or $\theta \neq \theta_0$ can be derived along the lines of this chapter.

The derivation of the Pitman asymptotic relative efficiency is based on testing the null hypothesis against local (contiguous) alternatives, that is, alternatives of the form H_1: $\theta = \theta_n$ where

$$\theta_n = \theta_0 + k/\sqrt{n} \qquad k > 0 \tag{3.1.1}$$

so that in the limit θ_n tends to θ_0.

Recall that the John-Draper transformation with central parameter θ is defined in (2.2.4) for $\lambda \neq 0$ as

$$h(Y-\theta,\lambda) = \operatorname{sign}(Y-\theta)[(|Y-\theta|+1)^{\lambda} - 1]/\lambda .$$

The main assumption under this transformation is that for some θ, σ and λ,

$$h(Y-\theta,\lambda) = \sigma\epsilon \tag{3.1.2}$$

where ϵ has a standard symmetric distribution, that is,

$$\text{i. } \epsilon \overset{\mathcal{L}}{=} -\epsilon \qquad \text{ii. } E(\epsilon) = 0 \qquad \text{iii. } E(\epsilon^2) = 1 \tag{3.1.3}$$

and also satisfies

iv. the cdf F_ϵ is continuous at 0 .

A test statistic denoted by $T_n(\lambda)$ is defined as follows:

$$T_n(\lambda) = \sqrt{n}\,\bar{h}_n(\theta_0,\lambda)/\,\hat{\sigma}_n(\theta_0,\lambda)$$

where,

$$\bar{h}_n(\theta_0,\lambda) = \sum_{i=1}^{n} h(Y_i-\theta_0,\lambda)/n$$
$$\hat{\sigma}_n^2(\theta_0,\lambda) = \sum_{i=1}^{n} (h(Y_i-\theta_0,\lambda) - \bar{h}(\theta_0,\lambda))^2/(n-1) \; . \tag{3.1.4}$$

Since $h(Y-\theta_0,1) = Y-\theta_0$, thus $T_n(1)$ is used to denote the t-test in the original scale. $T_n(\lambda_*)$ is used to denote the t-test when the true transformation is known and $T_n(\hat{\lambda}_n)$ is the t-test using the MLE of λ . On the other hand if we use σ instead of $\hat{\sigma}_n(\theta_0,\lambda)$ we denote the test by $\tilde{T}_n(\lambda)$. We use $\Psi_n(\theta)$ and $\tau_n^2(\theta)$ to denote either $E_\theta(T_n(\lambda))$ and $\mathrm{var}_\theta(T_n(\lambda))$, or some approximations of them such that the regularity conditions C1 and C2 in Appendix B hold in the limit.

3.2 Asymptotics When λ is Known

In this section we derive the asymptotic properties of the test statistics $T_n(\lambda_*)$ and $T_n(1)$ under both the null and contiguous alternative models and show that both of them are Pitman regular.

3.2.1 Asymptotic distribution of $T_n(\lambda_*)$ under H_o

Assume that under H_0 and for some λ_* , $h(Y-\theta_0,\lambda_*)$ satisfies the assumption of model (3.1.2), that is,

$$h(Y-\theta_0,\lambda_*) = \sigma\,\epsilon$$

where ϵ satisfies (3.1.3). Since $\bar{h}_n(\theta_0,\lambda_*) = \sigma\,\bar{\epsilon}$ and since by the WLLN $\bar{\epsilon} \xrightarrow{p} 0$, hence

$$\bar{h}_n(\theta_0,\lambda_*) \xrightarrow{p} 0 \; . \tag{3.2.1}$$

Since by the WLLN $\sum_{i=1}^{n} \epsilon_i^2/n \xrightarrow{p} 1$, hence

$$(1/n)\sum_{i=1}^{n} h^2(Y_i-\theta_0,\lambda_*) = \sigma^2 \sum_{i=1}^{n} \epsilon_i^2/n \xrightarrow{p} \sigma^2. \tag{3.2.2}$$

Now (3.2.1) and (3.2.2) imply that $\hat{\sigma}_n^2(\theta_0,\lambda_*) \xrightarrow{P} \sigma^2$ and by Fact 4 of Appendix A

$$\hat{\sigma}_n(\theta_0,\lambda_*) \xrightarrow{P} \sigma \, . \tag{3.2.3}$$

By the CLT

$$\tilde{T}_n(\lambda_*) = \sqrt{n}\,\bar{h}_n(\theta_0,\lambda_*)/\sigma \xrightarrow{\mathcal{L}} N(0,1) \, . \tag{3.2.4}$$

Using (3.2.3), (3.2.4) and Fact (2-ii) of Appendix A we conclude that

$$T_n(\lambda_*) = \sqrt{n}\,\bar{h}_n(\theta_0,\lambda_*)/\,\hat{\sigma}_n(\theta_0,\lambda_*) \xrightarrow{\mathcal{L}} N(0,1) \, . \tag{3.2.5}$$

3.2.2 Asymptotic distribution of $T_n(\lambda_*)$ under contiguous alternatives

As mentioned earlier we only need to consider local alternatives as those defined in (3.1.1), that is, alternatives of the form H_1: $\theta = \theta_n = \theta_0 + k_1/\sqrt{n}$, for some $k_1 > 0$. Under such alternatives we assume that $Y_{n1}, Y_{n2}, \ldots, Y_{nn}$ are independently and identically distributed (iid) with distribution function F_{θ_n} which depends on n. We also assume that for a given n

$$h(Y_{ni}-\theta_n,\lambda_*) = \sigma\,\epsilon_{ni} \tag{3.2.6}$$

where $\epsilon_{n1}, \epsilon_{n2}, \ldots, \epsilon_{nn}$ are iid F_ϵ which does not depend on n. The safest way to deal with Y_{ni} so that the double subscripts do not cause confusion is to transform Y_{ni} to ϵ_{ni}. Since the distribution of ϵ_{ni} does not depend on n, it is safe to write ϵ_{ni} as ϵ_i. In the proofs of parts (i) and (ii) of Lemma 3.2.1 we will use double subscripts but thereafter we change to single subscripts since double subscripts are too cumbersome.

The first step in deriving the asymptotic distribution of $T_n(\lambda_*)$ under contiguous alternatives is to express $h(Y-\theta_0,\lambda_*)$ in terms of $h(Y-\theta_n,\lambda_*)$. For $Y \le \theta_0$ we can write

$$h(Y-\theta_0,\lambda_*) = [1 - (\theta_n-Y+1)^{\lambda_*}\left(1 - (k_1/\sqrt{n})/(\theta_n-Y+1)\right)^{\lambda_*}]/\lambda_* .$$

The Maclaurin expansion of $(1-t)^{\lambda}$ for $|t| < 1$ is

$$(1-t)^{\lambda} = 1 - \lambda t + \sum_{j=2}^{\infty} \frac{(-1)^j}{j!} [\prod_{m=0}^{j-1}(\lambda-m)]\ t^j . \tag{3.2.7}$$

Since we are interested in asymptotics we can suppose $k_1/\sqrt{n} < 1$. Since $Y \le \theta_0$ implies $Y < \theta_n$, hence for all $Y \le \theta_0$, $\theta_n-Y+1 > 1$ and a Maclaurin expansion of $\left(1 - (k_1/\sqrt{n})/(\theta_n-Y+1)\right)^{\lambda_*}$ is absolutely convergent. Therefore

$$h(Y-\theta_0,\lambda_*) = h(Y-\theta_n,\lambda_*) + (k_1/\sqrt{n})(\theta_n-Y_i+1)^{\lambda_*-1} - (1/\lambda_*)\sum_{j=2}^{\infty}((-k_1/\sqrt{n})^j/j!)(\prod_{m=0}^{j-1}(\lambda_*-m))(\theta_n-Y+1)^{\lambda_*-j} . \tag{3.2.8}$$

Similarly for $Y \ge \theta_n$ it can be shown that

$$h(Y-\theta_0,\lambda_*) = h(Y-\theta_n,\lambda_*) + (k_1/\sqrt{n})\ (Y-\theta_n+1)^{\lambda_*-1} + (1/\lambda_*)\ \sum_{j=2}^{\infty}((k_1/\sqrt{n})^j/j!)(\prod_{m=0}^{j-1}(\lambda_*-m)\ (Y-\theta_n+1)^{\lambda_*-j} . \tag{3.2.9}$$

Let

$$A_n = \{Y : \theta_0 < Y < \theta_n\} \tag{3.2.10}$$

and let A_n^c denote the complement of A_n . From (3.2.8) and (3.2.9) and for all $Y \in A_n^c$ we express $h(Y-\theta_0,\lambda_*)$ as

$$h(Y-\theta_0,\lambda_*) = h(Y-\theta_n,\lambda_*) + (k_1/\sqrt{n})(|Y-\theta_n|+1)^{\lambda_*-1} + \sum_{j=2}^{\infty} \text{sign}(Y-\theta_n)^{j-1}((k_1/\sqrt{n})^j/j!)(\prod_{m=1}^{j-1}(\lambda_*-m))\ (|Y-\theta_n|+1)^{\lambda_*-j} . \tag{3.2.11}$$

Note that under model (3.2.6) we can write,

$$|Y - \theta_n| + 1 = (1 + \lambda_*\sigma|\epsilon|)^{1/\lambda_*} . \tag{3.2.12}$$

Also note that $\text{sign}(Y-\theta_n) = \text{sign}(\epsilon)$. Using (3.2.11) and (3.2.12) we express $h(Y-\theta_0,\lambda_*)$ as a function of ϵ, for $Y \in A_n^c$, in the following two forms that we will need later:

1. $h(Y-\theta\ ,\lambda_*) = \sigma\epsilon + R_1(\epsilon,\lambda_*,n)$ (3.2.13)

where

$$R_1(\epsilon,\lambda_*,n) = \sum_{j=1}^{\infty}\Big\{(\text{sign}(\epsilon))^{j-1}((k_1/\sqrt{n})^j/j!)\prod_{m=1}^{j-1}(\lambda_*-m)(1+\lambda_*\sigma|\epsilon|)^{1-(j/\lambda_*)}\Big\}. \tag{3.2.14}$$

2. $h(Y-\theta_0,\lambda_*) = \sigma\epsilon + (k_1/\sqrt{n})(1+\lambda_*\sigma|\epsilon|)^{1-(1/\lambda_*)} + R_2(\epsilon,\lambda_*,n)$ (3.2.15)

where

$$R_2(\epsilon,\lambda_*,n) = \sum_{j=2}^{\infty}\Big\{(\text{sign}(\epsilon))^{j-1}((k_1/\sqrt{n})^j/j!)\prod_{m=1}^{j-1}(\lambda_*-m)(1+\lambda_*\sigma|\epsilon|)^{1-(j/\lambda_*)}. \tag{3.2.16}$$

Let

$$B_n = \{\epsilon:\ [1-(1+k_1/\sqrt{n})^{\lambda_*}]/\lambda_*\sigma < \epsilon < 0\}. \tag{3.2.17}$$

Under the model (3.2.6) it is easy to see that $Y \in A_n$ iff $\epsilon \in B_n$.

The results of the following two lemmas will be used in deriving the asymptotic distribution of $T_n(\lambda_*)$. The proofs of these lemmas are given in Appendix C.

Lemma 3.2.1 Let A_n and B_n be as defined in (3.2.10) and (3.2.17) respectively. Then under model (3.2.6) as $n \longrightarrow \infty$,

i. $\Pr\{Y \in A_n\} = \Pr\{\epsilon \in B_n\} \longrightarrow 0$

ii. $\sqrt{n}\,\{(1/n)\sum_{Y_i \in A_n} h(Y_i-\theta_0,\lambda_*)\} \xrightarrow{p} 0$

iii. $(1/n)\sum_{Y_i \in A_n} h^2(Y_i-\theta_0,\lambda_*) \xrightarrow{p} 0$

iv. $\sqrt{n}\,[1/n \sum_{\epsilon_i \in B_n} \epsilon_i] \xrightarrow{p} 0$

v. $(1/n) \sum_{\epsilon_i \in B_n} \epsilon_i^2 \xrightarrow{p} 0$

vi. $(1/n) \sum_{\epsilon_i \in B_n} (1 + \lambda_* \sigma |\epsilon_i|)^{1-1/\lambda_*} \xrightarrow{p} 0$

<u>Lemma 3.2.2</u> Let $R_1(\epsilon,\lambda_*,n)$ and $R_2(\epsilon,\lambda_*,n)$ be as defined in (3.2.14) and (3.2.16) respectively. Then, as $n \longrightarrow \infty$,

i. $(1/n) \sum_{\epsilon_i \in B_n^c} \epsilon_i R_1(\epsilon_i,\lambda_*,n) \xrightarrow{p} 0$

ii. $(1/n) \sum_{\epsilon_i \in B_n^c} R_1^2(\epsilon_i,\lambda_*,n) \xrightarrow{p} 0$

iii) $\sqrt{n}\,(1/n) \sum_{\epsilon_i \in B_n^c} R_2(\epsilon_i,\lambda_*,n) \xrightarrow{p} 0$.

<u>Theorem 3.2.1</u> Under the alternative model (3.2.6) $\hat{\sigma}_n(\theta_0,\lambda_*)$ is a consistent estimator of σ .

<u>Proof:</u> From (3.2.14) write

$$h^2(Y_i-\theta_0,\lambda_*) = \sigma^2\epsilon_i^2 + 2\sigma\epsilon_i R_1(\epsilon_i,\lambda_*,n) + R_1^2(\epsilon_i,\lambda_*,n)$$

and hence

$$(1/n) \sum_{Y_i \in A_n^c} h^2(Y_i-\theta_0,\lambda_*) - \sigma^2(1/n) \sum_{\epsilon_i \in B_n^c} \epsilon_i^2 =$$

$$2\sigma(1/n) \sum_{\epsilon_i \in B_n^c} \epsilon_i R_1(\epsilon_i,\lambda_*,n) + (1/n) \sum_{\epsilon_i \in B_n^c} R_1^2(\epsilon_i,\lambda_*,n) .$$

It follows from Lemma (3.2.2) parts i and ii that

$$(1/n) \sum_{Y_i \in A_n^c} h^2(Y_i-\theta_0,\lambda_*) - \sigma^2(1/n) \sum_{\epsilon_i \in B_n^c} \epsilon_i^2 \xrightarrow{p} 0 . \quad (3.2.18)$$

By the WLLN and Lemma 3.2.1(v),

$$(1/n) \sum_{\epsilon_i \in B_n^c} \epsilon_i^2 = (1/n) \sum_{i=1}^{n} \epsilon_i^2 - (1/n) \sum_{\epsilon_i \in B_n} \epsilon_i^2 \xrightarrow{p} 1 - 0 . \quad (3.2.19)$$

By Fact (1) Appendix A, (3.2.18) and (3.2.19) imply

$$(1/n) \sum_{Y_i \in A_n^c} h^2(Y_i-\theta_0,\lambda_*) \xrightarrow{p} \sigma^2 . \quad (3.2.20)$$

Write

$$(1/n)\sum_{i=1}^{n} h^2(Y_i-\theta_0,\lambda_*) = (1/n)\sum_{Y_i\in A_n} h^2(Y_i-\theta_0,\lambda_*)$$

$$+ (1/n)\sum_{Y_i\in A_n^c} h^2(Y_i-\theta_0,\lambda_*) \ . \qquad (3.2.21)$$

Then Lemma 3.2.1 (iii) , (3.2.20) and (3.2.21) imply that

$$(1/n)\sum_{i=1}^{n} h^2(Y_i-\theta_0,\lambda_*) \xrightarrow{p} \sigma^2 \ . \qquad (3.2.22)$$

A similar argument shows that $\bar{h}_n(\theta_0,\lambda_*) \xrightarrow{p} 0$ and hence, by Fact (4) in Appendix A, that

$$\bar{h}_n^2(\theta_0,\lambda_*) \xrightarrow{p} 0 \ . \qquad (3.2.23)$$

Formula (3.1.4) can be expressed as

$$\hat{\sigma}_n^2(\theta_0,\lambda_*) = (n/(n-1))[(1/n)\sum_{i=1}^{n} h^2(Y_i-\theta_0,\lambda_*) - \bar{h}_n^2(\theta_0,\lambda_*)] \ . \qquad (3.2.24)$$

Now (3.2.22), (3.2.23) and (3.2.24) imply $\hat{\sigma}_n^2(\theta_0,\lambda_*) \xrightarrow{p} \sigma^2$ and hence

$$\hat{\sigma}_n(\theta_0,\lambda_*) \xrightarrow{p} \sigma \ . \quad [] \qquad (3.2.25)$$

In the following theorem we prove the asymptotic normality of the transformed t-test under contiguous alternatives.

<u>Theorem 3.2.2</u>

Let $\Psi = (k_1/\sigma)\, E[(1 + \lambda_*\sigma|\epsilon|)^{1-1/\lambda_*}]$. Under contiguous alternatives

$$T_n(\lambda_*) - \Psi \xrightarrow{\mathcal{L}} N(0,1) \ .$$

<u>Proof:</u> Since

$$T_n(\lambda_*) - \Psi = (\sigma/\hat{\sigma}_n)[\tilde{T}_n(\lambda_*) - \Psi] + [(\sigma/\hat{\sigma}_n) -1]\,\Psi \ ,$$

and since $(\sigma/\hat{\sigma}_n) \xrightarrow{p} 1$ by Theorem 3.2.1, it suffices to show $\tilde{T}_n(\lambda_*) - \Psi \xrightarrow{\mathcal{L}} N(0,1)$. Since

$$\sqrt{n}\,\bar{h}_n(\theta_0,\lambda_*) - \sqrt{n}\sum_{Y_i\in A_n^c} h(Y_i-\theta_0,\lambda_*)/n = \sqrt{n}\sum_{Y_i\in A_n} h(Y_i-\theta_0,\lambda_*)/n \ ,$$

by Lemma 3.2.1 (ii)

$$\sqrt{n}\,\bar{h}_n(\theta_0,\lambda_*) - (\sqrt{n}/n) \sum_{Y_i \in A_n^c} h(Y_i-\theta_0,\lambda_*) \xrightarrow{p} 0 \ . \qquad (3.2.26)$$

Similarly Lemma 3.2.1 (iv) implies

$$\sigma \sqrt{n}\,\bar{\epsilon}_n - \sigma \sqrt{n}\,(1/n \sum_{\epsilon_i \in B_n^c} \epsilon_i) \xrightarrow{p} 0 \ . \qquad (3.2.27)$$

Finally Lemma 3.2.1(vi) implies

$$(k_1/n) \sum_{i=1}^{n} (1+\lambda_*\sigma|\epsilon_i|)^{1-1/\lambda_*} - (k_1/n) \sum_{\epsilon_i \in B_n^c} (1+\lambda_*\sigma|\epsilon_i|)^{1-1/\lambda_*} \xrightarrow{p} 0 \ . \qquad (3.2.28)$$

By the CLT,

$$\sigma \sqrt{n}\,\bar{\epsilon}_n \xrightarrow{\mathscr{L}} N(0,\sigma^2) \ . \qquad (3.2.29)$$

By the WLLN

$$(k_1/n) \sum_{i=1}^{n} (1+\lambda_*\sigma|\epsilon|)^{1-1/\lambda_*} \xrightarrow{p} k_1\ E[(1+\lambda_*\sigma|\epsilon|)^{1-1/\lambda_*}] = \sigma\Psi \ . \qquad (3.2.30)$$

A proof that the above expectation is finite is given in Lemma C.1 of Appendix C. From (3.2.15) we can write

$$\sqrt{n} \sum_{Y_i \in A_n^c} h(Y_i-\theta_0,\lambda_*)/n - \sigma\sqrt{n} \sum_{\epsilon_i \in B_n^c} \epsilon_i/n - (k_1/n) \sum_{\epsilon_i \in B_n^c} (1+\lambda_*\sigma|\epsilon_i|)^{1-1/\lambda_*}$$

$$= \sqrt{n}\,(1/n) \sum_{\epsilon_i \in B_n^c} R_2(\epsilon_i,\lambda_*,n) \ . \qquad (3.2.31)$$

Apply Lemma 3.2.2 (iii) to conclude that the right hand side of (3.2.31) converges in probability to 0 . The result of the theorem follows from a chain of substitutions of (3.2.26) through (3.2.30) in (3.2.31). []

3.2.3 The transformed t-test is Pitman regular

The following theorem shows that the transformed t-test statistic using the John-Draper transformation is Pitman regular.

Theorem 3.2.3 The test statistic $T_n(\lambda_*)$ is Pitman regular with

$$R_n^2(\theta_0) = (n/\sigma^2)\{E[(1+\lambda_*\sigma|\epsilon|)^{1-1/\lambda_*}]\}^2 .$$

Proof: We first need candidates for $\Psi_n(\theta)$ and $\tau_n^2(\theta)$ in Appendix B. These are obtained as approximations for the functions $E_\theta(\bar{T}_n(\lambda_*))$ and $\mathrm{var}_\theta(\bar{T}_n(\lambda_*))$, respectively. Since we are interested only in contiguous alternatives, we approximate the above two functions for θ in a neighborhood of θ_0 . Consider a Taylor expansion of $\bar{T}_n(\lambda_*)$ about $\theta_0 = \theta$,

$$(\sqrt{n}/\sigma)\bar{h}_n(\theta_0,\lambda_*) \approx (\sqrt{n}/\sigma)\bar{h}_n(\theta,\lambda_*) + (\sqrt{n}/\sigma)\frac{\partial\bar{h}_n(\theta,\lambda_*)}{\partial\theta}(\theta_0-\theta) .$$

Under model (3.1.2) $h(Y_i-\theta,\lambda_*) = \sigma\,\epsilon_i$ ans so

$$\bar{T}_n(\lambda_*) \approx \sqrt{n}\,\bar{\epsilon} + (\sqrt{n}/\sigma)(\theta-\theta_0)[(1/n)\sum_{i=1}^{n}(1+\lambda_*\sigma|\epsilon_i|)^{1-1/\lambda_*}] . \tag{3.2.32}$$

Taking the expectation of the right side of (3.2.32) we define

$$\Psi_n(\theta) = (\sqrt{n}/\sigma)(\theta-\theta_0)\ E[(1+\lambda_*\sigma|\epsilon|)^{1-1/\lambda_*}] . \tag{3.2.33}$$

Taking the variance of the first term only, we define

$$\tau_n^2(\theta) = \mathrm{var}(\sqrt{n}\,\bar{\epsilon}) = 1 . \tag{3.2.34}$$

We now verify the seven regularity conditions of Appendix B.

C1. Note that $\Psi_n(\theta_0) = 0$. The asymptotic normality follows from (3.2.5).

C2. For $\theta = \theta_n$ (3.2.33) becomes

$$\Psi_n(\theta_n) = (k_1/\sigma)\ E[(1 + \lambda_*\sigma|\epsilon|)^{1-1/\lambda_*}] = \Psi ,$$

where Ψ is as in Theorem 3.2.2. So the result of the theorem verifies the asymptotic normality under contiguous alternatives.

C3. From (3.2.33) $\Psi_n(\theta)$ is differentiable for all θ .

C4. $\Psi_n'(\theta) = (\sqrt{n}/\sigma)\ E[\ (1+\lambda_*\sigma|\epsilon|)^{1-1/\lambda_*}] = \sqrt{n}\ \Psi/k_1$.

Since $(1+\lambda_*\sigma|\epsilon|)^{1-1/\lambda_*} > 0$ for all $\epsilon \in R^1$, hence

$$E[(1+\lambda_* |\epsilon|)^{1-1/\lambda_*}] > 0 . \quad \text{So} \quad \Psi_n'(\theta) > 0 .$$

C5. $\Psi_n'(\theta_0)/\sqrt{n} = \Psi/k_1 > 0$.

C6. $\Psi_n'(\theta)$ is the same for all θ hence

$$\sup_{\theta_0 \le \theta^* \le \theta_n} |\Psi_n'(\theta_n^*)/\Psi_n'(\theta_0) - 1| = 0 \quad \text{for all} \quad n .$$

C7. $\tau_n(\theta_n)/\tau_n(\theta_0) = 1$ for all n .

Hence the test statistic $T_n(\lambda_*)$ is Pitman regular and from Appendix B

$$R_n^2(\theta_0) = n \{E [(1+\lambda_*\sigma|\epsilon|)^{1-1/\lambda_*}]\}^2/\sigma^2 . \quad [] \tag{3.2.35}$$

3.2.4 Asymptotics for the t-test in original scale

In the following we derive the asymptotic distribution of the t-test statistic in the original scale denoted as $T_n(1)$ under both the null and alternative models and show that it is Pitman regular. Recall that the test statistic in the original scale is defined as

$$T_n(1) = \sqrt{n} (\bar{Y} - \theta_0)/S_Y \tag{3.2.36}$$

where $\bar{Y} = \sum_{i=1}^{n} Y_i/n$ and $S_Y^2 = \sum_{i=1}^{n} (Y_i - \bar{Y})^2/(n-1)$. We can write $T_n(1) = (\sigma_Y/S_Y)\bar{T}_n(1)$ where

$$\bar{T}_n(1) = \sqrt{n} (\bar{Y}-\theta_0)/\sigma_Y \tag{3.2.37}$$

and σ_Y^2 denotes the true variance of Y . Under model (3.1.2)

$$h(Y-\theta, \lambda_*) = \sigma \epsilon$$

$$Y = \theta + \text{sign}(\epsilon)[(1+\lambda_*\sigma|\epsilon|)^{1/\lambda^*} - 1] \tag{3.2.38}$$

Note that under symmetric distributions of ϵ , $\text{sign}(\epsilon)[(1+\lambda_*\sigma|\epsilon|)^{1/\lambda^*}-1]$ is an odd function and hence has zero expectation. Hence

$$\sigma_Y^2 = \mathrm{var}(Y) = \mathrm{var}(\theta + \mathrm{sign}(\epsilon)\ [(1+\lambda_*\sigma|\epsilon|)^{1/\lambda_*} - 1])$$
$$= E[\ [(1+\lambda_*\sigma|\epsilon|)^{1/\lambda_*} - 1]^2\ ]\ . \tag{3.2.39}$$

Under H_0, the CLT implies $\bar{T}_n(1) \xrightarrow{\mathcal{L}} N(0,1)$. We know $\sigma_Y/S_Y \longrightarrow 1$ and so

$$T_n(1) \xrightarrow{\mathcal{L}} N(0,1)\ . \tag{3.2.40}$$

Under H_1, $\theta = \theta_n = \theta_0 + k/\sqrt{n}$ and

$$\bar{T}_n(1) = \frac{\sqrt{n}(\bar{Y}-\theta_n+k/\sqrt{n})}{\sigma_Y} = \frac{\sqrt{n}(\bar{Y}-\theta_n)}{\sigma_Y} + \frac{(k/\sqrt{n})}{\sigma_Y}\ .$$

By the CLT, $\sqrt{n}\ (\bar{Y}-\theta_n)/\sigma_Y \xrightarrow{\mathcal{L}} N(0,1)$. Hence

$$T_n(1) \xrightarrow{\mathcal{L}} N(k/\sigma_Y,1) \quad \text{under } H_1. \tag{3.2.41}$$

<u>Theorem 3.2.4</u> The test statistic $T_n(1)$ is Pitman regular with $R_n^2(\theta_0) = n/\sigma_Y^2$.

<u>Proof :</u> We will verify the seven conditions of Appendix B using $\Psi_n(\theta) = (\sqrt{n}/\sigma_Y)(\theta-\theta_0)$ and $\tau_n^2(\theta) = 1$.

C1. Since $\Psi_n(\theta_0) = 0$, C1 follows from (3.2.40).

C2. Note $\Psi_n(\theta_n) = k/\sigma_Y$ for all n . C2 follows from (3.2.41).

C3. $\Psi_n(\theta) = (\sqrt{n}/\sigma_Y)(\theta-\theta_0)$ is differentiable with $\Psi_n'(\theta) = \sqrt{n}/\sigma_Y$.

C4. $\sqrt{n}/\sigma_Y > 0$.

C5. $\Psi_n'(\theta)/\sqrt{n} = 1/\sigma_Y > 0$.

C6. $\Psi_n'(\bar{\theta})/\Psi_n'(\theta_0) = 1$ for all $\bar{\theta}$ and all n .

C7. $\tau_n(\theta) = 1$ for all θ and all n .

Therefore $T_n(1)$ is Pitman regular and from Appendix B

$$R_n^2(\theta_0) = \Psi_n'^2(\theta_0)/\tau_n^2(\theta_0) = n/\sigma_Y^2 \quad [] \tag{3.2.42}$$

Now we compute the asymptotic relative efficiency of the t-test in the original scale, $T_{1n} = T_n(1)$ relative to the transformed t-test,

$T_{2n} = T_n(\lambda_*)$. From (3.2.35), $R_{2n}^2 = (n/\sigma^2)\{E[(1+\lambda_*\sigma|\epsilon|)^{1-1/\lambda_*}]\}^2$ and from (3.2.42) $R_{1n}^2 = n/\sigma_Y^2$. Therefore

$$\mathrm{ARE}(T_n(1)\ ,\ T_n(\lambda_*)) = \lim_{n\to\infty} R_{1n}^2/R_{2n}^2$$
$$= \sigma^2/\ \sigma_Y^2\{E[(1+\lambda_*\sigma|\epsilon|)^{1-1/\lambda_*}]\}^2 \qquad (3.2.44)$$

where σ_Y^2 can be evaluated using (3.2.39) .

3.3 Asymptotics When λ is Unknown

In practice the true value of the transformation parameter λ_* will not be known and a data-based estimator $\hat{\lambda}_n$ of λ (such as the MLE) may be considered. In this section we show that if $\hat{\lambda}_n$ is a consistent estimator of λ, then the test statistic $T_n(\hat{\lambda}_n)$ is asymptotically equivalent to the test statistic $T_n(\lambda_*)$ where

$$T_n(\hat{\lambda}_n) = \sqrt{n}\,\bar{h}_n(\theta_0,\hat{\lambda}_n)/\hat{\sigma}_n(\theta_0,\hat{\lambda}_n) \qquad (3.3.1)$$

By the asymptotic equivalence of the two test statistics $T_n(\hat{\lambda}_n)$ and $T_n(\lambda_*)$ we mean that the difference between them converges to 0 in probability under both the null and contiguous alternative models.

Lemma 3.3.1 Under both H_0 and H_1,

i. $\bar{h}_n(\theta_0,\hat{\lambda}_n) - \bar{h}_n(\theta_0,\lambda_*) \xrightarrow{p} 0$ (3.3.2)

ii. $(1/n)\sum_{i=1}^{n} h^2(Y_i-\theta_0,\hat{\lambda}_n) - (1/n)\sum_{i=1}^{n} h^2(Y_i-\theta_0,\lambda_*) \xrightarrow{p} 0$ (3.3.3)

The proof of the above lemma is given in Appendix C.

Theorem 3.3.1 Under both H_0 and H_1, $\hat{\sigma}_n^2(\theta_0,\hat{\lambda}_n)$ is a consistent estimator of σ^2 .

Proof: Note

$$\hat{\sigma}_n^2(\theta_0,\hat{\lambda}_n) = 1/(n-1)\Big[\ \sum_{i=1}^{n} h^2(Y_i-\theta_0,\hat{\lambda}_n) - n\bar{h}_n^2(\theta_0,\hat{\lambda}_n)\ \Big]\ .$$

In Section 3.2.1 and Theorem 3.2.2 we showed that $\sqrt{n}\,\bar{h}_n(\theta_0,\lambda_*)/\sigma$

has a limiting normal distribution under the null and alternative models respectively. Dividing by $\sqrt{n}$ and applying Fact 2(i) in Appendix A we conclude that

$$\bar{h}_n(\theta_0, \lambda_*) \xrightarrow{P} 0 .$$

By (3.3.2) and Fact 1 in Appendix A

$$\bar{h}_n(\theta_0, \hat{\lambda}_n) \xrightarrow{P} 0 .$$

By Fact 4 in Appendix A we conclude that

$$\bar{h}_n^2(\theta_0, \hat{\lambda}_n) \xrightarrow{P} 0 . \tag{3.3.4}$$

From Section 3.2.1 and Theorem 3.2.1 we have

$$(1/n) \sum_{i=1}^{n} h^2(Y_i - \theta_0, \lambda_*) \xrightarrow{P} \sigma^2$$

under both the null and alternative models. Hence by Lemma 3.3.2 (ii)

$$(1/n) \sum_{i=1}^{n} h^2(Y_i - \theta_0, \hat{\lambda}_n) \xrightarrow{P} \sigma^2 . \tag{3.3.5}$$

The result of the theorem follows from (3.3.4) and (3.3.5). []

<u>Lemma 3.3.2</u> If $\tilde{\lambda}_n \xrightarrow{P} \lambda_*$, then $\dfrac{\partial^2 \bar{h}_n(\theta_0, \tilde{\lambda}_n)}{\partial \lambda^2}$ is bounded in probability under both the null and alternative models.

<u>Lemma 3.3.3</u> Under both the null and alternative models

$$\frac{\partial \bar{h}_n(\theta_0, \lambda_*)}{\partial \lambda} \xrightarrow{P} 0 .$$

The proofs of these lemmas are given in Appendix C.

<u>Theorem 3.3.2</u> Under both H_0 and H_1 the test statistics $T_n(\hat{\lambda}_n)$ and $T_n(\lambda_*)$ are asymptotically equivalent in the sense that

$$T_n(\hat{\lambda}_n) - T_n(\lambda_*) \xrightarrow{P} 0 .$$

<u>Proof:</u>

Recall that

$$T_n(\hat{\lambda}_n) = \sqrt{n}\, \bar{h}_n(\theta_0, \hat{\lambda}_n) / \hat{\sigma}_n(\theta_0, \hat{\lambda}_n)$$

$$T_n(\lambda_*) = \sqrt{n}\,\bar{h}_n(\theta_0,\lambda_*)/\ \hat{\sigma}_n(\theta_0,\lambda_*)\ .$$

In Section 3.2 we showed that $\hat{\sigma}_n(\theta_0,\lambda_*)$ is a consistent estimator of σ under both hypotheses and a similar result for $\hat{\sigma}_n(\theta_0,\hat{\lambda}_n)$ is obtained in Theorem 3.3.1 above. Hence by Theorem A.2 of Appendix A the result follows if we show that

$$\sqrt{n}\,\bar{h}_n(\theta_0,\hat{\lambda}_n) - \sqrt{n}\,\bar{h}_n(\theta_0,\lambda_*) \xrightarrow{p} 0$$

under both models.

Consider a Taylor expansion of $\sqrt{n}\,\bar{h}_n(\theta_0,\hat{\lambda}_n)$ about $\hat{\lambda}_n = \lambda_*$

$$\sqrt{n}\,\bar{h}_n(\theta_0,\hat{\lambda}_n) = \sqrt{n}\,\bar{h}_n(\theta_0,\lambda_*) + \sqrt{n}\,(\hat{\lambda}_n - \lambda_*)\,\frac{\partial\bar{h}_n(\theta_0,\lambda_*)}{\partial\lambda} +$$

$$\sqrt{n}\,(\hat{\lambda}_n - \lambda_*)^2\,\frac{\partial^2\bar{h}_n(\theta_0,\tilde{\lambda}_n)}{\partial\lambda^2} \tag{3.3.6}$$

where $\tilde{\lambda}_n$ is such that $|\tilde{\lambda}_n - \lambda_*| \le |\hat{\lambda}_n - \lambda_*|$. From Appendix E $\hat{\lambda}_n \xrightarrow{p} \lambda_*$ and $\sqrt{n}\,(\hat{\lambda}_n - \lambda_*) \xrightarrow{p} N(0,I(\lambda_*))$, where $I(\lambda_*)$ denotes the information of λ . Hence, by the above two lemmas

$$\sqrt{n}\,(\hat{\lambda}_n-\lambda_*)\,\frac{\partial\bar{h}_n(\theta_0,\lambda_*)}{\partial\lambda} + \sqrt{n}\,(\hat{\lambda}_n-\lambda_*)^2\,\frac{\partial^2\bar{h}_n(\theta_0,\tilde{\lambda}_n)}{\partial\lambda^2} \xrightarrow{p} 0\ .$$

Therefore

$$\sqrt{n}\,\bar{h}_n(\theta_0,\hat{\lambda}_n) - \sqrt{n}\,\bar{h}_n(\theta_0,\lambda_*) \xrightarrow{p} 0\ . \qquad []$$

<u>Theorem 3.3.3</u> If $\hat{\lambda}_n$ is the MLE of λ then $T_n(\hat{\lambda}_n)$ is Pitman regular and $ARE(T_n(\hat{\lambda}_n),T_n(\lambda_*)) = 1$.

<u>Proof:</u>

Let $\Psi_n(\theta,\lambda_*)$ and $\tau_n(\theta,\lambda_*)$ be as defined in (3.2.33) and (3.2.34) respectively. Since by Theorem 3.2.3 $T_n(\lambda_*)$ is Pitman regular then, the result of the present theorem follows from Theorem 3.3.2 above and Theorem A.3 in Appendix A. []

3.4 Examples

In this section we use the derived formula of Pitman efficiency given by (3.2.44) and Theorem 3.3.3 to evaluate the asymptotic relative efficiency of $T_n(1)$ relative to $T_n(\hat{\lambda}_n)$ for the transformed-normal, transformed-contaminated-normal and transformed-Student's t models. Recall that in (3.1.2) we assume that $\mathrm{var}(\epsilon) = 1$. Except for the normal model we need to rescale ϵ so that this assumption is met. In all models we vary λ over the set $\{1/4, 1/3, 1/2, 1\}$. For the transformed normal model σ varies over the set $\{1/2, 1, 2, 3, 4, 5\}$. For the rest of the models σ assumes the above values multiplied by the factor required to make $\mathrm{var}(\epsilon) = 1$. We use numerical integration methods (Simpson rule) executed on GAUSS software to evaluate the two expectations involved in (3.2.44). Appendix E contains the program used under each model.

3.4.1 Transformed-normal model

The p.d.f. of ϵ is given as

$$f(\epsilon) = 1/\sqrt{2\pi}\, e^{-\epsilon^2/2} \quad \text{and} \quad \mathrm{var}(\epsilon) = 1 .$$

Table 3.1 below gives the results of evaluating $\mathrm{ARE}(T_n(1), T_n(\hat{\lambda}_n))$ for the above proposed values of λ and σ

Table 3.1 ARE($T_n(1), T_n(\hat{\lambda}_n)$) of the transformed-normal model

$\sigma \setminus \lambda$	1/4	1/3	1/2	1
.5	0.9092	0.9329	0.9667	1
1	0.7294	0.8044	0.9075	1
2	0.4043	0.5584	0.7906	1
3	0.2172	0.3884	0.6969	1
4	0.1214	0.2790	0.6244	1
5	0.0716	0.2076	0.5675	1

From the above table note that:

1. There is a considerable gain in efficiency when $T_n(\hat{\lambda}_n)$ instead of $T_n(1)$ is used.

2. The gain increases as λ decreases and/or σ increases.

3. The differences among the entries in the first row, corresponding to $\sigma = .5$, are not as much as those in the other rows. This gives an indication that when σ is small, $h(y,\lambda)$ is approximately linear in y. So, no matter what transformation is applied we get results that are close together.

3.4.2 Transformed-contaminated-normal model

Suppose that X has a contaminated normal distribution with contamination variance η^2, that is,

$$X = (1-B)\ X_1 + B\ X_2$$

where $X_1 \sim N(0,1)$, $X_2 \sim N(0,\eta^2)$ and $B \sim Bin(1,p)$. It follows that the p.d.f. of X is given as

$$f(x) = (1-p)\ \varphi(x_1) + p\ \varphi(x_2/\eta)/\eta$$

where φ is the p.d.f. of $N(0,1)$. It can be shown that

$$\mathrm{var}(X) = (1-p) + p\ \eta^2 .$$

Let $\epsilon = X/\sigma_X$, where σ_X refers to the standard deviation of X . Then $\mathrm{var}(\epsilon) = 1$. Tables 4.2 through 4.5 below give the asymptotic relative efficiency $ARE(T_n(1), T_n(\hat{\lambda}_n))$ for the different combinations of $p = .1$ and $.2$, and $\eta^2 = 16$ and 25 .

Table 3.2 $ARE(T_n(1), T_n(\hat{\lambda}_n))$ of the transformed-contaminated-normal model $p = .1$, $\eta^2 = 16$

$\sigma \diagdown \lambda$	1/4	1/3	1/2	1
.5	0.2236	0.3364	0.5550	1
1	0.0464	0.1248	0.3651	1
2	0.0050	0.0319	0.2132	1
3	0.0011	0.0130	0.1514	1
4	0.0004	0.0068	0.1186	1
5	0.0002	0.0041	0.0984	1

Table 3.3 $ARE(T_n(1), T_n(\hat{\lambda}_n))$ of the transformed-contaminated-normal model $p = .2$, $\eta^2 = 16$

$\sigma \backslash \lambda$	1/4	1/3	1/2	1
.5	0.2042	0.3114	0.5273	1
1	0.0421	0.1138	0.3411	1
2	0.0046	0.0292	0.1976	1
3	0.0010	0.0120	0.1403	1
4	0.0003	0.0063	0.1010	1
5	0.0001	0.0038	0.0913	1

Table 3.4 $ARE(T_n(1), T_n(\hat{\lambda}_n))$ of the transformed-contaminated-normal model $p = .1$, $\eta^2 = 25$

$\sigma \backslash \lambda$	1/4	1/3	1/2	1
.5	0.1182	0.2128	0.4335	1
1	0.0177	0.0636	0.2573	1
2	0.0015	0.0139	0.1382	1
3	0.0003	0.0053	0.0948	1
4	0.0001	0.0027	0.0728	1
5	0.0001	0.0016	0.0597	1

Table 3.5 ARE($T_n(1), T_n(\hat{\lambda}_n)$) of the transformed-contaminated-normal model p = .2 , η^2 = 25

$\sigma \backslash \lambda$	1/4	1/3	1/2	1
.5	0.1136	0.2053	0.4221	1
1	0.0173	0.0618	0.2500	1
2	0.0015	0.0137	0.1349	1
3	0.0003	0.0053	0.0929	1
4	0.0001	0.0027	0.6244	1
5	0.0001	0.0016	0.0589	1

From the above tables we note that

1. The entries in all tables show that $T_n(\hat{\lambda}_n)$ is asymptotically much more efficient than $T_n(1)$, and as above the efficiency increases as σ increases and/or λ decreases.

2. Comparisons of Table 3.2 with Table 3.4 and Table 3.3 with Table 3.5 show that there is not much difference in efficiency between the two proportions of contamination.

3. Comparisons of Tables 3.2 and 3.3 with 3.4 and 3.5 show that the gain in efficiency increases as the contamination variance increases.

3.4.3 Transformed Student's-t model

Suppose that X has a Student's-t distribution with ν degrees of freedom. The p.d.f. of X is then given by

$$f(x) = 1/\sqrt{\nu\pi}\ \frac{\Gamma(\nu+1)/2}{\Gamma(\nu/2)}\ (1 + t^2/\nu)^{-(\nu+1)/2}$$

and $\operatorname{var}(X) = \nu/(\nu-2)$, $\nu > 2$. Let $\epsilon = X \sqrt{(\nu-2)/\nu}$ then, $\operatorname{var}(\epsilon) = 1$. Tables 3.6, 3.7 and 3.8 below contain the asymptotic relative efficiency for the t-model with ν equals 10 , 20 and 30 respectively.

Table 3.6 $\mathrm{ARE}(T_n(1), T_n(\hat{\lambda}_n))$ of the transformed-Student's t model $\nu = 10$

$\sigma \backslash \lambda$	1/4	1/3	1/2	1
.5	0.7828	0.8338	0.9067	1
1	0.4991	0.6286	0.8056	1
2	0.1661	0.3436	0.6490	1
3	0.0591	0.2009	0.5440	1
4	0.0246	0.1277	0.4707	1
5	0.0117	0.0869	0.4172	1

Table 3.7 $ARE(T_n(1), T_n(\hat{\lambda}_n))$ of the transformed-Student's t model $\nu = 20$

$\sigma \backslash \lambda$	1/4	1/3	1/2	1
.5	0.8594	0.8931	0.9422	1
1	0.6367	0.7318	0.8647	1
2	0.2983	0.4637	0.7290	1
3	0.1396	0.3014	0.6289	1
4	0.0704	0.2061	0.5550	1
5	0.0385	0.1479	0.4988	1

Table 3.8 $ARE(T_n(1), T_n(\hat{\lambda}_n))$ of the transformed-Student's t model $\nu = 30$

$\sigma \backslash \lambda$	1/4	1/3	1/2	1
.5	0.8782	0.9080	0.9513	1
1	0.6710	0.7586	0.8804	1
2	0.3358	0.4975	0.7512	1
3	0.1658	0.3317	0.6531	1
4	0.0870	0.2311	0.5796	1
5	0.0490	0.1680	0.5230	1

From the above tables we note that the same directions for the previous two models continue to hold for the t-model. Further it may be noted that the ARE is smaller for smaller ν and vice versa.

4. ASYMPTOTIC RESULTS FOR TEST STATISTICS UNDER SKEWED MODELS

In Chapter 3, the performance of the transformed t-test relative to the untransformed t-test was investigated under the assumption that the original observations have a heavy-tailed symmetric distribution. In that chapter we considered the John-Draper family of transformations which symmetrically deals with both tails and hence we get a symmetric distribution for the transformed variable.

In the present chapter we drop the symmetry assumption and assume that when the two-domain transformation is considered, there exists some value of the transformation parameters that bring the transformed variable into normality. As in Chapter 3, the main purpose of this chapter is to evaluate the Pitman asymptotic relative efficiency of the t-test in the original scale relative to the transformed t-test using the maximum likelihood estimators of the transformation parameters. We use the transformed t-test assuming the true transformation is known as the basic model to which we relate the asymptotic results of both the untransformed test and the transformed test using the MLE's of the transformation parameters. In Section 4.1 we introduce some notation that will be frequently used throughout the chapter. Section 4.2 contains proofs of the different asymptotic results required for the derivation of the Pitman efficiency of the untransformed t-test relative to the t-test after applying a known transformation. In Section 4.3 we derive the asymptotic results for the transformed t-test using the maximum likelihood estimators of the transformation parameters and the asymptotic relative efficiency of this test relative to the t-test with a known transformation. The chapter is concluded in Section 4.4 with some numerical evaluations of the asymptotic relative efficiency under the transformed normal model for different combinations of the transformation parameters.

4.1 Definitions and Notation

Throughout this chapter $h(Y-\theta,\underline{\lambda})$ is used to denote the two-domain transformation introduced in Section 2.2 for $\underline{\lambda} \neq 0$ as

$$h(Y-\theta,\underline{\lambda}) = \begin{cases} (1 - (\theta-Y+1)^{\lambda_1})/\lambda_1 & Y \leq \theta \\ ((Y-\theta+1)^{\lambda_2} - 1)/\lambda_2 & Y > \theta \end{cases}$$

where $\underline{\lambda}$ denotes the 2x1 vector of transformation parameters $(\lambda_1,\lambda_2)^t$. We always assume that both λ_1 and λ_2 are positive. As in Chapter 3, θ is used to denote the median of the distribution of Y.

The main model we assume in this chapter is

$$h(Y-\theta,\underline{\lambda}_*) = \sigma\epsilon \tag{4.1.1}$$

where, for some λ_*, θ and σ, ϵ is assumed to have a standard normal distribution. The asymptotic results are developed for the problem of testing

$$H_0\colon \theta = \theta_0 \qquad \text{versus} \qquad H_1\colon \theta = \theta_n > \theta_0$$

where

$$\theta_n = \theta_0 + k_1/\sqrt{n} \tag{4.1.2}$$

for some positive number k_1.

The test statistics we consider in this chapter are the same as those introduced in Section 3.1 with $h(Y-\theta,\underline{\lambda})$ being the two-domain family instead of the John-Draper family. Let $f_Y(y,\theta,\underline{\lambda}_*,\sigma)$ denote the pdf of Y,

$$f_Y(y,\theta,\underline{\lambda}_*,\sigma) = (1/\sigma)\ f_\epsilon(h(y-\theta,\underline{\lambda})/\sigma)\ \frac{\partial h(y-\theta,\underline{\lambda}_*)}{\partial y}\ , \qquad (4.1.3)$$

where $f_\epsilon(Z)$ is the pdf of the standard normal distributio.

Let $\omega = (\underline{\lambda},\sigma)$ and let $U(\epsilon,\omega)$ denote the 3x1 score vector with k-th component given by

$$U_k(\epsilon,\omega) = \frac{\partial(\ln f_Y(y,\theta,\underline{\lambda},\sigma))}{\partial\omega_k}\ .$$

Let $I(\omega)$ denote the 3x3 information matrix with (j,k) entry given by

$$I_{jk}(\omega) = E\{-\frac{\partial^2(\ln f_Y(y,\theta,\underline{\lambda},\sigma))}{\partial\omega_j\ \partial\omega_k}\}\ .$$

Let $I^{\lambda\lambda}(\omega)$ denote the upper left 2x2 block of $I^{-1}(\omega)$. Let $U_*(\epsilon)$ denote $U(\epsilon,\omega_*)$ and let I^* denote $I^{\lambda\lambda}(\omega_*)$ where $\omega_* = (\underline{\lambda}_*,\sigma)$.

4.2 Asymptotics When $\underline{\lambda}$ is Known

In this section we derive the asymptotic properties of the transformed t-test $T_n(\underline{\lambda}_*)$ and the t-test in the original scale $T_n(1)$ under both the null and alternative models and show that both of them are Pitman regular. Also, we give an expression for $ARE(T_n(1),T_n(\underline{\lambda}_*))$.

4.1.1 Asymptotic distribution of $T_n(\underline{\lambda}_*)$ under H_0

Assume that under H_0 and for some $\underline{\lambda}_*$ and σ

$$h(Y-\theta_0,\underline{\lambda}_*) = \sigma\epsilon \qquad (4.2.1)$$

where ϵ has a standard normal distribution. Then

$$T_n(\underline{\lambda}_*) = \sqrt{n}\ \bar{h}_n(\theta_0,\underline{\lambda}_*)/\ \hat{\sigma}_n(\theta_0,\underline{\lambda}_*) = \sqrt{n}\ \bar{\epsilon}_n/\hat{\sigma}_\epsilon\ .$$

By the CLT and the fact that $\hat{\sigma}_\epsilon/\sigma \longrightarrow 1$, it follows that $\sqrt{n}\ \bar{\epsilon}_n/\hat{\sigma}_\epsilon \xrightarrow{\mathcal{L}} N(0,1)$. Therefore under H_0

$$T_n(\underline{\lambda}_*) \xrightarrow{\mathcal{L}} N(0,1)\ . \qquad (4.2.2)$$

4.2.2 Asymptotic distribution of $T(\underline{\lambda}_*)$ under contiguous alternatives

Given $k_1 > 0$, under the alternative model, we assume that

$$h(Y-\theta_n, \underline{\lambda}_*) = \sigma\,\epsilon\ . \qquad (4.2.3)$$

The first step in deriving the asymptotic distribution of $T_n(\underline{\lambda}_*)$ under the alternative model is to express $h(Y-\theta_0, \underline{\lambda}_*)$ in terms of $h(Y-\theta_n, \underline{\lambda}_*)$. For $Y \le \theta_0$,

$$h(Y-\theta_0, \underline{\lambda}_*) = [1 - (\theta_0 - Y+1)^{\lambda_{1*}}]/\ \lambda_{1*} = [1 - (\theta_n - Y + 1 - k_1/\sqrt{n}\ )^{\lambda_{1*}}]/\lambda_{1*}$$

$$= [1-(\theta_n - Y+1)^{\lambda_{1*}}\left(1 - \frac{k_1/\sqrt{n}}{(\theta_n - Y+1)}\right)^{\lambda_{1*}}]/\lambda_{1*}\ .$$

Since we are considering asymptotic results, we can assume $k_1/\sqrt{n} < 1$. Since $Y-\theta_0 \le 0$ implies $Y-\theta_n \le 0$ hence, a Maclaurin expansion of $\left(1 - \frac{k_1/\sqrt{n}}{(\theta_n - Y+1)}\right)^{\lambda_{1*}}$ is absolutely convergent and

$$h(Y-\theta_0, \underline{\lambda}_*) = h(Y-\theta_n, \underline{\lambda}_*) + (k_1/\sqrt{n}\ )(\theta_n - Y+1)^{\lambda_{1*}-1}$$

$$- (1/\lambda_{1*}) \sum_{j=2}^{\infty} ((-k_1/\sqrt{n}\ )^{j}/j!)(\prod_{m=0}^{j-1}(\lambda_{1*}-m))(\theta_n - Y_i+1)^{\lambda_{1*}-j}\ .$$

Similarly for $Y > \theta_n$ it can be shown that,

$$h(Y-\theta_0, \underline{\lambda}_*) = h(Y-\theta_n, \underline{\lambda}_*) + (k_1/\sqrt{n}\ )(Y-\theta_n+1)^{\lambda_{2*}-1}$$

$$+ (1/\lambda_{2*}) \sum_{j=2}^{\infty} ((k_1/\sqrt{n}\ )^{j}/j!)(\prod_{m=0}^{j-1}(\lambda_{2*}-m))(Y-\theta_n+1)^{\lambda_{2*}-j}\ .$$

Let A_n and B_n be defined as in (3.2.10) and (3.2.17) respectively. Under model (4.2.3) and for all $Y_i \in A_n^c$ we can write

$$h(Y-\theta_0, \underline{\lambda}_*) = \sigma\epsilon + R_3(\epsilon, \underline{\lambda}_*, n) \qquad (4.2.4)$$

where

$$R_3(\epsilon,\underline{\lambda}_*,n) = -(1/\lambda_{1*})\sum_{j=1}^{\infty}\{((-k_1/\sqrt{n})^j/j!)(\prod_{m=0}^{j-1}(\lambda_{1*}-m))(1-\lambda_{1*}\sigma\epsilon)^{1-j/\lambda_{1*}}$$

$$I_{(-\infty,0)}(\epsilon)\} + (1/\lambda_{2*})\sum_{j=1}^{\infty}\{((k_1/\sqrt{n})^j/j!)(\prod_{m=0}^{j-1}(\lambda_{2*}-m)(1+\lambda_{2*}\sigma\epsilon)^{1-j/\lambda_{2*}}I_{(0,\infty)}(\epsilon)\}$$

(4.2.5)

or we can write

$$h(Y-\theta_0,\underline{\lambda}_*) = \sigma\epsilon + (k_1/\sqrt{n})(g(\epsilon,\underline{\lambda}_*)) + R_4(\epsilon,\underline{\lambda}_*,n) \qquad (4.2.6)$$

where

$$g(\epsilon,\underline{\lambda}_*) = (1-\lambda_{1*}\sigma\epsilon)^{1-1/\lambda_{1*}}I_{(-\infty,0)}(\epsilon) + (1+\lambda_{2*}\sigma\epsilon)^{1-1/\lambda_{2*}}I_{(0,\infty)}(\epsilon)$$

(4.2.7)

and

$$R_4(\epsilon,\underline{\lambda}_*,n) = -\sum_{j=2}^{\infty}\{((-k_1/\sqrt{n})^j/j!)(\prod_{m=1}^{j-1}(\lambda_{1*}-m))(1-\lambda_{1*}\sigma\epsilon)^{1-j/\lambda_{1*}}$$

$$I_{(-\infty,0)}(\epsilon)\} + \sum_{j=2}^{\infty}\{((k_1/\sqrt{n})^j/j!)(\prod_{m=1}^{j-1}(\lambda_{2*}-m)(1+\lambda_{2*}\sigma\epsilon)^{1-j/\lambda_{2*}}I_{(0,\infty)}(\epsilon)\}.$$

(4.2.8)

In the following two theorems we show that under the alternative model $\hat{\sigma}_n(\theta_0,\underline{\lambda}_*)$ is a consistent estimator of σ and $T_n(\underline{\lambda}_*)$ tends to a normal distribution.

Theorem 4.2.1 Letting $\hat{\sigma}_n^2(\theta_0,\underline{\lambda}_*)$ be defined as in (4.1.3), then under the alternative model given by (4.2.3) $\hat{\sigma}_n(\theta_0,\underline{\lambda}_*)$ is a consistent estimator of σ.

Proof From (4.2.4) write

$$h^2(Y-\theta_0,\underline{\lambda}_*) = \sigma^2\epsilon^2 + 2\sigma\epsilon\, R_3(\epsilon,\underline{\lambda}_*,n) + R_3^2(\epsilon,\underline{\lambda}_*,n)$$

hence,

$$(1/n) \sum_{Y_i \in A_n^c} h^2(Y_i-\theta_0,\underline{\lambda}_*) - (\sigma^2/n) \sum_{\epsilon_i \in B_n^c} \epsilon_i^2 =$$

$$(2\sigma/n) \sum_{\epsilon_i \in B_n^c} \epsilon_i R_3(\epsilon_i,\underline{\lambda}_*) + (1/n) \sum_{\epsilon_i \in B_n^c} R_3^2(\epsilon_i,\underline{\lambda}_*,n) \ . \qquad (4.2.9)$$

From (4.2.5) note that the cross product term in $R_3^2(\epsilon_i,\underline{\lambda}_*,n)$ is 0 since $I_{(-\infty,0)}(\epsilon_i)\, I_{(0,\infty)}(\epsilon_i) = 0$ for all ϵ_i. Also, from (4.2.5) note that each component of $R_3(\epsilon_i,\underline{\lambda}_*,n)$ has the same form as that of $R_1(\epsilon_i,\lambda_*,n)$ given by (3.2.14). Hence from Lemma 3.2.2(i) we get

$$(1/n) \sum_{\epsilon_i \in B_n^c} \epsilon_i R_3(\epsilon_i,\underline{\lambda}_*,n) \xrightarrow{p} 0$$

and from Lemma 3.2.2(ii) we get

$$(1/n) \sum_{\epsilon_i \in B_n^c} R_3^2(\epsilon_i,\underline{\lambda}_*,n) \xrightarrow{p} 0 \ .$$

Therefore it follows from (4.2.9) that

$$(1/n) \sum_{Y_i \in A_n^c} h^2(Y_i-\theta_0,\underline{\lambda}_*) - (\sigma^2/n) \sum_{\epsilon_i \in B_n^c} \epsilon_i^2 \xrightarrow{p} 0 \ . \qquad (4.2.10)$$

By the WLLN $(1/n) \sum_{i=1}^{n} \epsilon_i^2/n \xrightarrow{p} 1$ and by Lemma 3.2.1(v),

$$(1/n) \sum_{i=1}^{n} \epsilon_i^2/n - (1/n) \sum_{\epsilon_i \in B_n^c} \epsilon_i^2 \xrightarrow{p} 0$$

hence,

$$(\sigma^2/n) \sum_{\epsilon_i \in B_n^c} \epsilon_i^2 \xrightarrow{p} \sigma^2 \qquad (4.2.11)$$

From a similar argument like that of Lemma 3.2.1(iii) we can show

$$(1/n) \sum_{i=1}^{n} h^2(Y_i-\theta_0,\underline{\lambda}_*) - (1/n) \sum_{Y_i \in A_n^c} h^2(Y_i-\theta_0,\underline{\lambda}_*) \xrightarrow{p} 0 \ .$$

$$(4.2.12)$$

Hence by (4.2.10) , (4.2.11) and (4.2.12) we get

$$(1/n) \sum_{i=1}^{n} h^2(Y_i-\theta_0,\underline{\lambda}_*) \xrightarrow{p} \sigma^2 \ .$$

Also, it can be shown that, $\bar{h}_n^2(\theta_0,\underline{\lambda}_*) \xrightarrow{p} 0$. Therefore under contiguous alternatives $\hat{\sigma}_n^2(\theta_0,\underline{\lambda}_*) \xrightarrow{P} \sigma^2$. By Fact 4 in Appendix A $\hat{\sigma}_n(\theta_0,\underline{\lambda}_*) \xrightarrow{P} \sigma$. []

<u>Theorem 4.2.2</u> Let $\Psi = (k_1/\sigma)\, E\{g(\epsilon,\underline{\lambda}_*)\}$ where $g(\epsilon,\underline{\lambda}_*)$ is defined in (4.2.7). Under the alternative model given by (4.2.3)

$$T_n(\underline{\lambda}_*) - \Psi \xrightarrow{\mathcal{L}} N(0,1) \ .$$

<u>Proof</u> As argued in the proof of Theorem 3.2.2 it suffices to show $\bar{T}_n(\underline{\lambda}_*) - \Psi \xrightarrow{\mathcal{L}} N(0,1)$.

$$\sqrt{n} \sum_{Y_i \in A_n^c} h(Y_i-\theta_0,\underline{\lambda}_*)/n = \sigma\sqrt{n}\Big(\sum_{\epsilon_i \in B_n^c} \epsilon_i/n\Big) + k_1 \sum_{\epsilon_i \in B_n^c} g(\epsilon_i,\underline{\lambda}_*)/n$$
$$+ (1/\sqrt{n}) \sum_{\epsilon_i \in B_n^c} R_4(\epsilon_i,\lambda_*,n) \ . \qquad (4.2.13)$$

Since each component of $R_4(\epsilon_i,\underline{\lambda}_*,n)$ as defined in (4.2.8) has the same form as $R_2(\epsilon_i,\underline{\lambda}_*,n)$ given by (3.2.16), hence by Lemma 3.2.2(iii) $(1/\sqrt{n}) \sum_{\epsilon_i \in B_n^c} R_4(\epsilon_i,\underline{\lambda}_*,n) \xrightarrow{P} 0$. From (4.2.13)

$$\sqrt{n} \sum_{Y_i \in A_n^c} h(Y_i-\theta_0,\underline{\lambda}_*)/n - \sigma\sqrt{n} \sum_{\epsilon_i \in B_n^c} \epsilon_i/n - k_1 \sum_{\epsilon_i \in B_n^c} g(\epsilon_i,\underline{\lambda}_*)/n$$
$$\xrightarrow{P} 0 \qquad (4.2.14)$$

In the proofs of parts (ii), (iv) and (vi) of Lemma 3.2.1 replace λ_* by λ_{2*} to conclude

$$\sqrt{n}\, \bar{h}_n(\theta_0,\underline{\lambda}_*) - \sqrt{n} \sum_{Y_i \in A_n^c} h(Y_i-\theta_0,\underline{\lambda}_*)/n \xrightarrow{P} 0 \ , \qquad (4.2.15)$$

$$\sqrt{n} \sum_{\epsilon_i \in B_n^c} \epsilon_i/n \xrightarrow{\mathcal{L}} N(0,1) \ , \qquad (4.2.16)$$

and

$$k_1 \sum_{\epsilon_i \in B_n^c} g(\epsilon_i,\underline{\lambda}_*)/n \xrightarrow{P} k_1\, E\{g(\epsilon_i,\underline{\lambda}_*)\} \ . \qquad (4.2.17)$$

(4.2.14) through (4.2.17) imply,

$$T_n(\underline{\lambda}_*) - \Psi = \frac{\sqrt{n}\, \bar{h}_n(\theta_0,\underline{\lambda}_*) - k_1\, E\{g(\epsilon_i,\underline{\lambda}_*)\}}{\sigma} \xrightarrow{\mathcal{L}} N(0,1) \ . \quad []$$

4.2.3 The test statistic $T_n(\underline{\lambda}_*)$ is Pitman regular

We first derive candidates for $\Psi_n(\theta)$ and $\tau_n^2(\theta)$ in Appendix B. Consider a Taylor expansion of $\bar{T}_n(\underline{\lambda}_*)$ about $\theta_0 = \theta$, where θ is in a neighborhood of θ_0.

$$(\sqrt{n}/\sigma)\ \bar{h}_n(\theta_0,\underline{\lambda}_*) \approx (\sqrt{n}/\sigma)\ \bar{h}_n(\theta,\underline{\lambda}_*) + (\sqrt{n}/\sigma)(\theta_0-\theta)\frac{\partial\bar{h}_n(\theta,\underline{\lambda}_*)}{\partial\theta}$$

$$= (1/\sigma)\sqrt{n}\ \bar{h}_n(\theta,\underline{\lambda}_*) - (\sqrt{n}/\sigma)(\theta-\theta_0)\sum_{i=1}^{n}[-(\theta-Y_i+1)^{\lambda_{1*}-1} I_{(-\infty,\theta)}(Y_i)$$

$$- (Y_i-\theta+1)^{\lambda_{2*}-1} I_{(\theta,\infty)}(Y_i)]/n$$

Under the model $h(Y_i-\theta,\underline{\lambda}_*) = \sigma\epsilon_i$

$$\bar{T}_n(\underline{\lambda}_*) \approx \sqrt{n}\ \bar{\epsilon}_n + (\sqrt{n}/\sigma)(\theta-\theta_0)(1/n\sum_{i=1}^{n} g(\epsilon_i,\underline{\lambda}_*)) .$$

Define

$$\Psi_n(\theta) = (\sqrt{n}/\sigma)(\theta-\theta_0)\ E\{g(\epsilon,\underline{\lambda}_*)\} \tag{4.2.18}$$

where $g(\epsilon,\underline{\lambda}_*)$ is given by (4.2.7), and define

$$\tau_n^2(\theta) = 1 . \tag{4.2.19}$$

<u>Theorem 4.2.3</u> Let $\Psi_n(\theta,\underline{\lambda}_*)$ and $\tau_n^2(\theta,\underline{\lambda}_*)$ be as defined in (4.2.18) and (4.2.19) respectively then, $T_n(\underline{\lambda}_*)$ is Pitman regular and

$$R_n^2(\theta_0) = (n/\sigma^2)\ \{E[g(\epsilon,\underline{\lambda}_*)]\}^2 .$$

<u>Proof</u> To prove the theorem we verify the seven regularity conditions of Appendix B.

C1. Note that $\Psi_n(\theta_0) = 0$ and $T_n(\underline{\lambda}_*) \xrightarrow{\mathcal{L}} N(0,1)$. Therefore, from (4.2.2)

$$\frac{T_n(\underline{\lambda}_*) - \Psi_n(\theta_0)}{\tau_n(\theta_0)} \xrightarrow{\mathcal{L}} N(0,1) .$$

C2. Under H_1, $\theta = \theta_n$ and (4.1.18) becomes

$\Psi_n(\theta_n) = k_1\, E\{g(\epsilon,\underline{\lambda}_*)\}/\sigma = \Psi$, where Ψ is as in Theorem 4.2.2. Hence the result of this theorem verifies the asymptotic normality under contiguous alternatives.

C3. From (4.2.18) $\Psi_n(\theta)$ is differentiable for all θ .

C4. $\Psi_n'(\theta) = (\sqrt{n}/\sigma)\, E\{g(\epsilon,\underline{\lambda}_*)\} = \sqrt{n}\, \Psi/k_1$. From (4.2.7),

$$g(\epsilon,\underline{\lambda}_*) = (1-\lambda_{1*}\sigma\epsilon)^{1-1/\lambda_{1*}} I_{(-\infty,0)}(\epsilon) + (1+\lambda_{2*}\sigma\epsilon)^{1-1/\lambda_{2*}} I_{(0,\infty)}(\epsilon)$$

Which is positive every where. Hence $E\{g(\epsilon,\underline{\lambda}_*)\} > 0$ and $\Psi_n'(\theta) > 0$.

C5. $\Psi_n'(\theta_0)/\sqrt{n} = \Psi/k_1 > 0$.

C6. $\Psi_n'(\theta)$ is the same for all θ hence,

$$\sup_{\theta_0 \le \theta_n^* \le \theta_n} |\Psi_n'(\theta_n^*)/\Psi_n'(\theta_0) - 1| = 0 \quad \text{for all } n .$$

C7. $\tau_n(\theta_n)/\tau_n(\theta_0) = 1$ for all n .

Therefore $T_n(\underline{\lambda}_*)$ is Pitman regular and from Appendix B

$$R_n^2(\theta_0) = n\, \{E[g(\epsilon,\underline{\lambda}_*)]\}^2/\sigma^2 . \quad [] \tag{4.2.20}$$

4.2.4 Asymptotics of the t-test in the original scale

Recall that the t-test in the original scale is given by $T_n(1) = \sqrt{n}\,(\bar{Y}-\theta_0)/\hat{\sigma}_Y$. Under the model $h(Y-\theta,\underline{\lambda}_*) = \sigma\epsilon$ we can express Y as

$$Y = \begin{cases} \theta - [\ (1-\lambda_{1*}\sigma\epsilon)^{1/\lambda_{1*}} - 1\ ] & \epsilon \le 0 \\ \theta + [\ (1+\lambda_{2*}\sigma\epsilon)^{1/\lambda_{2*}} - 1\ ] & \epsilon > 0 \end{cases} .$$

This can be written as

$$Y-\theta = (1-\lambda_{1*}\sigma\epsilon)^{1/\lambda_{1*}} I_{(-\infty,0)}(\epsilon) + (1+\lambda_{2*}\sigma\epsilon)^{1/\lambda_{2*}} I_{(0,\infty)}(\epsilon) . \tag{4.2.21}$$

Define $\xi(\underline{\lambda}_*,\sigma) = E_\theta(Y-\theta)$.

<u>Theorem 4.2.4</u> Let $\Psi_n(\theta) = (\sqrt{n}/\sigma_Y)[(\theta-\theta_0) + \xi(\underline{\lambda}_*,\sigma)]$ and let

$\tau_n^2(\theta) = 1$. Then, the test statistic $\tilde{T}_n(1)$ is Pitman regular with,

$$R_n^2(\theta_0) = n/\sigma_Y^2 .$$

Proof The test statistic is Pitman regular if it satisfies the seven regularity conditions of Appendix B .

C1. Under H_0 $\Psi_n(\theta_0) = \sqrt{n}/\sigma_Y \, \xi(\underline{\lambda}_*,\sigma)$. hence

$$\tilde{T}_n(1) - \Psi_n(\theta_0) = \frac{\sqrt{n}\,(\bar{Y} - \theta_0)}{\sigma_Y} - \frac{\sqrt{n}\,\xi(\underline{\lambda}_*,\sigma)}{\sigma_Y}$$

$$= \sqrt{n}\,\frac{[\bar{Y} - (\theta_0+\xi(\underline{\lambda}_*,\sigma))]}{\sigma_Y} = \sqrt{n}\,\frac{[\bar{Y} - E_{\theta_0}(Y)]}{\sigma_Y} .$$

Hence by the CLT

$$\tilde{T}_n(1) - \Psi_n(\theta_0) \xrightarrow{\mathscr{L}} N(0,1) .$$

C2. replace θ_0 by θ_n in the above argument for C1 .

C3. $\Psi_n(\theta)$ is differentiable for all θ .

C4. $\Psi_n'(\theta) = \sqrt{n}\,/\sigma_Y > 0$.

C5. $\Psi_n'(\theta_0)/\sqrt{n} = 1/\sigma_Y > 0$.

C6. from C4 $\Psi_n'(\theta)$ is the same for all θ hence,

$$\sup_{\theta_0 \le \tilde{\theta}_n \le \theta_n} |\Psi_n'(\tilde{\theta}_n)/\Psi_n'(\theta_0) - 1| = 1 \quad \text{for all } n .$$

C7. $\tau_n^2(\theta) = 1$ for all n and θ .

Hence $T_n(1)$ is Pitman regular and from C4 and C7 ,

$$R_n^2(\theta_0) = n/\ \sigma_Y^2 . \quad [] \tag{4.2.22}$$

From (4.2.20) and (4.2.21) we get the Pitman asymptotic relative efficiency of the untransformed t-test relative to the transformed t-test with a known transformation as,

$$\mathrm{ARE}(\tilde{T}_n(1),T_n(\underline{\lambda}_*)) = \frac{\sigma^2/\ E^2\{g(\epsilon,\underline{\lambda}_*)\}}{\sigma_Y^2} \tag{4.2.23}$$

where $g(\epsilon,\underline{\lambda}_*)$ and σ_Y^2 follow from (4.2.7) and (4.2.21) respectively.

4.3 Asymptotics when $\underline{\lambda}$ is Unknown

In this section we show that if $\hat{\underline{\lambda}}_n$ is the maximum likelihood estimator of $\underline{\lambda}$, then the test statistic $T_n(\hat{\underline{\lambda}}_n)$ is Pitman regular and we derive an expression for $ARE(T_n(\underline{\lambda}_*), T_n(\hat{\underline{\lambda}}_n))$.

Lemma 4.3.1 Under both H_0 and H_1,

i. $\bar{h}_n^2(\theta_0, \hat{\underline{\lambda}}_n) - \bar{h}_n^2(\theta_0, \underline{\lambda}_*) \xrightarrow{P} 0$

ii. $(1/n) \sum_{i=1}^{n} h^2(Y_i - \theta_0, \hat{\underline{\lambda}}_n) - (1/n) \sum_{i=1}^{n} h^2(Y_i - \theta_0, \underline{\lambda}_*) \xrightarrow{P} 0$

The proof of this lemma is given in Appendix D.

Theorem 4.3.1 $\hat{\sigma}_n(\theta_0, \hat{\underline{\lambda}}_n)$ is a consistent estimator of σ under both the null and alternative models.

Proof Since by the results of Sections 4.2.1 and 4.2.2

$$\hat{\sigma}^2(\theta_0, \underline{\lambda}_*) \xrightarrow{P} \sigma^2$$

under both the null and alternative models, the proof is immediate from Lemma 4.3.1.

Lemma 4.3.2 If $\bar{\underline{\lambda}}_n$ is a consistent estimator of $\underline{\lambda}$, then

$$(1/n) \sum_{i=1}^{n} \frac{\partial^2 h(Y_i - \theta_0, \bar{\lambda}_{1n})}{\partial \lambda_1^2} \quad \text{and} \quad (1/n) \sum_{i=1}^{n} \frac{\partial^2 h(Y_i - \theta_0, \bar{\lambda}_{2n})}{\partial \lambda_2^2}$$

are bounded in probability under the null and alternative models.

Lemma 4.3.3 Under both the null and alternative models

$$(1/n) \sum_{i=1}^{n} \frac{\partial h(Y_i - \theta_0, \underline{\lambda}_*)}{\partial \underline{\lambda}}$$

converges in probability to the same limit,

$$E[S(\epsilon, \underline{\lambda}_*)] = \begin{bmatrix} E\{ [-\lambda_{1*}^{-2}(1-\lambda_{1*}\sigma\epsilon)\ln(1-\lambda_{1*}\sigma\epsilon) - \lambda_{1*}^{-1}\sigma\epsilon] I_{(-\infty,0)}(\epsilon) \} \\ \\ E\{ [\lambda_{2*}(1+\lambda_{2*}\sigma\epsilon)\ln(1+\lambda_{2*}\sigma\epsilon) - \lambda_{2*}^{-1}\sigma\epsilon] I_{(0,\infty)}(\epsilon) \} \end{bmatrix}$$

The proofs of the above two lemmas are given in Appendix D.

<u>Theorem 4.3.2</u> Let $\Psi = 0$ and let

$$\tau^2 = \mathrm{var}[\epsilon + U_*^t I^* E\{S(\epsilon, \underline{\lambda}_*)\}/\sigma]$$

where U_* and I^* are defined in Section 4.1. Under the null model

$$\frac{T_n(\hat{\lambda}_n) - \Psi}{\tau} \xrightarrow{\mathscr{L}} N(0,1) .$$

<u>Proof</u> Consider a Taylor expansion of $\bar{h}_n(\theta_0, \hat{\underline{\lambda}}_n)$ about $\hat{\underline{\lambda}}_n = \underline{\lambda}_*$.

$$\sqrt{n} \sum_{i=1}^{n} h(Y_i - \theta_0, \hat{\underline{\lambda}}_n)/n = \sqrt{n} \sum_{i=1}^{n} h(Y_i - \theta_0, \underline{\lambda}_*)/n + \sqrt{n}(\hat{\underline{\lambda}}_n - \underline{\lambda}_*)^t \Big(\sum_{i=1}^{n} \frac{\partial h(Y_i - \theta_0, \underline{\lambda}_*)}{\partial \underline{\lambda}} \Big)/n$$

$$+ (\sqrt{n}/2n)(\hat{\lambda}_{1n} - \lambda_{1*})^2 \sum_{i=1}^{n} \frac{\partial^2 h(Y_i - \theta_0, \tilde{\lambda}_{1n})}{\partial \lambda_1^2}$$

$$+ (\sqrt{n}/2n)(\hat{\lambda}_{2n} - \lambda_{2*})^2 \sum_{i=1}^{n} \frac{\partial^2 h(Y_i - \theta_0, \tilde{\underline{\lambda}}_{1n})}{\partial \lambda_2^2} \qquad (4.3.3)$$

where $\tilde{\lambda}_{1n}$ and $\tilde{\lambda}_{2n}$ are such that, $|\tilde{\lambda}_{1n} - \lambda_{1*}| \le |\hat{\lambda}_{1n} - \lambda_{1*}|$ and $|\tilde{\lambda}_{2n} - \lambda_{2*}| \le |\hat{\lambda}_{2n} - \lambda_{2*}|$. Since by Appendix E $(\hat{\lambda}_{kn} - \lambda_{k*}) \xrightarrow{p} 0$ and $\sqrt{n}\,(\hat{\lambda}_{kn} - \lambda_{k*})$ has a limiting normal distribution, and by Lemma 4.3.3 the second order derivatives in the above expansion are bounded in probability, therefore the two terms including these derivatives tend in probability to 0 . Denote the sum of these two terms by R_n so that $R_n \xrightarrow{p} 0$. Under H_0

$$\sqrt{n} \sum_{i=1}^{n} h(Y_i - \theta_0, \hat{\underline{\lambda}}_n)/n = \sqrt{n}(1/n) \sum_{i=1}^{n} \sigma\epsilon_i + \sqrt{n}(1/n)\ (\hat{\underline{\lambda}}_n - \underline{\lambda}_*)^t$$

$$\Big(\sum_{i=1}^{n} \frac{\partial h(Y_i - \theta_0, \underline{\lambda}_*)}{\partial \underline{\lambda}} \Big)/n + R_n . \qquad (4.3.4)$$

Let

$$Q_n = \sqrt{n}(1/n) \sum_{i=1}^{n} \sigma\epsilon_i + \sqrt{n}(\hat{\underline{\lambda}}_n - \underline{\lambda}_*)^t\ E\{(S(\epsilon, \underline{\lambda}_*)\} . \qquad (4.3.5)$$

Since by Lemma 4.3.3

$$(1/n) \sum_{i=1}^{n} \frac{\partial h(Y_i-\theta_0,\underline{\lambda}_*)}{\partial \underline{\lambda}} - E\{S(\epsilon,\underline{\lambda}_*)\} \xrightarrow{p} 0 ,$$

hence subtracting (4.3.5) from (4.3.4) and noting that $\sqrt{n}\ (\hat{\underline{\lambda}}_n-\underline{\lambda}_*)^t$ tends to a limiting normal distribution by Theorem E.1 in Appendix E we get

$$\sqrt{n} \sum_{i=1}^{n} h(Y_i-\theta_0,\hat{\underline{\lambda}}_n)/n - Q_n \xrightarrow{p} 0 . \tag{4.3.6}$$

Let

$$W_i = \sigma\epsilon_i + U^t(\epsilon_i,\underline{\lambda}_*)\ I^* E\ \{S(\epsilon,\underline{\lambda}_*)\} \tag{4.3.7}$$

From Lemma E.5 we have $\sqrt{n}\ (\hat{\underline{\lambda}}_n-\underline{\lambda}_*) - \sqrt{n}\ I^{\lambda\lambda} \sum_{i=1}^{n} U(\epsilon_i,\underline{\lambda}_*)/n \xrightarrow{p} 0$.

From (4.3.5) and (4.3.7) by subtraction we get

$$\sqrt{n}\ (1/n) \sum_{i=1}^{n} W_i - Q_n \xrightarrow{p} 0 . \tag{4.3.8}$$

From (4.3.6) and (4.3.8) we get

$$\sqrt{n}(1/\sigma n) \sum_{i=1}^{n} h(Y_i-\theta_0,\hat{\underline{\lambda}}_n) - \sqrt{n}(1/\sigma n) \sum_{i=1}^{n} W_i \xrightarrow{p} 0 . \tag{4.3.9}$$

Since it is assumed that $E(\epsilon_i) = 0$ and it is known that $E\{U(\epsilon_i,\underline{\lambda}_*)\} = 0$, hence $E\{W_i\} = \Psi = 0$. From (4.3.7) $\mathrm{var}(W_i/\sigma) = \tau^2$.
By the CLT $\sqrt{n}(1/\sigma n) \sum_{i=1}^{n} W_i \xrightarrow{\mathcal{L}} N(0,1)$. From (4.3.9)

$$(\sqrt{n} \sum_{i=1}^{n} h(Y_i-\theta_0,\hat{\underline{\lambda}}_n)/n)/(\sigma\ \tau_n(\theta_0)) \xrightarrow{\mathcal{L}} N(0,1) .$$

Since by Theorem 4.3.1 $\hat{\sigma}(\theta_0,\hat{\underline{\lambda}}_n)/\sigma \xrightarrow{p} 1$, therefore,

$$\frac{T_n(\hat{\lambda}_n) - \Psi}{\tau} \xrightarrow{\mathcal{L}} N(0,1) . \quad []$$

<u>Theorem 4.3.3</u> Let $\Psi = (k_1/\sigma)\ E\{g(\epsilon,\underline{\lambda}_*)\}$ where, $g(\epsilon,\underline{\lambda}_*)$ is defined in (4.2.7) and let $\tau^2 = \mathrm{var}(\epsilon + U_*^t I^*\ E\{S(\epsilon,\underline{\lambda}_*)\}/\sigma)$. Under the contiguous alternative model (4.2.3)

$$\frac{T_n(\hat{\lambda}_n) - \Psi}{\tau} \xrightarrow{\mathcal{L}} N(0,1) .$$

Proof

Consider the Taylor expansion given by (4.3.3). Since the result of Lemma 4.3.3 holds under the alternative hypothesis, hence using the same argument as in the previous theorem the two terms including the second order derivatives go in probability to 0 . As in the proof of Theorem 4.2.2, under the alternative model

$$\sqrt{n}\,\bar{h}_n(\theta_0,\underline{\lambda}_*)/\sigma - [\sqrt{n}\,\bar{\epsilon}_n + (k_1/\sigma)E\{g(\epsilon,\underline{\lambda}_*)\}] \xrightarrow{p} 0 .$$

By Lemma 4.3.3 under the alternative model,

$$(1/n)\sum_{i=1}^{n}\frac{\partial h(Y_i-\theta_0,\underline{\lambda}_*)}{\partial\underline{\lambda}} - E\{(S(\epsilon,\underline{\lambda}_*)\} \xrightarrow{p} 0 .$$

Using an argument similar to that given in the proof of the previous theorem we can write

$$\sqrt{n}\,\bar{h}_n(\theta_0,\hat{\underline{\lambda}}_n)/\sigma = \sqrt{n}(1/n)\sum_{i=1}^{n}[\epsilon_i + U^t(\epsilon_i,\underline{\lambda}_*)I^{-1}(\underline{\lambda}_*)E\{S(\epsilon,\underline{\lambda}_*)\}/\sigma]$$
$$+ (k_1/\sigma)E\{g(\epsilon,\underline{\lambda}_*)\} + R_n ,$$

where $R_n \xrightarrow{p} 0$. Let

$$W_i = \epsilon_i + U^t(\epsilon_i,\underline{\lambda}_*)I^{*}E\{S(\epsilon,\underline{\lambda}_*)\}/\sigma + (k_1/\sigma)E\{g(\epsilon,\underline{\lambda}_*)\},$$

then,

$$E(W_i) = k_1/\sigma\ E\{g(\epsilon,\underline{\lambda}_*)\}$$

and

$$\mathrm{var}(W_i/\sigma) = \tau^2 .$$

Hence Theorem 4.3.1 and the CLT imply,

$$\frac{T_n(\hat{\lambda}_n) - \Psi}{\tau} \xrightarrow{\mathcal{L}} N(0,1) \quad []$$

Theorem 4.3.4 The test statistic $T_n(\hat{\underline{\lambda}}_n)$ is Pitman regular.

Proof

let $\Psi_n(\theta) = (\sqrt{n}/\sigma)(\theta-\theta_0)\ E\{g(\epsilon,\underline{\lambda}_*)\}$ and let

$\tau_n^2(\theta) = \mathrm{var}[\epsilon + U_*^t\ I^{*}E\{S(\epsilon,\underline{\lambda})\}/\sigma]$. Note that $\Psi_n(\theta_n)$ and $\tau_n(\theta)$

under H_0 and H_1 are the same as those defined in Theorems 4.3.2 and 4.3.3 respectively. Hence conditions C1 and C2 of Appendix B follow by these two theorems respectively. It remains to verify C3 through C7.

C3. $\Psi_n(\theta)$ is differentiable for all θ

C4. $\Psi_n'(\theta) = (1/\sigma)E\{g(\epsilon,\underline{\lambda}_*)\} > 0$, because $g(\epsilon,\underline{\lambda}_*) = -\dfrac{\partial \bar{h}_n(\theta,\underline{\lambda}_*)}{\partial\theta}$

which is positive everywhere.

$$\text{C5.}\quad \frac{\Psi_n'(\theta_0)}{\sqrt{n}\ \tau_n(\theta_0)} = \frac{E\{g(\epsilon,\underline{\lambda}^*)\}/\sigma}{[\mathrm{var}(\ \epsilon + U_*^t\ I^* E\{S(\epsilon,\underline{\lambda}_*)\}/\sigma]^{1/2}} > 0\ .$$

C6. $\Psi_n'(\theta)$ is the same for all θ hence,

$$\sup_{\theta_0 \le \tilde{\theta}_n \le \theta_n} \left| \Psi_n'(\tilde{\theta}_n)/\Psi_n'(\theta_0) - 1 \right| = 1 \qquad \text{for all n}$$

C7. $\tau_n^2(\theta_n)/\tau_n^2(\theta_0) = 1$ for all n

Therefore $T_n(\hat{\underline{\lambda}}_n)$ is Pitman regular. []

From Appendix B

$$R_{3n}^2 = \frac{n\ \ E^2\{g(\epsilon,\underline{\lambda}^*)\}/\sigma}{[\mathrm{var}(\epsilon + U_*^t\ I^* E\{S(\epsilon,\underline{\lambda}_*)\}/\sigma]} \tag{4.3.10}$$

From (4.2.20) and (4.3.8) we get,

$$\mathrm{ARE}(T_n(\underline{\lambda}_*),T_n(\hat{\underline{\lambda}}_n)) = \frac{\mathrm{var}(\sigma\epsilon + U_*^t I^* E\{S(\epsilon,\underline{\lambda})\})}{\sigma^2} \tag{4.3.11}$$

4.4 Examples

In this section we use numerical integration to evaluate the asymptotic relative efficiency of the t-test in the original scale relative to the transformed t-test using the MLE of the transformation parameters. The model we consider is the transformed-normal model by the two-domain family. This is done in two steps. In the first step we evaluate the asymptotic relative efficiency of the t-test in the original scale relative to the transformed t-test when the transformation is known. Next, we evaluate the asymptotic relative efficiency of the transformed t-test when the transformation is unknown and the MLE's of the transformation parameters are used, relative to the corresponding test when the transformation is known. the product of the above two efficiencies gives the required efficiency. Tables 4.1 through 4.4 below give the asymptotic relative efficiency for fixed λ_2 and different values of λ_1 and σ . From these tables note that:

1. There is no gain in efficiency when the transformed t-test is used with small values of σ . The gain starts to be considerable if σ is at least 1/2 .

2. The gain in efficiency as λ_1 varies has a parabolic shape with the smallest gain being when λ_1 is slightly greater than λ_2 .

3. When $\lambda_1 = \lambda_2$ the two-domain family reduces to the John-Draper family and the columns representing this situation are the same as those of Table 3.1.

Table 4.1 $ARE(\tilde{T}_n(1), T_n(\underline{\lambda}_*))$ $\lambda_2 = 1/4$

$\sigma \backslash \lambda_1$	1/4	1/3	1/2	3/4	1
0.1	0.9955	0.996	0.9965	0.9965	0.9955
0.25	0.974	0.9769	0.98	0.979	0.973
0.5	0.9092	0.92	0.9296	0.9236	0.9023
1	0.7294	0.7615	0.7814	0.752	0.6935
2	0.4043	0.4596	0.4688	0.3964	0.3145
3	0.2172	0.2668	0.257	0.1866	0.1289
4	0.1214	0.1581	0.1407	0.089	0.0549
5	0.0716	0.0971	0.0797	0.0449	0.0252

Table 4.2 $ARE(\tilde{T}_n(1), T_n(\underline{\lambda}_*))$ $\lambda_2 = 1/3$

$\sigma \backslash \lambda_1$	1/4	1/3	1/2	3/4	1
0.1	0.9960	0.9965	0.9972	0.9973	0.9965
0.25	0.9769	0.9802	0.9842	0.9844	0.9794
0.5	0.9200	0.9329	0.9464	0.9449	0.9270
1	0.7615	0.8044	0.8412	0.8265	0.7741
2	0.4596	0.5584	0.6218	0.5650	0.4671
3	0.2668	0.3884	0.4512	0.3677	0.2683
4	0.1581	0.2790	0.3313	0.2411	0.1577
5	0.0971	0.2076	0.2483	0.1625	0.0966

Table 4.3 ARE($\tilde{T}_n(1), T_n(\underline{\lambda}_*)$) $\lambda_2 = 1/2$

$\sigma \backslash \lambda_1$	1/4	1/3	1/2	3/4	1
0.1	0.9965	0.9972	0.9981	0.9986	0.9981
0.25	0.9800	0.9842	0.9897	0.9922	0.9891
0.5	0.9296	0.9464	0.9667	0.9740	0.9631
1	0.7814	0.8412	0.9075	0.9241	0.8893
2	0.4688	0.6218	0.7906	0.8137	0.7268
3	0.2570	0.4512	0.6969	0.7145	0.5882
4	0.1407	0.3313	0.6244	0.6315	0.4801
5	0.0797	0.2483	0.5675	0.5629	0.3972

Table 4.4 ARE($\tilde{T}_n(1), T_n(\underline{\lambda}_*)$) $\lambda_2 = 1$

$\sigma \backslash \lambda_1$	1/4	1/3	1/2	3/4	1
0.1	0.9955	0.9965	0.9981	0.9995	1
0.25	0.9730	0.9794	0.9891	0.9975	1
0.5	0.9023	0.9270	0.9631	0.9920	1
1	0.6935	0.7741	0.8893	0.9774	1
2	0.3145	0.4671	0.7268	0.9447	1
3	0.1289	0.2683	0.5882	0.9135	1
4	0.0549	0.1577	0.4801	0.8851	1
5	0.0252	0.0966	0.3972	0.8593	1

Tables 4.5 through 4.8 below give the asymptotic relative efficiency of the transformed t-test using the MLE's of the transformation parameters relative to the corresponding test using a known transformation. In this situation it should be expected to find that the resulting asymptotic relative efficiency is at least 1 but, as the tables show we always get the evaluated efficiency smaller than 1 . The program given on page 169 of Appendix F used to evaluate this efficiency have been written in two different ways to make sure that it does not have any mistakes and we get the same results. Also each statistic calculated in the program including the asymptotic relative efficiency itself have been simulated for the special case where $\lambda_1 = \lambda_2 = \sigma = 1$ using 20,000 runs of samples of size 100 . It was found

that every simulated formula lies within 1 standard error of its corresponding true value. When we tried to add some more terms to the Taylor expansion given by (4.3.6) and on which (4.3.10) is based we found that all higher order terms tends to 0 in the limit and does not have any effect on the involved expressions. We tried to neglect the adjustment due to estimating σ when $I^{\lambda\lambda}$ is evaluated but we get the same type of result. This unusual result can be added to a similar type of results discussed by Freedman and Stephen (1984).

Table 4.5 $ARE(T_n(\underline{\lambda}_*), T_n(\hat{\lambda}_n))$ $\lambda_2 = 1/4$

σ \ λ_1	1/4	1/3	1/2	3/4	1
0.1	0.8556	0.8553	0.8548	0.8542	0.8536
0.25	0.8533	0.8528	0.8518	0.8505	0.8496
0.5	0.8500	0.8490	0.8475	0.8459	0.8449
1	0.8443	0.8428	0.8408	0.8394	0.8393
2	0.8356	0.8336	0.8315	0.8315	0.8334
3	0.8289	0.8267	0.8251	0.8265	0.8301
4	0.8236	0.8213	0.8202	0.8230	0.8280
5	0.8191	0.8168	0.8162	0.8202	0.8265

Table 4.6 $ARE(T_n(\underline{\lambda}_*), T_n(\hat{\lambda}_n))$ $\lambda_2 = 1/3$

σ \ λ_1	1/4	1/3	1/2	3/4	1
0.1	0.8553	0.8551	0.8546	0.8539	0.8533
0.25	0.8528	0.8522	0.8511	0.8499	0.8488
0.5	0.8490	0.8480	0.8463	0.8445	0.8434
1	0.8428	0.8411	0.8387	0.8368	0.8363
2	0.8336	0.8310	0.8281	0.8270	0.8280
3	0.8267	0.8236	0.8206	0.8205	0.8230
4	0.8213	0.8178	0.8149	0.8157	0.8194
5	0.8168	0.8131	0.8103	0.8120	0.8167

Table 4.7 $ARE(T_n(\underline{\lambda}_*), T_n(\hat{\lambda}_n))$ $\lambda_2 = 1/2$

σ \ λ_1	1/4	1/3	1/2	3/4	1
0.1	0.8548	0.8546	0.8541	0.8534	0.8527
0.25	0.8518	0.8511	0.8500	0.8486	0.8475
0.5	0.8475	0.8463	0.8443	0.8422	0.8408
1	0.8408	0.8387	0.8356	0.8328	0.8315
2	0.8315	0.8281	0.8236	0.8206	0.8202
3	0.8251	0.8206	0.8153	0.8125	0.8129
4	0.8202	0.8149	0.8091	0.8065	0.8076
5	0.8162	0.8103	0.8041	0.8018	0.8035

Table 4.8 $ARE(T_n(\underline{\lambda}_*), T_n(\hat{\lambda}_n))$ $\lambda_2 = 1$

σ \ λ_1	1/4	1/3	1/2	3/4	1
0.1	0.8536	0.8533	0.8527	0.8520	0.8513
0.25	0.8496	0.8488	0.8475	0.8458	0.8443
0.5	0.8449	0.8434	0.8408	0.8378	0.8356
1	0.8393	0.8363	0.8315	0.8267	0.8236
2	0.8334	0.8280	0.8202	0.8130	0.8091
3	0.8301	0.8230	0.8129	0.8042	0.7999
4	0.8280	0.8194	0.8076	0.7979	0.7934
5	0.8265	0.8167	0.8035	0.7930	0.7883

Multiplying each entry of Table 4.1 by its corresponding entry of Table 4.5 and so forth, we get the last 4 tables which give the asymptotic relative efficiency of the t-test in the original scale relative to the transformed t-test under the two-domain family.

Table 4.9 $ARE(\bar{T}_n(1), T_n(\hat{\lambda}_n))$ $\lambda_2 = 1/4$

$\sigma \backslash \lambda_1$	1/4	1/3	1/2	3/4	1
0.1	0.8517	0.8519	0.8519	0.8512	0.8497
0.25	0.8311	0.8331	0.8347	0.8327	0.8267
0.5	0.7728	0.7811	0.7879	0.7813	0.7624
1	0.6159	0.6418	0.6570	0.6312	0.5820
2	0.3378	0.3831	0.3898	0.3296	0.2621
3	0.1800	0.2206	0.2121	0.1542	0.1070
4	0.1000	0.1298	0.1154	0.0733	0.0455
5	0.0587	0.0793	0.0650	0.0368	0.0208

Table 4.10 $ARE(\bar{T}_n(1), T_n(\hat{\lambda}_n))$ $\lambda 2 = 1/3$

σ \ λ_1	1/4	1/3	1/2	3/4	1
0.1	0.8519	0.8521	0.8522	0.8516	0.8503
0.25	0.8331	0.8354	0.8377	0.8366	0.8313
0.5	0.7811	0.7911	0.8009	0.7980	0.7819
1	0.6418	0.6766	0.7055	0.6916	0.6474
2	0.3831	0.4640	0.5149	0.4672	0.3867
3	0.2206	0.3199	0.3703	0.3017	0.2208
4	0.1298	0.2282	0.2700	0.1967	0.1292
5	0.0793	0.1688	0.2012	0.1319	0.0789

Table 4.11 $ARE(\bar{T}_n(1), T_n(\hat{\lambda}_n))$ $\lambda_2 = 1/2$

σ \ λ_1	1/4	1/3	1/2	3/4	1
0.1	0.8519	0.8522	0.8524	0.8522	0.8511
0.25	0.8347	0.8377	0.8413	0.8419	0.8383
0.5	0.7879	0.8009	0.8163	0.8203	0.8098
1	0.6570	0.7055	0.7583	0.7696	0.7395
2	0.3898	0.5149	0.6511	0.6677	0.5961
3	0.2121	0.3703	0.5682	0.5806	0.4781
4	0.1154	0.2700	0.5052	0.5093	0.3877
5	0.0650	0.2012	0.4563	0.4513	0.3191

Table 4.12 ARE($\bar{T}_n(1), T_n(\hat{\lambda}_n)$) $\lambda_2 = 1$

σ \ λ_1	1/4	1/3	1/2	3/4	1
0.1	0.8497	0.8503	0.8511	0.8516	0.8513
0.25	0.8267	0.8313	0.8383	0.8437	0.8443
0.5	0.7624	0.7819	0.8098	0.8311	0.8356
1	0.5820	0.6474	0.7395	0.8080	0.8236
2	0.2621	0.3867	0.5961	0.7681	0.8091
3	0.1070	0.2208	0.4781	0.7347	0.7999
4	0.0455	0.1292	0.3877	0.7062	0.7934
5	0.0208	0.0789	0.3191	0.6815	0.7883

5. SIMULATION STUDY

In this chapter we present the results of simulations of the level and power of the testing procedures discussed in the previous chapters to test for the median of a certain distribution. These are: the t-test calculated from the original observations, the transformed t-test using the John-Draper or the two-domain family, the trimmed t-test and the trimmed transformed t-test with 10% and 20% of the observations on each side being trimmed. This simulation study is intended to serve two purposes. The first is to use finite sample sizes to support the asymptotic results of Chapters 3 and 4 . This is done by comparing the simulation results of the first two test statistics above. The second purpose is to try to give some general trends through overall comparisons among the different testing procedures. For example, given some implications from the data about symmetry, tail heaviness and degree of spread, can we state that a certain test may be recommended because it is the best or because it is safer under the set of data at hand?

In Section 5.1 we give a general description of the simulation study. Section 5.2 contains a discussion of the conclusions that can be drawn from the study. These conclusions are based on a large number of runs. Details of only a part of these runs are given in the remaining three sections of this chapter. Section 5.3 is devoted to the simulated power and level of the different test statistics under the normal model. Section 5.4 is devoted to symmetric non-normal models transformed by the John-Draper transformation. In Section 5.5 we consider skewed models where the two-domain family is applied.

5.1 Description of the Simulation Study

As mentioned above we report the level and power of the different testing procedures as output of the simulation runs. The input for each run is some combination of a number of factors. In the following we discuss each of these factors.

1. The model used to generate the data. The simulation results are reported both for symmetric models where the John-Draper family is applied and for skewed models where the two-domain family is applied. The symmetric models we consider are, the normal model, the transformed normal model, the contaminated normal model, the transformed contaminated normal model, the Student's t model and the transformed Student's t model. Under the transformed models mentioned above the data are generated by applying the inverse of the John-Draper transformation for some λ smaller than 1 to the associated symmetric model so that we generate data sets with heavier tails than those from these symmetric models. The skewed models we consider are the transformed normal model, the Gamma model and the extreme-value model. The data for the transformed normal model in the last case is generated by applying the inverse of the two-domain family to normal data using different values of the transformation parameters λ_1 and λ_2.

2. Transformation parameters. For symmetric transformed models, λ is chosen to be some value from the set $\{0 , 1/4 , 1/3 , 1/2\}$. Note that $\lambda = 0$ corresponds to the log-transformation and the case $\lambda = 1$ corresponds to the no-transformation situation. For the transformed normal model under the two-domain family, we fix λ_2 at 1/4 and run successive runs for λ_1 equal 1/2 , 3/4 and 1 .

3. <u>Scale parameter.</u> Recall that our model assumption under transformations is $h(Y-\theta,\lambda) = \sigma\epsilon$ where ϵ has some specified distribution. When ϵ has a standard normal distribution, σ represents the standard deviation of $h(Y-\theta,\lambda)$. In this case, to study the effect of the degree of spread of the data, σ is taken to be 2, 8, 18, 32 and 50 . The reason for this choice is to get some results for data that represents the difference between two normal samples each with standard deviation respectively 1 , 2 , 3 , 4 , and 5 . Under non-normal distributions of ϵ , σ is taken to be 1 , 2 or 3 , which represents some variety of scale multipliers of the standard deviation of ϵ to allow for models with different spread.

4. <u>Sample size.</u> Different runs are made for samples of size 10 , 20 and 50 . These sizes are chosen to represent small, moderate and large samples.

5. <u>Alternative models.</u> In each run we start by simulating the significance level corresponding to the .05 nominal level. Then we simulate the powers under alternative models obtained by successively adding to the data .2 or .1 times the standard deviation of the generated random variable, until the simulated power is over .95 .

6. <u>Number of simulations.</u> The number of simulations under the null model is 30,000 . This makes the standard error of the simulated .05 nominal level under a normal model approximately equal to .00125 so that the limits of a 99% confidence interval for the level are approximately (.0467 , .0533) . Given the above number of simulations we may consider any test with simulated level in this interval to have correct level. Any test with simulated level considerably greater than the upper limit may be considered as an invalid test. The number of

simulations under alternative models is 5,000 . This makes the maximum estimate of the standard error of the simulated power under a normal model, when it is .5 , equal to .0071 with ± .0183 giving the limits of a 99% confidence interval. The only exception from the above rule occurs under the Student's t with two degrees of freedom model where its variance is infinity and we calculate the power by successively adding .2 to the data. To make a pairwise comparison between any two of the test statistics we propose to use McNemar's test for paired data.

The GAUSS software is used to generate the data under the different models specified above and to run all programs. All data sets from the above models can be generated from normal or uniform random variables. GAUSS has a normal random number generator based on the fast acceptance-rejection algorithm (see Kinderman and Ramage, 1976). Also, GAUSS has a uniform random number generator based on the multiplicative congruential method (see Kennedy and Gentle, 1980). All data sets are generated using the number 9831815 as an initial seed. The maximum likelihood estimators of the transformation parameters under both the John-Draper and the two-domain families are evaluated by solving for values of the parameters at which the first derivative of a normal loglikelihood is zero. The equations are solved using module 10 of the GAUSS procedures, which finds the roots of a system of nonlinear equations using Broyden's secant method (see Dennis and Schamble, 1983).

5.2 General Simulation Results

As mentioned earlier the details given in the next three sections are only a part of the full simulation study. In this section we try to give the main conclusions from the whole study without including a large number of tables. The results listed below are classified according to the models they represent as, results for the normal model, results for symmetric non-normal models and results for skewed models.

5.2.1 Normal model

Under the normal model the t-test in the original scale of the data is the optimal test procedure to consider. Comparisons of other test procedures with this optimal test give an indication about the performance of these alternative procedures under the true model. Although it is not expected that such alternative procedures will perform as good as the optimal procedure under normality, it may be required that they perform reasonably well with the hope that they would have a better performance under non-normal situations.

We simulated the level and power of the different procedures using normal data with variance 2 and 50 for the three sample sizes 10 , 20 and 50 . The main results from these simulations are as follows:

1. The simulated significance level is found to belong to the 99% confidence interval (.0467 , .0533) under all situations except in two cases corresponding to the trimmed and to the transformed-trimmed tests with proportion 20% (see Tables 5.1 and 5.2) .

2. The power of the transformed t-test is almost the same as that of the optimal test. The maximum difference in all runs is about .008 .

3. Trimming under normality is somewhat harmful. It decreases the power by as much as .03 for 10% trimming and as much as .105 for 20% trimming.

The above results indicate that the 20% trimmed tests are the worst procedures under the normal model. Also, an experimenter may be reluctant to discard 40% of the data. However, as will be shown latter there are some situations in which these two tests have the best simulated powers.

5.2.2 Symmetric non-normal models

The results we report here correspond to the models considered in Section 5.4. These models are all symmetric with distributions that have tails heavier than those of a normal distribution. In the following we give an outline of the general trends found from the simulation runs, then we discuss the effect of the different factors considered in Section 5.1 on the level and power of the different testing procedures.

1. The maximum simulated significance level under all the simulated models and all the different testing procedures is found to be .0529 which indicate that all these procedures have approximately the correct level.

2. Under very heavy tailed models like the transformed Student's t and the transformed contaminated-normal models the untransformed tests appear to be very conservative. In some runs the simulated significance levels of these tests were as low as .02 and in many cases they are found to be around .03 .

3. The simulated significance level of the transformed t-test and the transformed, 10% trimmed t-test always appear to be fairly close to the nominal level; the smallest simulated level is found to be greater than .047 . The simulated level of the transformed, 20% trimmed t-test is sometimes as low as .04113; however, in most cases it is also found to be close to .05 .

4. The t-test in the original scale under non-normal heavy-tailed symmetric models is found to be the worst testing procedure. Under the very heavy-tailed models referred to above, and for samples of size 20 , the simulated power of the transformed t-test is sometimes more than 6 times that of the t-test in the original scale. This result supports the results of Chapter 3. However, under other models and for some special values of the parameters, as will be discussed latter, there is almost no difference among all the different testing procedures.

5. A comparison of the transformed t-test with the trimmed tests shows that the trimmed tests have more power than the transformed t-test especially when the proportion of trimming is 20% . When the proportion is 10%, it is sometimes the case that the transformed t-test has more simulated power than the trimmed test for alternatives that are close to the true model.

6. From a comparison between the trimmed test and the transformed-trimmed test with 10% proportion of trimming it is found that under heavy-tailed models like the contaminated normal and Student's t models there is a slight difference between the two tests. Under such models the simulated power of the transformed-trimmed test exceeds that of the trimmed test by at most .03 . However, under transformed heavy-tailed models and under the transformed normal model, the simulated power of the transformed-trimmed test could be above that of the trimmed test by as much as .20 .

7. Comparisons of the trimmed test and the transformed-trimmed test with proportion 20% show that the trimmed test has more simulated power than the transformed-trimmed test for samples of size 10 . The difference in powers could be as much as .07 . Under larger sample

sizes this difference does not exceed .02 . However, under transformed heavy-tailed models and transformed normal models the simulated power of the transformed- trimmed test with proportion 20% could be greater than that of the corresponding trimmed test by as much as .125 .

8. In a trial to investigate the effect of small values of the scale parameter, a simulation run of the transformed normal model with $\lambda = .5$ and $\sigma^2 = .25$ showed that all the six testing procedures perform about equally well.

We now discuss the effect of the different factors involved in the simulation model on the results. It is found that the differences between the simulated powers of the transformed tests and those of the untransformed tests increase with the sample size. They also increase as the heaviness of the tails of the distribution used to generate the data increases. Note that, for the transformed-symmetric distributions, smaller values of the transformation parameter produce heavier tails. Also, these differences increase as the degree of spread of the data increases.

To conclude our discussion we state that the superiority of the transformed tests over the untransformed tests appears under certain combinations of all the above factors together, and in particular under large values of the scale parameter (to increase the spread of the data) with small values of the transformation parameter.

5.2.3 Skewed models

The simulation results given in this part correspond to the skewed models of Section 5.5 . These are the Gamma models with shape parameters 3 , 4 , and 5, the extreme value model and the transformed-normal model using the inverse of the two-domain family. Before summarizing the simulation results associated with these models, we give some implications about two problems that arise under skewed models.

The first problem concerns the location parameter the hypothesis statement involves. That is, are we interested in testing the mean or the median? Under symmetric models both are the same and this distinction need not be considered. However, under skewed models they differ and what is valid for one need not be valid for the other. For example, if we are testing for the mean using the t-test, even if the distribution of the original observations is skewed, the Central Limit Theorem implies that the t-statistic $\sqrt{n}\,(\bar{Y} - \mu)/\hat{\sigma}$ converges in law to N(0,1). If the mean μ is replaced by the median θ and $\theta \neq \mu$ then $\sqrt{n}\,(\bar{Y} - \theta)/\hat{\sigma}$ diverges in the direction of $\mu - \theta$.

The second problem concerns what type of alternative is considered. That is, are we considering right-sided alternatives (RHA) or left-sided alternatives (LHA). Under symmetry such a distinction need not be made. However, as indicated in Chapter 2, under skewed models the distribution of the t-statistic is skewed and hence it will be expected that for one type of alternative the t-test will be conservative while for the other type the t-test may not be valid in the sense that its level may be far greater than the nominal level.

Table 5.1 below gives the simulated significance levels of the transformed t-test and the t-test in the original scale under both types of alternatives and for the different models of Section 5.5.

Table 5.1 Simulated significance level corresponding to .05 nominal level

MODEL	Original RHA	Original LHA	Transformed RHA	Transformed LHA
Transforme-normal $\lambda_1=.25$ $\lambda_2=.5$	0.0332	0.0712	0.0339	0.0265
Transforme-normal $\lambda_1=.25$ $\lambda_2=.75$	0.0235	0.0983	0.0245	0.0261
Transforme-normal $\lambda_1=.25$ $\lambda_2=.5$	0.0181	0.1279	0.0243	0.0259
Gamma(4)	0.1343	0.0177	0.0177	0.0226
Gamma(5)	0.1219	0.0195	0.0220	0.0298
Extreme-Value	0.0168	0.1326	0.0266	0.0224

Note that the Gamma models are skewed to the right while the rest of the models have their longer tail to the left. The table shows that the simulated levels for alternatives in the direction of the longer tail is always much higher than the nominal level and the t-test may be judged to be invalid under this situation. For this reason we made our runs for alternatives in the other direction.

In the following we summarize the main results of Section 5.5.

1. Although the simulated level of the transformed t-test is a little bit greater than that of the t-test in the original scale, both are much smaller than .05 .

2. The transformed t-test has level $\leq .05$ under both types of alternatives (see Table 5.1) .

3. The transformed-trimmed test with proportion 10% has better level. In all runs it ranged over the interval (.041 , .054) . The corresponding test with proportion 20% in two cases gave simulated levels .057, and .055 .

4. The transformed t-test always has better simulated power than that of the test in the original scale. The difference under the Gamma and the extreme value models could as much as .05 and under the transformed normal model could be as much as .10 .

5. The transformed-trimmed tests have better simulated power than the untransformed tests by as much as .20 .

6. There does not appear to be much difference in simulated power between the transformed-trimmed tests using the two proportions of trimming.

7. In some runs powers of the trimmed test, using equal proportion of trimming on each tail, are simulated. Comparisons of these powers with those of the transformed-trimmed tests indicate that the transformed-trimmed tests have better simulated powers under alternatives that are close to the true model. Otherwise, they are almost the same.

5.3 Simulation Results for the Normal Model

Tables 5.2 , 5.3 and 5.4 below present the simulated level and power of the different test statistics applied to normal data. The column corresponding to the test $T_n(1)$ presents the simulated power for the uniformly most powerful unbiased test. The reason for giving these tables is to check the performance of the different test statistics when no correction for departure from normality is needed. From these tables note the following:

1. The simulated level of the test statistic $T_{\cdot 2}(\hat{\lambda})$ in Table 5.2 is .0451 and that of $T_{\cdot 2}(1)$ in Table 5.3 is .0542 which are respectively somewhat smaller and greater than the .05 nominal level. Except for these two cases the simulated levels are close to .05 .

2. The simulated power of the transformed t-test is almost the same as that of the uniformly most powerful unbiased test. The maximum difference between both tests in Table 5.3 is .0082 .

3. For the trimmed tests with proportion 10% there is almost no difference ($< .01$) in the simulated power between the transformed and the untransformed tests. However, the simulated power of these two tests could be smaller than that of the untrimmed tests by as much as .04 .

4. The trimmed tests with proportion 20% have the worst simulated powers in all three tables. This power could be below that of the untrimmed tests by as much as .085 .

5. When the McNemar test is used to compare the simulated level and power, the t-test in the original scale and the transformed t-test, there does not appear any significant difference between the two tests.

Table 5.2 Normal model $\sigma^2 = 2$ n = 20 $\alpha = .05$

θ/σ	$T_n(1)$	$T_n(\hat{\lambda})$	$T_{\cdot 1}(1)$	$T_{\cdot 1}(\hat{\lambda})$	$T_{\cdot 2}(1)$	$T_{\cdot 2}(\hat{\lambda})$
0	0.0488	0.0489	0.0492	0.0492	0.0512	0.0451
0.2	0.2112	0.2076	0.2024	0.2038	0.2006	0.1792
0.4	0.5330	0.5322	0.5088	0.5102	0.4702	0.4384
0.6	0.8228	0.8218	0.7954	0.7972	0.7594	0.7362
0.8	0.9622	0.9616	0.9468	0.9480	0.9294	0.9186
1	0.9962	0.9954	0.9934	0.9934	0.9864	0.9836
1.2	0.9994	0.9992	0.9994	0.9992	0.9982	0.9980

Table 5.3 Normal model $\sigma^2 = 50$ n = 10 $\alpha = .05$

θ/σ	$T_n(1)$	$T_n(\hat{\lambda})$	$T_{\cdot 1}(1)$	$T_{\cdot 1}(\hat{\lambda})$	$T_{\cdot 2}(1)$	$T_{\cdot 2}(\hat{\lambda})$
0	0.0480	0.0478	0.0505	0.0523	0.0542	0.0505
0.2	0.1396	0.1392	0.1396	0.1434	0.1406	0.1274
0.4	0.3166	0.3130	0.3042	0.3080	0.2732	0.2574
0.6	0.5386	0.5348	0.5120	0.5188	0.4646	0.4482
0.8	0.7558	0.7476	0.7166	0.7240	0.6510	0.6370
1	0.8960	0.8912	0.8630	0.8690	0.8032	0.7972
1.2	0.9652	0.9632	0.9464	0.9482	0.9086	0.9038
1.4	0.9908	0.9898	0.9804	0.9812	0.9624	0.9582

Table 5.4 Normal model $\sigma^2 = 50$ n = 50 $\alpha = .05$

θ/σ	$T_n(1)$	$T_n(\hat{\lambda})$	$T_{\cdot 1}(1)$	$T_{\cdot 1}(\hat{\lambda})$	$T_{\cdot 2}(1)$	$T_{\cdot 2}(\hat{\lambda})$
0	0.0500	0.0496	0.0508	0.0503	0.0512	0.0513
0.1	0.1650	0.1640	0.1614	0.1624	0.1580	0.1584
0.2	0.4026	0.4000	0.3798	0.3792	0.3552	0.3564
0.3	0.6650	0.6622	0.6450	0.6460	0.6148	0.6168
0.4	0.8744	0.8740	0.8552	0.8542	0.8240	0.8242
0.5	0.9706	0.9698	0.9612	0.9602	0.9446	0.9446
0.6	0.9974	0.9976	0.9940	0.9938	0.9878	0.9878

5.4 Simulation Results for Symmetric Non-Normal Models

5.4.1 Transformed normal model

The data used in simulating the power of the test statistics in Tables 5.5 through 5.10 below are generated as follows:

Let ϵ denote a random variable with a standard normal distribution. For some λ and σ let

$$Y = h^{-1}(\epsilon, \lambda, \sigma)$$

where h is the John-Draper transformation. In Tables 5.5 through 5.8 $\lambda = 1/3$, in Table 5.9 $\lambda = 1/2$ and in Table 5.10 $\lambda = 0$. σ^2 equals 2 in Table 5.5 and 50 in the rest of the tables. The sample size n is 10 and 20 in Tables 5.6 and 5.7 and is 50 otherwise. From the above specification of the parameters, a comparison of Table 5.5 and Table 5.8 reflects the effect of changing the variance (spread). Comparisons among tables 5.6, 5.7 and 5.8 reflect the effect of the different sample sizes. Finally, comparisons of Tables 5.8, 5.9 and 5.10 together give an indication about the effect of the different values of the transformation parameter. From these tables note that:

1. The simulated significance levels of the transformed tests are close to .05 while they are generally smaller than .05 for the untransformed tests. Table 5.10 represents an extreme case due to the combination of small value of λ (= 0) with large value of σ^2 (= 50) resulting in very small simulated significance levels for the untransformed tests.

2. Except for Table 5.6 the transformed t-test in general has more simulated power than the t-test in the original scale over the range of alternatives covered by this study. For example in Table 5.5 where $\sigma^2 = 2$ we could observe a difference in these two powers by as much as .085 while from Table 5.8 this difference could be about .30 . Under the extreme case of Table 5.10 where there does not appear to be any power for the t-test the wide difference between the powers is very clear.

3. In Table 5.6 the power of the transformed t-test under alternatives that are closer to the true model exceeds that of the t-test in the original scale by as much as .06 . Under alternatives that are far from the true model it is observed that the power of the untransformed t-test exceeds that of the transformed test by about .04 . The average percentage of observations more than the hypothesized value $\theta/\sigma \geq 6$ is calculated and is found to be at most 15% . Under such situations we are almost applying the shifted Box-Cox transformation with shift parameter 1 rather than the John-Draper transformation.

4. The simulated power of the transformed-trimmed test with proportion 10% is always greater than that of the corresponding untransformed-trimmed test. The difference in some cases is not considerable as in Table 5.5 . In some other cases it could be as much as .18 like in Table 5.8 or even much more as in Table 5.10 .

5. Tables 5.8 and 5.10 show situations under which the transformed-trimmed test with proportion 20% has more simulated power than the corresponding untransformed-trimmed test. This happens when the value of the transformation parameter is small, the variance is large and the sample size is large. Although the tables below show that the former test is better than the second, this is not always the case. There are situations in which the trimmed test has more simulated power than the transformed-trimmed test as the tables of Section 5.3 indicate.

6. A comparison of Table 5.5 with Table 5.8 shows that except for the simulated level the entries in the second table are much smaller due to changing the variance from 2 to 50 . However, a significant difference in simulated power between the transformed and untransformed tests is associated with the large variance.

7. From Tables 5.6 , 5.7 and 5.8 which correspond to samples of size 10 , 20 and 50 respectively, note that larger sample sizes increase the gap between the simulated power of the transformed tests and the untransformed tests.

8. From Tables 5.9 , 5.8 and 5.10 which correspond to λ equal to 1/2 , 1/3 and 0 respectively, note that better simulated powers of the transformed tests over the untransformed tests are associated with smaller values of the transformation parameter.

9. When the McNemar test statistic is calculated for the

differences between the simulated powers of the transformed t-test and the t-test in the original scale it is always found that differences as small as .01 are statistically significant.

Table 5.5 Transformed normal model $\lambda = 1/3$ $\sigma^2 = 2$

$n = 50$ $\alpha = .05$

θ/σ	$T_n(1)$	$T_n(\hat{\lambda})$	$T_{\cdot 1}(1)$	$T_{\cdot 1}(\hat{\lambda})$	$T_{\cdot 2}(1)$	$T_{\cdot 2}(\hat{\lambda})$
0	0.0490	0.0493	0.0495	0.0503	0.0492	0.0482
0.2	0.1714	0.1996	0.2064	0.2216	0.2366	0.2384
0.4	0.4098	0.4824	0.5156	0.5410	0.5752	0.5784
0.6	0.6654	0.7496	0.7992	0.8184	0.8494	0.8504
0.8	0.8568	0.9112	0.9452	0.9546	0.9688	0.9680
1	0.9590	0.9798	0.9924	0.9944	0.9968	0.9970

Table 5.6 Transformed normal model $\lambda = 1/3$ $\sigma^2 = 50$

$n = 10$ $\alpha = .05$

θ/σ	$T_n(1)$	$T_n(\hat{\lambda})$	$T_{\cdot 1}(1)$	$T_{\cdot 1}(\hat{\lambda})$	$T_{\cdot 2}(1)$	$T_{\cdot 2}(\hat{\lambda})$
0	0.0340	0.0480	0.0292	0.0523	0.0211	0.0494
0.4	0.0544	0.0870	0.0638	0.1120	0.0666	0.1170
0.8	0.0816	0.1286	0.1160	0.1806	0.1446	0.2068
1.2	0.1158	0.1724	0.1810	0.2444	0.2422	0.2920
1.6	0.1598	0.2210	0.2530	0.3144	0.3338	0.3690
2	0.2072	0.2660	0.3218	0.3778	0.4204	0.4432
2.4	0.2536	0.3094	0.3854	0.4374	0.4964	0.5036
3	0.3250	0.3706	0.4780	0.5192	0.5912	0.6258
3.6	0.3926	0.4290	0.5632	0.5962	0.6750	0.6976
4.2	0.4606	0.4790	0.6402	0.6640	0.7376	0.7512
4.8	0.5204	0.5252	0.7000	0.7170	0.7918	0.7964
5.4	0.5794	0.5698	0.7484	0.7592	0.8312	0.8342
6	0.6228	0.6068	0.7892	0.7998	0.8638	0.8654
6.6	0.6666	0.6426	0.8276	0.8330	0.8898	0.8872
7.2	0.7010	0.6770	0.8558	0.8614	0.9094	0.9060
7.8	0.7342	0.7054	0.8800	0.8822	0.9256	0.9230
8.4	0.7624	0.7292	0.8976	0.8996	0.9390	0.9334
9	0.7912	0.7518	0.9134	0.9112	0.9494	0.9432
9.6	0.8084	0.7698	0.9244	0.9238	0.9584	0.9520
10.2	0.8300	0.7894	0.9354	0.9338	0.9646	0.9572

Table 5.7 Transformed normal model $\lambda = 1/3$ $\sigma^2 = 50$

$n = 20$ $\alpha = .05$

θ/σ	$T_n(1)$	$T_n(\hat{\lambda})$	$T_{.1}(1)$	$T_{.1}(\hat{\lambda})$	$T_{.2}(1)$	$T_{.2}(\hat{\lambda})$
0	0.0428	0.0489	0.0356	0.0490	0.0296	0.0519
0.2	0.0542	0.0692	0.0522	0.0818	0.0590	0.0904
0.4	0.0648	0.1008	0.0778	0.1258	0.1020	0.1464
0.6	0.0818	0.1358	0.1072	0.1678	0.1558	0.2086
0.8	0.0986	0.1712	0.1470	0.2190	0.2192	0.2744
1	0.1200	0.2074	0.1892	0.2746	0.2938	0.3414
1.2	0.1382	0.2470	0.2376	0.3262	0.3636	0.4096
1.4	0.1672	0.2842	0.2870	0.3806	0.4344	0.4684
1.6	0.1972	0.3232	0.3408	0.4340	0.5002	0.5318
1.8	0.2292	0.3580	0.3906	0.4822	0.5566	0.5816
2	0.2606	0.3882	0.4458	0.5276	0.6162	0.6308
2.2	0.2880	0.4186	0.4910	0.5660	0.6648	0.6726
2.4	0.3172	0.4460	0.5352	0.6126	0.7114	0.7142
3	0.4042	0.5386	0.6622	0.7182	0.8184	0.8276
3.6	0.4928	0.6126	0.7590	0.7972	0.8860	0.8908
4.2	0.5794	0.6732	0.8268	0.8520	0.9272	0.9284
4.8	0.6502	0.7248	0.8750	0.8920	0.9522	0.9512
5.4	0.7068	0.7672	0.9124	0.9224	0.9694	0.9704
6	0.7590	0.8100	0.9378	0.9444	0.9816	0.9812
6.6	0.8046	0.8400	0.9580	0.9582	0.9866	0.9866

Table 5.8 Transformed normal model $\lambda = 1/3$ $\sigma^2 = 50$

$n = 50$ $\alpha = .05$

θ/σ	$T_n(1)$	$T_n(\hat{\lambda})$	$T_{.1}(1)$	$T_{.1}(\hat{\lambda})$	$T_{.2}(1)$	$T_{.2}(\hat{\lambda})$
0	0.0478	0.0496	0.0438	0.0505	0.0400	0.0512
0.2	0.0574	0.0912	0.0712	0.1084	0.0850	0.1178
0.4	0.0788	0.1516	0.1136	0.1836	0.1604	0.2182
0.6	0.1044	0.2182	0.1686	0.2744	0.2580	0.3420
0.8	0.1314	0.2948	0.2346	0.3734	0.3788	0.4730
1	0.1660	0.3668	0.3090	0.4812	0.5042	0.5914
1.2	0.2036	0.4436	0.3984	0.5786	0.6138	0.6982
1.4	0.2410	0.5104	0.4844	0.6588	0.7124	0.7792
1.6	0.2928	0.5784	0.5692	0.7326	0.7920	0.8406
1.8	0.3360	0.6398	0.6406	0.7950	0.8496	0.8912
2	0.3818	0.6920	0.7094	0.8424	0.8964	0.9252
2.2	0.4308	0.7350	0.7654	0.8798	0.9272	0.9498
2.4	0.4746	0.7762	0.8124	0.9090	0.9498	0.9658

Table 5.9 Transformed normal model $\lambda = 1/2$ $\sigma^2 = 50$

$n = 50$ $\alpha = .05$

θ/σ	$T_n(1)$	$T_n(\hat{\lambda})$	$T_{\cdot 1}(1)$	$T_{\cdot 1}(\hat{\lambda})$	$T_{\cdot 2}(1)$	$T_{\cdot 2}(\hat{\lambda})$
0	0.04943	0.0496	0.0487	0.0505	0.0476	0.0511
0.2	0.0934	0.1244	0.1184	0.1380	0.1374	0.158
0.4	0.1756	0.2412	0.2350	0.2834	0.2950	0.3324
0.6	0.2896	0.3914	0.3992	0.4738	0.5026	0.5434
0.8	0.4262	0.5498	0.5782	0.6466	0.6964	0.7306
1	0.5526	0.6858	0.7252	0.7846	0.8392	0.8542
1.2	0.6798	0.7918	0.8454	0.8842	0.9262	0.935
1.4	0.7872	0.8728	0.9226	0.9432	0.9680	0.972
1.6	0.8644	0.9276	0.9664	0.9754	0.9902	0.9906
1.8	0.9238	0.9630	0.9866	0.9916	0.9972	0.9976
2	0.9618	0.9830	0.9962	0.9976	0.9994	0.9994

Table 5.10 Transformed normal model $\lambda = 0$ $\sigma^2 = 50$

$n = 50$ $\alpha = .05$

θ/σ	$T_n(1)$	$T_n(\hat{\lambda})$	$T_{\cdot 1}(1)$	$T_{\cdot 1}(\hat{\lambda})$	$T_{\cdot 2}(1)$	$T_{\cdot 2}(\hat{\lambda})$
0	0.0067	0.0493	0.0023	0.0505	0.0025	0.0510
0.4	0.0076	0.0830	0.0026	0.0950	0.0040	0.1178
0.8	0.0076	0.1172	0.0026	0.1438	0.0076	0.1890
1	0.0076	0.1328	0.0028	0.1642	0.0092	0.2210
2	0.0076	0.2024	0.0028	0.2656	0.0272	0.3706
3	0.0076	0.2532	0.0036	0.3480	0.0525	0.4888
4	0.0076	0.2950	0.0044	0.4160	0.0850	0.5690
5	0.0076	0.3342	0.0048	0.4734	0.1204	0.6376
6	0.0076	0.3682	0.0062	0.5262	0.1642	0.6934
7	0.0076	0.3968	0.0074	0.5658	0.2026	0.7374
8	0.0076	0.4210	0.0088	0.5980	0.2450	0.7692
9	0.0076	0.4454	0.0098	0.6302	0.2758	0.7986
10	0.0076	0.4662	0.0120	0.6576	0.3100	0.8232
16	0.0076	0.5584	0.0264	0.7726	0.4720	0.9022
20	0.0076	0.6002	0.0396	0.8152	0.5568	0.9328
24	0.0076	0.6330	0.0540	0.8488	0.6228	0.9486

5.4.2 Contaminated-normal model

Tables 5.11 and 5.12 below present the simulation results for the contaminated-normal model. From these two tables we note that:

1. Except for the simulated level of the t-test in the original scale from Table 5.11 all the levels are fairly close to .05 .

2. There appears to be a significant difference between the simulated power of the transformed t-test and that of the t-test in the original scale. This difference from Table 5.11 could be as much as .085 and from Table 5.12 could as much as .11 .

3. Within the trimmed tests the differences in simulated powers among the transformed and untransformed tests are always within .02 .

Table 5.11 Contaminated-normal $Y = .9N(0,1)+.1N(0,25)$

$n=20 \quad \lambda = 1 \quad \alpha = .05$

θ/σ	$T_n(1)$	$T_n(\hat{\lambda})$	$T_{.1}(1)$	$T_{.1}(\hat{\lambda})$	$T_{.2}(1)$	$T_{.2}(\hat{\lambda})$
0	0.0423	0.0481	0.0462	0.0482	0.0500	0.0523
0.1	0.1342	0.1536	0.1610	0.1654	0.1626	0.1536
0.2	0.2962	0.3408	0.3704	0.3814	0.3630	0.3478
0.3	0.4906	0.5458	0.6208	0.6318	0.6184	0.5980
0.4	0.6336	0.7160	0.8222	0.8290	0.8178	0.8070
0.5	0.7484	0.8362	0.9320	0.9346	0.9330	0.9258
0.6	0.8226	0.9032	0.9718	0.9740	0.9840	0.9810
0.7	0.8824	0.9388	0.9860	0.9882	0.9954	0.9942
0.8	0.9220	0.9568	0.9936	0.9936	0.9986	0.9982
0.9	0.9522	0.9684	0.9970	0.9956	0.9988	0.9996

Table 5.12 Contaminated-normal $Y = .8N(0,1)+.2N(0,25)$

$n=20 \quad \lambda = 1 \quad \alpha = .05$

θ/σ	$T_n(1)$	$T_n(\hat{\lambda})$	$T_{.1}(1)$	$T_{.1}(\hat{\lambda})$	$T_{.2}(1)$	$T_{.2}(\hat{\lambda})$
0	0.0450	0.0477	0.0408	0.0468	0.0474	0.0513
0.1	0.1344	0.1630	0.1732	0.1898	0.1838	0.1916
0.2	0.2860	0.3574	0.4164	0.4390	0.4410	0.4530
0.3	0.4462	0.5560	0.6760	0.6964	0.7112	0.7278
0.4	0.5940	0.7066	0.8364	0.8476	0.8894	0.8938
0.5	0.7202	0.8098	0.9176	0.9258	0.9654	0.9664
0.6	0.8146	0.8698	0.9570	0.9624	0.9874	0.9876
0.7	0.9318	0.9332	0.9890	0.9882	0.9984	0.9984

Tables 5.13 through 5.16 below give the simulation results for the transformed-contaminated-normal data. The transformation parameter used in applying the inverse of the John-Draper family is 1/3. In the last two tables, before applying the inverse transformation we multiply the data by 3 to allow for more dispersion. From the tables note that:

1. The simulated levels of the untransformed tests are considerably smaller than .05 which indicates that the untransformed tests are conservative under the present model. On the other hand the simulated level of the transformed tests are fairly close to .05 .

2. In all the four tables there is a wide gap between the transformed t-test and the t-test in the original scale. This difference starts to appear for alternatives that are close to the true

model, and could be as much as .60 in Table 5.16.

3. Within the trimmed tests with proportion 10% the transformed tests has more simulated power than the untransformed test. This difference could be as much as .24 like in Table 5.16.

4. Within the trimmed tests with proportion 20% , there is not much difference between the simulated powers of the transformed test and the untransformed test in Tables 5.13 and 5.14. The difference is clear in Table 5.16 where it could be as much as .085 .

Table 5.13 Transformed-contaminated normal ϵ = .8N(0,1)+.2N(0,16)

$Y = h_{\lambda}^{-1}(\epsilon)$ n=20 $\lambda = 1/3$ $\alpha = .05$

θ/σ	$T_n(1)$	$T_n(\hat{\lambda})$	$T_{\cdot 1}(1)$	$T_{\cdot 1}(\hat{\lambda})$	$T_{\cdot 2}(1)$	$T_{\cdot 2}(\hat{\lambda})$
0	0.0298	0.0476	0.0350	0.0474	0.0411	0.0504
0.2	0.0962	0.1930	0.1986	0.2452	0.2510	0.2798
0.4	0.2118	0.4120	0.4734	0.5416	0.5914	0.6138
0.6	0.3300	0.5876	0.7060	0.7658	0.8280	0.8418
0.8	0.4254	0.7084	0.8338	0.8816	0.9438	0.9466
1	0.5010	0.7818	0.9040	0.9354	0.9786	0.9788
1.2	0.5730	0.8282	0.9384	0.9626	0.9902	0.9904
1.4	0.6288	0.8602	0.9562	0.9740	0.9954	0.9948
1.6	0.6718	0.8786	0.9686	0.9826	0.9984	0.9984

Table 5.14 Transformed-contaminated normal ϵ = .8N(0,1)+.2N(0,16)

$Y = h_{\lambda}^{-1}(\epsilon)$ n=50 λ = 1/3 α = .05

θ/σ	$T_n(1)$	$T_n(\hat{\lambda})$	$T_{.1}(1)$	$T_{.1}(\hat{\lambda})$	$T_{.2}(1)$	$T_{.2}(\hat{\lambda})$
0	0.0392	0.0499	0.0404	0.0477	0.0447	0.0493
0.1	0.0730	0.1582	0.1490	0.1838	0.1902	0.2100
0.2	0.1220	0.3324	0.3660	0.4448	0.4756	0.4982
0.3	0.1832	0.5496	0.6060	0.6958	0.7352	0.7574
0.4	0.2500	0.7238	0.7836	0.8636	0.9034	0.9154
0.5	0.3150	0.8380	0.8932	0.9498	0.9722	0.9788
0.6	0.3776	0.9148	0.9458	0.9826	0.9922	0.9936
0.7	0.4372	0.9548	0.9730	0.9944	0.9980	0.9980

Table 5.15 Transformed-contaminated normal $\epsilon = .8N(0,1)+.2N(0,16)$

$Y = h_{\lambda}^{-1}(3\epsilon)$ n=20 $\lambda = 1/3$ $\alpha = .05$

θ/σ	$T_n(1)$	$T_n(\hat{\lambda})$	$T_{\cdot 1}(1)$	$T_{\cdot 1}(\hat{\lambda})$	$T_{\cdot 2}(1)$	$T_{\cdot 2}(\hat{\lambda})$
0	0.0216	0.0474	0.0262	0.0474	0.0306	0.0507
0.2	0.0406	0.1278	0.0892	0.1566	0.1344	0.1904
0.4	0.0656	0.2322	0.2042	0.3102	0.3244	0.3858
0.6	0.0998	0.3376	0.3358	0.4754	0.5192	0.5696
0.8	0.1356	0.4334	0.4630	0.6156	0.6764	0.7138
1	0.1744	0.5052	0.5692	0.7030	0.7836	0.8076
1.2	0.2102	0.5598	0.6534	0.7870	0.8630	0.8802
1.4	0.2436	0.6064	0.7238	0.8342	0.9102	0.9188
1.6	0.2746	0.6470	0.7756	0.8730	0.9414	0.9484
1.8	0.3062	0.6780	0.8078	0.8996	0.9620	0.9650
2	0.3308	0.7060	0.8366	0.9196	0.9710	0.9750
2.2	0.3562	0.7242	0.8578	0.9342	0.9790	0.9814

Table 5.16 Transformed-contaminated normal $\epsilon = .8N(0,1)+.2N(0,16)$

$Y = h_{\lambda}^{-1}(3\epsilon)$ n=50 $\lambda = 1/3$ $\alpha = .05$

θ/σ	$T_n(1)$	$T_n(\hat{\lambda})$	$T_{.1}(1)$	$T_{.1}(\hat{\lambda})$	$T_{.2}(1)$	$T_{.2}(\hat{\lambda})$
0	0.0339	0.0495	0.0332	0.0475	0.0389	0.0491
0.1	0.0402	0.1122	0.0724	0.1270	0.1082	0.1444
0.2	0.0478	0.2012	0.1430	0.2596	0.2436	0.3132
0.3	0.0596	0.3046	0.2400	0.4214	0.4154	0.5012
0.4	0.0742	0.4236	0.3546	0.5746	0.5872	0.6706
0.5	0.0852	0.5324	0.4682	0.7068	0.7196	0.7906
0.6	0.0984	0.6192	0.5736	0.7998	0.8232	0.8824
0.7	0.1128	0.6970	0.6690	0.8670	0.8922	0.9330

5.4.3 Student's t model

Tables 5.17 and 5.18 below give the simulation results for the Student's t model with 2 and 3 degrees of freedom respectively. From these tables we note that:

1. The simulated significant levels are not too far from .05 . However, the levels of the transformed t-test and the transformed-trimmed with proportion 10% are closer to the above nominal level.

2. In Table 5.17 there is some difference in the simulated power between the transformed t-test and the t-test in the original scale. In Table 5.18 this difference does not exceed .04 .

3. Within the trimmed tests there are only small differences among the simulated powers of the transformed tests and the corresponding untransformed tests.

Table 5.17 Student's t model df = 2 n = 20 α = .05

θ	$T_n(1)$	$T_n(\hat{\lambda})$	$T_{.1}(1)$	$T_{.1}(\hat{\lambda})$	$T_{.2}(1)$	$T_{.2}(\hat{\lambda})$
0	0.0419	0.0488	0.0443	0.0495	0.0471	0.0456
0.2	0.1176	0.1380	0.1414	0.1528	0.1576	0.1548
0.4	0.2444	0.2908	0.3330	0.3516	0.3556	0.3484
0.6	0.4086	0.4844	0.5632	0.5778	0.5974	0.5814
0.8	0.5692	0.6598	0.7510	0.7660	0.7930	0.7818
1	0.6928	0.7780	0.8758	0.8846	0.9044	0.8954
1.2	0.7858	0.8680	0.9406	0.9456	0.9620	0.9572
1.4	0.8432	0.9146	0.9742	0.9752	0.9862	0.9830
1.6	0.8858	0.9462	0.9884	0.9894	0.9950	0.9942
1.8	0.9130	0.9628	0.9938	0.9934	0.9974	0.9972

Table 5.18 Student's t model df = 3 n = 20 α = .05

θ/σ	$T_n(1)$	$T_n(\hat{\lambda})$	$T_{\cdot 1}(1)$	$T_{\cdot 1}(\hat{\lambda})$	$T_{\cdot 2}(1)$	$T_{\cdot 2}(\hat{\lambda})$
0	0.0445	0.0467	0.0451	0.0476	0.0469	0.0430
0.1	0.1238	0.1338	0.1418	0.1474	0.1466	0.1380
0.2	0.2736	0.2928	0.3170	0.3236	0.3176	0.3032
0.3	0.4652	0.4906	0.5370	0.5456	0.5470	0.5288
0.4	0.6386	0.6774	0.7368	0.7422	0.7530	0.7380
0.5	0.7854	0.8156	0.8752	0.8784	0.8810	0.8696
0.6	0.8728	0.8964	0.9486	0.9486	0.9538	0.9470
0.7	0.9294	0.9452	0.9834	0.9824	0.9854	0.9822
0.8	0.9586	0.9730	0.9954	0.9952	0.9956	0.9944
0.9	0.9742	0.9834	0.9982	0.9984	0.9992	0.9988
1	0.9828	0.9922	0.9992	0.9990	0.9992	0.9992

Tables 5.19 through 5.23 below represent the simulation results for the transformed Student's t model where the inverse of the John-Draper transformation is applied to the t with 2 degrees of freedom data. The value $\lambda = 1/3$ is used in Tables 5.19 , 5.20 and 5.21, $\lambda = 1/2$ in Table 5.22 and $\lambda = 0$ in Table 5.23. From these tables note that:

1. While the simulated levels of the transformed tests appear to be close to .05 , those of the untransformed tests are considerably smaller than .05 especially the simulated level of the t-test in the original scale.

2. The gap between the simulated powers of the transformed t-test and the t-test in the original scale is very wide in all the tables.

3. Within the trimmed tests with proportion 10% , the simulated power of the transformed test is always greater than that of the untransformed test. The maximum difference could be as much as .30 or .35 as Tables 5.22 and 5.23 show.

4. Within the trimmed tests with proportion 20% , in some cases there is not much difference between the simulated powers of the transformed and untransformed tests like in Tables 5.19 and 5.22 . In some other cases the difference could be as much as .10 like in Tables 5.21 and 5.23

Table 5.19 Transformed Student's t model df = 2

$\lambda = 1/3 \quad \sigma = 1 \quad n = 50 \quad \alpha = .05$

θ/σ	$T_n(1)$	$T_n(\hat{\lambda})$	$T_{.1}(1)$	$T_{.1}(\hat{\lambda})$	$T_{.2}(1)$	$T_{.2}(\hat{\lambda})$
0	0.0287	0.0502	0.0419	0.0506	0.0441	0.0475
0.1	0.0430	0.0950	0.0846	0.1050	0.1030	0.1114
0.2	0.0608	0.1580	0.1498	0.1926	0.1936	0.2120
0.3	0.0786	0.2446	0.2364	0.3004	0.3208	0.3408
0.4	0.0988	0.3354	0.3388	0.4256	0.4664	0.4888
0.5	0.1218	0.4298	0.4454	0.5506	0.5988	0.6246
0.6	0.1500	0.5286	0.5556	0.7670	0.6212	0.7448
0.7	0.1784	0.6196	0.6534	0.8678	0.7166	0.8366
0.8	0.2058	0.7054	0.7378	0.8416	0.8880	0.9032
1	0.2656	0.8338	0.8618	0.9346	0.9634	0.9698
1.2	0.3176	0.9068	0.9262	0.9760	0.9890	0.9916

Table 5.20 Transformed Student's t model df = 2

$\lambda = 1/3 \quad \sigma = 2 \quad n = 50 \quad \alpha = .05$

θ/σ	$T_n(1)$	$T_n(\hat{\lambda})$	$T_{-1}(1)$	$T_{-1}(\hat{\lambda})$	$T_{-2}(1)$	$T_{-2}(\hat{\lambda})$
0	0.0247	0.0500	0.0364	0.0506	0.0400	0.0477
0.1	0.0310	0.0846	0.0638	0.0930	0.0824	0.0994
0.2	0.0370	0.1280	0.1010	0.1570	0.1444	0.1708
0.3	0.0444	0.1886	0.1476	0.2336	0.2236	0.2674
0.4	0.0542	0.2546	0.1994	0.3198	0.3270	0.3852
0.5	0.0616	0.3228	0.2670	0.4198	0.4336	0.4960
0.6	0.0686	0.3888	0.3354	0.5160	0.5422	0.5992
0.7	0.0804	0.4600	0.4032	0.6032	0.6326	0.6940
0.8	0.0910	0.5274	0.4738	0.6854	0.7216	0.7756
1	0.1144	0.6478	0.6088	0.8118	0.8444	0.8838
1.2	0.1406	0.7452	0.7132	0.9900	0.8252	0.9504
1.4	0.1660	0.8220	0.8002	0.9418	0.9620	0.9736
1.6	0.1878	0.8750	0.8578	0.9679	0.9790	0.9887
1.8	0.2188	0.5109	0.9000	0.9822	0.9896	0.9956

Table 5.21 Transformed Student's t model df = 2

$\lambda = 1/3 \quad \sigma = 3 \quad n = 50 \quad \alpha = .05$

θ/σ	$T_n(1)$	$T_n(\hat{\lambda})$	$T_{.1}(1)$	$T_{.1}(\hat{\lambda})$	$T_{.2}(1)$	$T_{.2}(\hat{\lambda})$
0	0.0233	0.0502	0.0331	0.0502	0.0362	0.0477
0.1	0.0258	0.0804	0.0512	0.0858	0.0696	0.0904
0.2	0.0294	0.1104	0.0742	0.1364	0.1088	0.1476
0.3	0.0330	0.156	0.1026	0.1970	0.1708	0.2262
0.4	0.0382	0.2082	0.1356	0.2634	0.2370	0.3122
0.5	0.0418	0.2612	0.1702	0.3358	0.3174	0.4106
0.6	0.0468	0.3142	0.2132	0.4136	0.4018	0.5006
0.7	0.0514	0.3644	0.2624	0.4898	0.4846	0.5806
0.8	0.0554	0.4144	0.3126	0.5606	0.5608	0.6546
1	0.0660	0.5164	0.4076	0.6826	0.6984	0.7828
1.2	0.0792	0.6026	0.5026	0.7812	0.8012	0.8658
1.4	0.0912	0.6820	0.5940	0.8498	0.8698	0.9258
1.6	0.1060	0.5022	0.6712	0.9030	0.9240	0.9584
1.8	0.1186	0.7980	0.7334	0.9356	0.9518	0.9732
2	0.1332	0.8408	0.7872	0.9580	0.9688	0.9852
2.2	0.1500	0.8682	0.8298	0.9716	0.9790	0.9926

Table 5.22 Transformed Student's t model df = 2

$\lambda = 1/2$ $\sigma = 2$ n = 20 $\alpha = .05$

θ/σ	$T_n(1)$	$T_n(\hat{\lambda})$	$T_{.1}(1)$	$T_{.1}(\hat{\lambda})$	$T_{.2}(1)$	$T_{.2}(\hat{\lambda})$
0	0.0293	0.0488	0.0359	0.0495	0.0381	0.0456
0.2	0.0554	0.0978	0.0814	0.1164	0.1020	0.1194
0.4	0.0892	0.1726	0.1672	0.2184	0.2256	0.2484
0.6	0.1394	0.2642	0.2778	0.3472	0.3768	0.3976
0.8	0.1874	0.3540	0.4012	0.4838	0.5220	0.5386
1	0.2454	0.4472	0.5270	0.5974	0.6622	0.6654
1.2	0.3022	0.5308	0.6224	0.6954	0.7714	0.7700
1.4	0.3604	0.5984	0.7086	0.7710	0.8422	0.8384
1.6	0.4120	0.6536	0.7720	0.8272	0.8920	0.8876
1.8	0.4518	0.7022	0.8250	0.8680	0.9238	0.9200
2	0.4958	0.7428	0.8668	0.9074	0.9508	0.9470
0.2	0.5338	0.7730	0.8976	0.9274	0.9654	0.9614

Table 5.23 Transformed Student's t model df = 2

$\lambda = 0$ $\sigma = 2$ $n = 20$ $\alpha = .05$

θ/σ	$T_n(1)$	$T_n(\hat{\lambda})$	$T_{.1}(1)$	$T_{.1}(\hat{\lambda})$	$T_{.2}(1)$	$T_{.2}(\hat{\lambda})$
0.2	0.0070	0.0784	0.0182	0.0838	0.0436	0.0932
0.4	0.0086	0.1036	0.0260	0.1308	0.0728	0.1624
0.6	0.0094	0.1356	0.0442	0.1848	0.1342	0.2498
0.8	0.0114	0.1736	0.0646	0.2454	0.1988	0.3240
1	0.0122	0.2092	0.0904	0.2998	0.2734	0.3996
1.2	0.0148	0.2424	0.1184	0.3548	0.3406	0.4640
1.4	0.0164	0.2736	0.1442	0.4054	0.4020	0.5200
1.6	0.0188	0.2968	0.1724	0.4472	0.4566	0.5674
1.8	0.0216	0.3248	0.2002	0.4918	0.5068	0.6116
2	0.0244	0.3516	0.2278	0.5272	0.5548	0.6456
2.2	0.0270	0.3770	0.2544	0.5596	0.5914	0.6818

5.5 Simulation Results for Skewed Models

As pointed out earlier, the distribution of the t-statistic is not symmetric about 0 when the distribution of the data is skewed. In this section, besides reporting the results of the simulated level and power of the different test statistics, we give some examples of the frequency and cumulative distributions of the t-statistic in the original scale and the t-statistic after transformation.

5.5.1 Transformed-normal model

The data used in the runs under this model are generated from a standard normal variable by applying the inverse of the two-domain transformation for some λ_1 and λ_2 . We fix $\lambda_1 = 1/4$ and consider runs for $\lambda_2 = 1/2$, 3/4 and 1 . Table 5.24 below gives the frequency and cumulative distributions of the t-statistic calculated from the original and the transformed data, for $\lambda_1 = 1/4$ and $\lambda_2 = 3/4$ and Figure 5.1 gives the shape of the frequency distributions.

Table 5.24 Frequency and cumulative distributions of the t-statistic under the transformed-normal model $\lambda_1 = 1/4$ $\lambda_2 = 3/4$

values of T	transformed freq.	transformed cum.	original freq.	original cum.
$(-\infty, -4)$	0	0	0.0002	0.0002
(-4,-3.5)	0	0	0.0009	0.0011
(-3.5,-3)	0.0002	0.0002	0.0030	0.0042
(-3,-2.5)	0.002	0.0022	0.0124	0.0166
(-2.5,-2)	0.0091	0.0113	0.0397	0.0563
(-2,-1.5)	0.0313	0.0426	0.0948	0.1511
(-1.5,-1)	0.0777	0.1203	0.1676	0.3187
(-1,-.5)	0.1538	0.2741	0.2029	0.5216
(-.5, 0)	0.2254	0.4995	0.1855	0.7071
(0 , .5)	0.2241	0.7236	0.1346	0.8417
(.5 , 1)	0.1563	0.8789	0.0811	0.9229
(1 , 1.5)	0.0793	0.9592	0.0418	0.9647
(1.5 , 2)	0.0301	0.9893	0.0207	0.9854
(2 , 2.5)	0.0087	0.9980	0.0093	0.9947
(2.5 , 3)	0.0017	0.9997	0.0034	0.9981
(3 , 3.5)	0.0003	1	0.0012	0.9993
(3.5 , 4)	0	1	0.0005	0.9998
$(4, \infty)$	0	1	0.0002	1

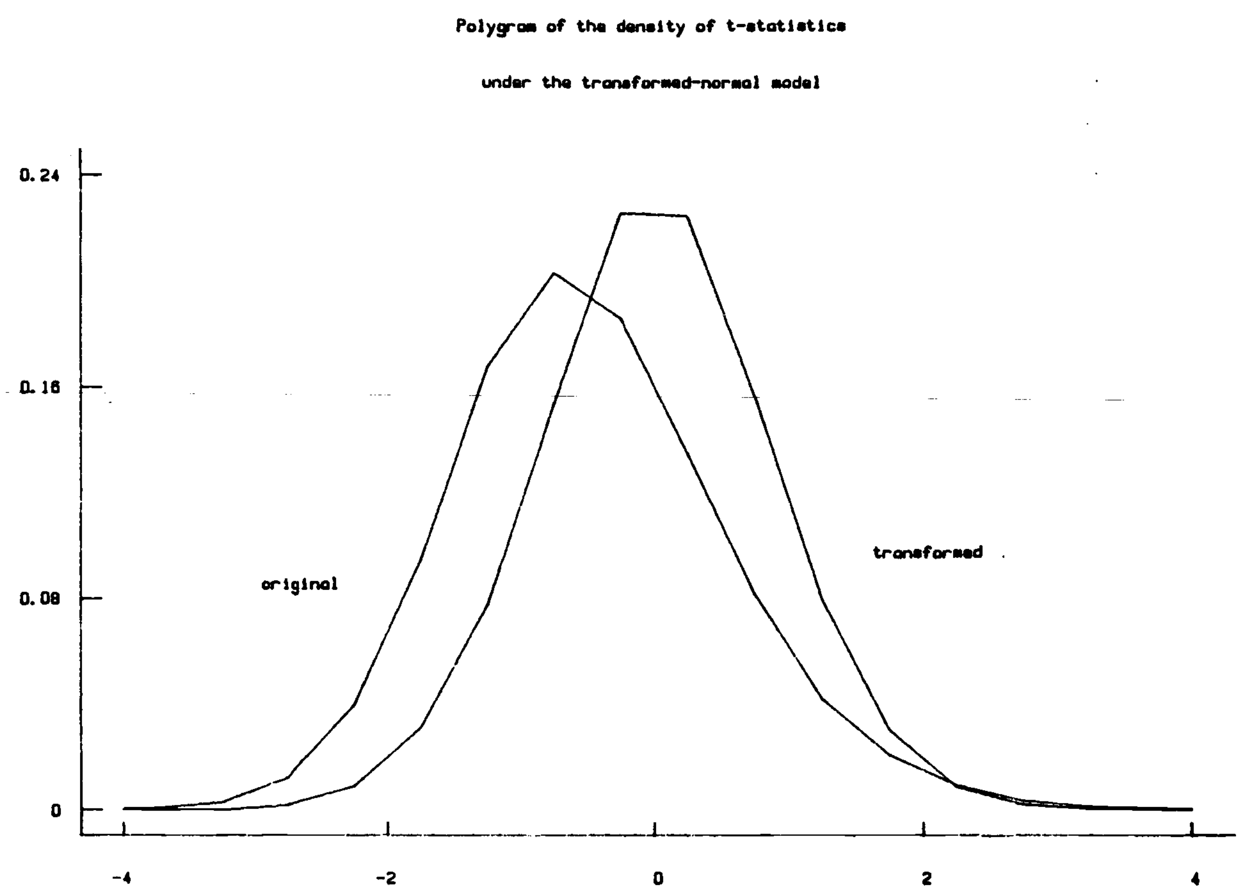
Polygram of the density of t-statistics
under the transformed-normal model
0.24
0.16
0.08
0
-4
-2
0
2
4
original
transformed

Figure 5.1

From the above table and figure note the following:

1. The distribution of the t-statistic in the original scale is not symmetric about 0 . The median from Table 5.24 is about -.55 .

2. The distribution of the transformed t-statistic is fairly symmetric about 0.

3. The distribution of the transformed t-statistic has shorter tails than those of a Student's t-distribution with 19 degrees of freedom. For example, while $Pr(t(19) < -2.5) = .0109$, the corresponding simulated probability is only .0022 , and while $Pr(t(19) < -2) = .0300$, the corresponding simulated probability is .0113 . On the upper tail, the simulated probabilities of $t > 2$ and $t > 2.5$ are .0107 and .0020 respectively.

Tables 5.25 through 5.27 below give the simulated level and power of the t-test in the original scale, the transformed t-test and the transformed-trimmed t-test. From these tables note that:

1. The simulated levels of the untrimmed tests are considerably smaller than .05 . The simulated level of the transformed-trimmed test with proportion 20% is close to the nominal level.

2. the simulated power of the transformed t-test exceeds that of the t-test in the original scale. The difference between the two simulated powers could be as much as .084 in Table 5.25, .075 in Table 5.26 and .12 in Table 5.27.

3. The transformed-trimmed test statistics have their simulated powers considerably greater than that of the t-test in the original scale and also greater than that of the transformed test without trimming.

Table 5.25 Transformed-normal model $\sigma^2 = 1$ n = 20

$\lambda_1 = .25$ $\lambda_2 = .5$

θ/σ	$T_n(1)$	$T_n(\hat{\lambda})$	$T_{.1}(\hat{\lambda})$	$T_{.2}(\hat{\lambda})$
0	0.0332	0.0339	0.0417	0.0489
0.2	0.2254	0.2730	0.3322	0.3532
0.3	0.4014	0.4766	0.5494	0.5772
0.4	0.5842	0.6682	0.7388	0.7574
0.5	0.7406	0.8098	0.8594	0.8806
0.6	0.8506	0.9014	0.9284	0.9426
0.7	0.9242	0.9494	0.9668	0.9762
0.8	0.9618	0.9800	0.9878	0.9894

Table 5.26 Transformed-normal model $\sigma^2 = 1$ n = 20

$\lambda_1 = .25$ $\lambda_2 = .75$

θ/σ	$T_n(1)$	$T_n(\hat{\lambda})$	$T_{.1}(\hat{\lambda})$	$T_{.2}(\hat{\lambda})$
0	0.0235	0.0240	0.0421	0.0492
0.2	0.1988	0.2364	0.3156	0.3528
0.3	0.3738	0.4366	0.5312	0.5752
0.4	0.5608	0.6374	0.7258	0.7588
0.5	0.7214	0.7876	0.8490	0.8804
0.6	0.8398	0.8854	0.9226	0.9438
0.7	0.9170	0.9430	0.9622	0.9756
0.8	0.9584	0.9744	0.9842	0.9870

Table 5.27 Transformed-normal model $\sigma^2 = 1$ n = 20

$\lambda_1 = .25 \quad \lambda_2 = 1$

θ/σ	$T_n(1)$	$T_n(\hat{\lambda})$	$T_{\cdot 1}(\hat{\lambda})$	$T_{\cdot 2}(\hat{\lambda})$
0	0.0181	0.0242	0.0424	0.0495
0.2	0.1766	0.2460	0.3286	0.3610
0.3	0.3488	0.4534	0.5498	0.5842
0.4	0.5386	0.6534	0.7422	0.7652
0.5	0.7046	0.8000	0.8592	0.8840
0.6	0.8308	0.8952	0.9292	0.9472
0.7	0.9078	0.9498	0.9664	0.9772

5.5.2 Gamma model

In this subsection we present the simulation results for the Gamma models with shape parameters 3 , 4 and 5 . Table 5.28 and Figure 5.2 below give the frequency distribution of the t-statistic under the original and transformed scales. Note that:

1. The distribution of the t-statistic is not symmetric about 0 . The median from Table 5.28 is about .65 .

2. The median of the distribution of the t-statistic from the transformed data is about -.06 .

3. The frequency distribution of the transformed t-test appears to be much more symmetric than that obtained from the data in the original scale. However, the tails appear to be shorter than those of

Table 5.28 Frequency and cumulative distributions of the t-statistic under the Gamma model with shape parameter 4

values of T	transformed		original	
	freq.	cum.	freq.	cum.
$(-\infty,-4)$	0	0	0.0002	0.0002
(-4,-3.5)	0	0	0.0007	0.0009
(-3.5,-3)	0.0003	0.0003	0.0009	0.0019
(-3,-2.5)	0.0029	0.0032	0.0025	0.0043
(-2.5,-2)	0.0112	0.0144	0.0060	0.0103
(-2,-1.5)	0.0373	0.0517	0.0151	0.0254
(-1.5,-1)	0.0879	0.1396	0.0323	0.0578
(-1,-.5)	0.1683	0.3079	0.0636	0.1214
(-.5 , 0)	0.2291	0.5370	0.1129	0.2343
(0 , .5)	0.2135	0.7505	0.1747	0.4090
(.5 , 1)	0.1437	0.8942	0.2105	0.6195
(1 , 1.5)	0.0681	0.9623	0.1840	0.8035
(1.5 , 2)	0.0259	0.9882	0.1149	0.9184
(2 , 2.5)	0.0087	0.9969	0.0541	0.9724
(2.5 , 3)	0.0029	0.9998	0.0195	0.9919
(3 , 3.5)	0.0002	1	0.0063	0.9982
(3.5 , 4)	0	1	0.0014	0.9996
$(4 , \infty)$	0	1	0.0004	1

Polygram of the density of t-statistics

under Gama model with shape parameter 4

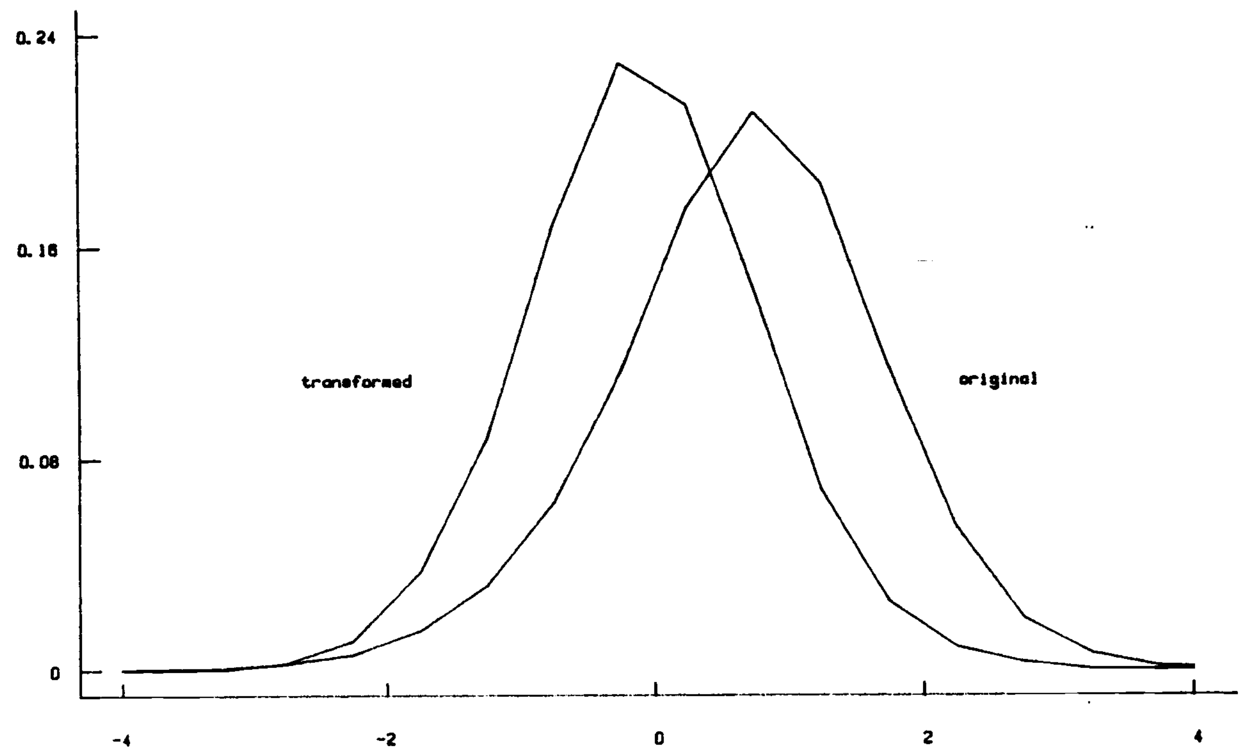

Figure 5.2

a student's t distribution with 19 degrees of freedom.

In Tables 3.29 through 3.31 we give the simulated power of the different test statistics for the three Gamma models mentioned above. We subtract the true median from each observation so that the null hypothesis is to test $\theta = 0$. From these tables note that:

1. The simulated levels of the transformed-trimmed tests are somewhat over the nominal level in two of the tables especially when 20% from the observations on each tail are trimmed.

2. The transformed t-test has greater simulated power than that of the t-test in the original scale for alternatives that are close to the true model. The difference in the simulated powers decreases as the value of the shape parameter increases. Under alternatives that are away from the true model the t-test in the original scale becomes slightly better than the transformed t-test.

3. The test statistics based on the trimmed-transformed samples have considerably greater simulated powers than the untrimmed tests. There is not much difference in the simulated power between the two percentage of trimming.

Table 5.29 Gamma model with shape parameter 3

median = 2.67406 n = 20 α = .05

θ/σ	$T_n(1)$	$T_n(\hat{\lambda})$	$T_{\cdot 1}(\hat{\lambda})$	$T_{\cdot 2}(\hat{\lambda})$
0	0.0132	0.0322	0.0544	0.0571
0.2	0.1640	0.2182	0.2964	0.3018
0.3	0.3314	0.3932	0.4900	0.4936
0.4	0.5188	0.5824	0.6730	0.6774
0.5	0.7036	0.7404	0.8084	0.8112
0.6	0.8594	0.8668	0.9154	0.9198
0.7	0.9296	0.9360	0.9560	0.9584

Table 5.30 Gamma model with shape parameter 4

median = 3.67206 n = 20 α = .05

θ/σ	$T_n(1)$	$T_n(\hat{\lambda})$	$T_{\cdot 1}(\hat{\lambda})$	$T_{\cdot 2}(\hat{\lambda})$
0	0.0177	0.0226	0.0377	0.0469
0.2	0.1792	0.2205	0.3153	0.4134
0.3	0.3500	0.3846	0.4796	0.4844
0.4	0.5532	0.5660	0.6602	0.6638
0.5	0.7332	0.7336	0.8042	0.8104
0.6	0.8682	0.8512	0.8972	0.9002
0.7	0.9422	0.9326	0.9570	0.9582

table 5.31 Gamma model with shape parameter 5

median = 4.670913 n = 20 α = .05

θ/σ	$T_n(1)$	$T_n(\hat{\lambda})$	$T_{\cdot 1}(\hat{\lambda})$	$T_{\cdot 2}(\hat{\lambda})$
0	0.0195	0.0298	0.0517	0.0551
0.2	0.1898	0.2126	0.2856	0.2946
0.3	0.3688	0.3986	0.4870	0.4864
0.4	0.5750	0.5880	0.6692	0.6750
0.5	0.7566	0.7562	0.8124	0.8144
0.6	0.8806	0.8586	0.9010	0.9026
0.7	0.9534	0.9322	0.9548	0.9606
0.8	0.9826	0.9714	0.9844	0.9848

5.3.3 Extreme-Value model

Table 5.32 below gives the frequency and cumulative distributions of the t-test in the original scale and the transformed t-test under the extreme-value model. Figure 5.3 gives the shape of the frequency distributions of the above two tests. We can conclude the following:

1. The distribution of the t-statistic in the original scale is not symmetric about 0 . Its median is about -.7 .

2. The frequency distribution of the transformed t-statistic indicates that the distribution of the frequencies is not far from symmetry about 0 . The median is approximately .0558 .

Table 5.32 Frequency and cumulative distributions of the t-statistic under the Extreme-Value model

value of T	transformed freq.	transformed cum.	original freq.	original cum.
$(-\infty,-4)$	0	0	0.0003	0.0003
(-4,-3.5)	0	0	0.0014	0.0017
(-3.5,-3)	0.0001	0.0001	0.0051	0.0068
(-3,-2.5)	0.0010	0.0011	0.0185	0.0253
(-2.5,-2)	0.0053	0.0064	0.0515	0.0768
(-2,-1.5)	0.0268	0.0332	0.1176	0.1944
(-1.5,-1)	0.0742	0.1074	0.1849	0.3794
(-1,-.5)	0.1469	0.2543	0.2113	0.5907
(-.5, 0)	0.2203	0.4746	0.1731	0.7638
(0 , .5)	0.2268	0.7014	0.114	0.8778
(.5 , 1)	0.1658	0.8672	0.0662	0.944
(1 , 1.5)	0.0913	0.9585	0.0319	0.9759
(1.5 , 2)	0.0331	0.9916	0.0131	0.9891
(2 , 2.5)	0.0073	0.9989	0.007	0.996
(2.5 , 3)	0.0009	0.9998	0.0025	0.9986
(3 , 3.5)	0.0002	1	0.0008	0.9993
(3.5 , 4)	0	1	0.0002	0.9995
$(4 , \infty)$	0	1	0.0005	1

Polygram of the density of t-statistics

under the extreme-value model

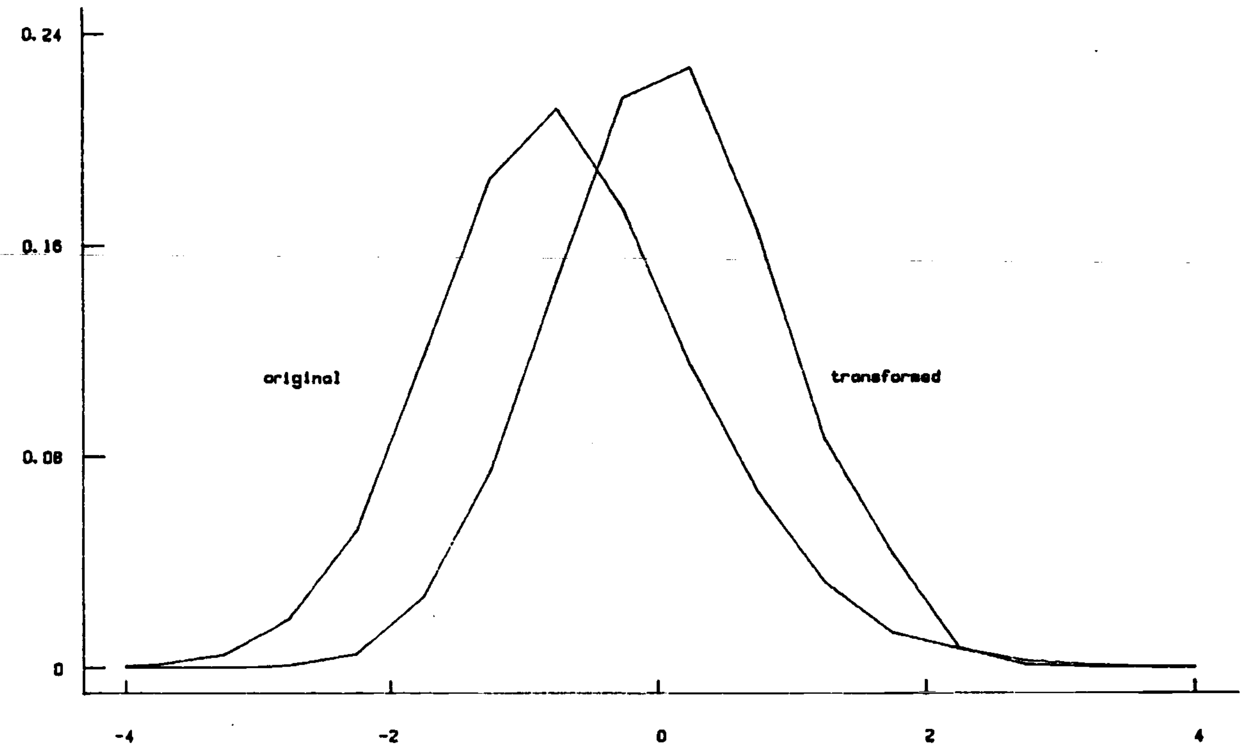

Figure 5.3

Table 5.33 gives the simulated power and level of the test statistics under the extreme value model. This table shows that the same conclusions we made under the transformed-normal and Gamma models hold for the present model. The transformed t-test is more powerful than the t-test in the original scale. The trimmed test statistics are considerably better than the untrimmed tests and the simulated power is not affected by the percentage of trimming.

Table 5.33 Extreme-Value median = ln(ln(2)) n = 20

θ/σ	$T_n(1)$	$T_n(\hat{\lambda})$	$T_{\cdot 1}(\hat{\lambda})$	$T_{\cdot 2}(\hat{\lambda})$
0	0.0168	0.0224	0.0390	0.0478
0.2	0.0996	0.1398	0.2020	0.2116
0.3	0.1908	0.2432	0.3246	0.3348
0.4	0.3120	0.3620	0.4702	0.4772
0.5	0.4600	0.5160	0.6160	0.6240
0.6	0.6104	0.6514	0.7286	0.7414
0.7	0.7350	0.7558	0.8194	0.8316
0.8	0.8262	0.8448	0.8926	0.8984
0.9	0.8970	0.9060	0.9404	0.9432

BIBLIOGRAPHY

Andrews, D. C. (1971). A note on the selection of data transformations. Biometrika 58, 249-254.

Atkinson, A. C. (1983). Diagnostic regression analysis and shifted power transformation. Technometrics 25, 23-33.

__________ (1985). Plots, transformations and regression. Clarendon Press. Oxford.

Bickel, P. and Doksum, K. (1981). An analysis of transformations revisited, J. Amer. Statist. Assoc. 76, 296-311.

Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations. J. R. Statist. Soc. B26, 211-246.

__________ (1982). An analysis of transformations revisited, rebuted. J. Amer. Statist. Assoc. 79, 209-210.

Carroll R. J. and Ruppert, D. (1984). Power transformations when fitting theoretical models to data. J. Amer. Statist. Assoc. 79, 312-313.

__________ (1988). Transformations and weighting in regression. Chapman and Hall. New York.

Doksum, K. and Wong C. W. (1980). Statistical tests after transformation. Tech. Report.

__________ (1983). Statistical tests based on transformed data. J. Amer. Statist. Assoc. 78, 411-417.

Freedman, D.A. and Stephne, C.P. (1984). Bootstrapping a regression equation: some empirical results. J. Amer. Statist. Assoc. 79, 97-106.

Geary, R. C. (1947). Testing for normality. Biometrika 34, 209-242.

Hinkley, D. V. (1975). On power transformations to symmetry. Biometrika 62, 101-111.

__________ (1977). On quick choice of power transformations. Appl. Statist. 26, 67-69.

Hinkley, D. V. and Runger, G. (1984). The analysis of transformed data. J. Amer. Statist. Assoc. 79, 302-308.

Hoaglin, D. C. , Mostller, F. and Tukey, J. W. (1983). Understanding robust and exploratory data analysis. John Wiley, New York.

Huber, P. J. (1970). Studentizing robust estimates. In Non-Parametric Techniques in Statistical Inference (M.L. Puri, ed.) Cambridge: Cambridge University Press.

__________ (1967). Behavior of maximum likelihood estimates under nonstandard conditions. Proceedings of the Fifth Berkely Symposium on mathematical statistics and probability. 221-233.

John, J. A. and Draper, N. R. (1980). An alternative family of transformations. Appl. Statist. 29, 190-197.

Laha, R. G. and Rohatgi, V. K. (1979). Probability theory. John Wiley & Sons, New York.

Lehmann, E. L. (1983). Theory of point estimation. John Wiley, New York.

__________ (1986). Testing statistical hypotheses. John Wiley, New York.

Quesenberry, C. P. and Quesenberry, C. J. (1982). On the distribution of residuals from fitted parametric models. J. Statist. Comp. and Sim., 15, 129-140.

Rao, C. R. (1973). Linear statistical inference and its applications. John Wiley & Sons, New York.

Serfling, R. J. (1980). Approximation theorems of mathematical statistics. John Wiley & Sons, New York.

Smith, K. T. (1971). Primer of modern analysis. Bogden & Quigley, Inc. Publishers.

Stigler, S. M. (1974). The asymptotic distribution of the trimmed mean. Anns. Statist. , 472-477.

Stukle, T. A. (1988). Generalized logistic models. J. Amer. Statist. Assoc. 83, 426-431.

Tan, W. Y. (1982). Sampling distributions and robustness of t, F and variance ratio in two samples and anova models with respect to departure from normality. Commun. Statist. Theor. Meth. 22, 2485-2511.

Taylor, J. M. G. (1986). The retransformed mean after fitted power transformation. J. Amer. Statist. Assoc. 81, 114-118.

Tiku, M. L. (1980). Robustness of MMl estimators based on censored samples and robust test statistics. J Statist. Plann. Inf. 4, 123-143.

Tiku, M. L., Tan, W. Y. and Balakrishman, N. (1986). Robust inference Marcel Dekker Inc., New York.

Van Zwet, W. R. (1964). Convex transformations of random variables. Amsterdam, Mathematisch Centrum.

Weisberg, S. (1985). Applied linear regression. John Wiley, New York.

Yuen, K. K. and Murthy, V. K. (1974). Percentage points of the distribution of the t statistic when the parent is student's t. Technometrics 16, 495-498.

APPENDICES

APPENDIX A

Some Asymptotic Results

In this appendix we present some facts and theorems about convergence in probability and in law which are frequently used in the proofs given in Chapters 3 and 4. Proofs of the stated facts can be found in Rao (1973, p.122-124).

Let $\{A_n , B_n\}$, $n = 1,2,...$ be a sequence of pairs of random variables.

<u>Fact 1</u> If $A_n - B_n \xrightarrow{p} 0$ and if $B_n \xrightarrow{\mathcal{L}} B$ then

$$A_n \xrightarrow{\mathcal{L}} B$$

<u>Fact 2</u> If $A_n \xrightarrow{p} A$ and $B_n \xrightarrow{p} b$ then

i. $A_nB_n \xrightarrow{p} Ab$

ii. $A_n/B_n \xrightarrow{p} A/b \qquad (b \neq 0)$

iii. $A_n + B_n \xrightarrow{p} A + b$

<u>Fact 3</u>

i. $A_n \xrightarrow{p} A$ imply $A_n \xrightarrow{p} A$

ii. $(A_n \xrightarrow{p} A$ iff $A_n \xrightarrow{p} A)$ iff A is constant

<u>Fact 4</u> Let g be a continuous function. If $A_n - B_n \xrightarrow{p} 0$ and if $B_n \xrightarrow{p} B$ then

i. $g(B_n) \xrightarrow{p} g(B) \qquad$ ii. $g(A_n) \xrightarrow{p} g(B)$

<u>Lemma A.1</u> If $B_n \xrightarrow{p} b$ then B_n is bounded in probability.

<u>Proof</u> Since $B_n \xrightarrow{p} b$ then given $\delta > 0$ there exists n such that

$$\Pr\{ |B_n - b| \leq 1 \} \geq 1-\delta .$$

If $b \geq 0$ let $M = b+1$ otherwise, let $M = |b - 1|$. It follows that

$$Pr\{|B_n| \leq M\} \geq 1-\delta \ . \qquad []$$

<u>Lemma A.2</u> If $Pr\{A_n\} \longrightarrow 1$ and if $Pr\{B_n\} \longrightarrow 1$ then

$$Pr\{A_n \cap B_n\} \longrightarrow 1 \ .$$

<u>Proof</u> $Pr\{A_n \cap B_n\} = Pr\{A_n\} - Pr\{A_n \cap B_n^c\} \geq Pr\{A_n\} - Pr\{B_n^c\} \longrightarrow 1$. []

<u>Lemma A.3</u> If $A_n B_n$ is bounded in probability and $B_n \xrightarrow{p} b$, then A_n is bounded in probability.

<u>Proof</u> $A_n B_n$ is bounded in probability implies that for some M' $Pr\{|A_n B_n| \leq M'\} \longrightarrow 1$. $B_n \xrightarrow{p} b$ implies $Pr\{|B_n| \leq b/2\} \longrightarrow 1$. Therefore

$$\begin{aligned} Pr\{|A_n| \leq M\} &\geq Pr\{|A_n B_n| \leq M' \text{ and } |B_n| \leq b/2\} \\ &= Pr\{|A_n| \leq M'/|B_n| \text{ and } |B_n| \leq b/2\} \\ &> Pr\{|A_n| \leq 2M'/|b| \text{ and } |B_n| \leq b/2\} \longrightarrow 1 \ . \qquad [] \end{aligned}$$

<u>Theorem A.1</u> If $A_n \xrightarrow{p} 0$ and if for some $M > 0$

$$Pr\{|B_n| \leq M\} \xrightarrow{p} 1 \text{ then}$$

$$A_n B_n \xrightarrow{p} 0 \ .$$

<u>Proof:</u> Since $A_n \xrightarrow{p} 0$, given ϵ and δ there exists n_1 such that $Pr\{|A_n| \leq \epsilon/M\} > 1-\delta/2$ for all $n \geq n_1$. Since

$$Pr\{|B_n| \leq M\} \longrightarrow 1 \text{ there exists } n_2 \text{ such that}$$

$$Pr\{|B_n| \leq M\} \geq 1-\delta/2 \qquad \text{for all } n \geq n_2 \ .$$

Let $n_0 = \max(n_1, n_2)$ then for all $n \geq n_0$

$$Pr\{|A_n| \leq M\} > 1-\delta/2 \text{ and } Pr\{|B_n| \leq M\} \geq 1-\delta/2 \ .$$

Since $|A_n| \leq \epsilon/M$ and $|B_n| \leq M$ imply $|A_n B_n| \leq \epsilon$, therefore

$$\begin{aligned} Pr\{|A_n B_n| \leq \epsilon\} &\geq Pr\{|A_n| \leq \epsilon/M , |B_n| \leq M\} \\ &= 1 - Pr\{|A_n| \leq \epsilon/M , |B_n| \leq M\}^c \\ &\geq Pr\{|A_n| \leq \epsilon/M\} + Pr\{|B_n| \leq M\} - 1 \\ &> 1-\delta/2 + 1 - \delta/2 - 1 = 1 - \delta \qquad [] \end{aligned}$$

Theorem A.2 Let $T_{1n} = A_{1n}/B_{1n}$ and $T_{2n} = A_{2n}/B_{2n}$. Suppose that $B_{1n} - B_{2n} \xrightarrow{P} 0$ and $B_{2n} \xrightarrow{P} \beta$ (B_{1n} , B_{2n} and β are positive for all n). If $A_{1n} - A_{2n} \xrightarrow{P} 0$ and $A_{1n} \xrightarrow{\mathcal{L}} A$ then

$$T_{1n} - T_{2n} \xrightarrow{\mathcal{L}} 0 .$$

Proof:

$$\begin{aligned} T_{1n} - T_{2n} &= A_{1n}/B_{1n} - A_{2n}/B_{2n} \\ &= (1/B_{1n} - 1/B_{2n})\, A_{1n} + 1/B_{2n}\,(A_{1n} - A_{2n}) \\ &\xrightarrow{\mathcal{L}} (1/\beta - 1/\beta)\, A + 1/\beta\,(0) = 0 . \quad [] \end{aligned}$$

Theorem A.3 Suppose T_{1n} is a test statistic that is Pitman regular (Appendix B) with $\Psi_n(\mu) = E_\mu(T_{1n})$ and $\tau_n^2(\mu) = \mathrm{var}_\mu(T_{1n})$. Suppose further that $\tau_n(\mu) \xrightarrow{P} \tau(\mu) > 0$. If T_{2n} is another test statistic of the same hypothesis then

i. T_{2n} is Pitman regular ii. ARE (T_{1n} , T_{2n}) = 1 .

Proof:

For T_{2n} choose the same $\Psi_n(\mu)$ and $\tau_n^2(\mu)$ as those for T_{1n} then, C3 through C7 of the regularity conditions of Appendix B are satisfied by the assumption that T_{1n} is Pitman regular and the choice of $\Psi_n(\mu)$ and $\tau_n^2(\mu)$. It remains to check C1 and C2 . Since

$$(T_{2n} - \Psi_n(\mu))/\tau_n(\mu) = (T_{1n} - \Psi_n(\mu))/\tau_n(\mu) - (T_{1n} - T_{2n})/\tau_n(\mu)$$

hence

$$(T_{2n} - \Psi_n(\mu))/\tau_n(\mu) - (T_{1n} - \Psi_n(\mu))/\tau_n(\mu) = (T_{1n} - T_{2n})/\tau_n(\mu)$$

By assumption

$$\tau_n(\mu) \xrightarrow{P} \tau(\mu) > 0 \quad \text{and} \quad T_{1n} - T_{2n} \xrightarrow{P} 0$$

hence

$$(T_{2n} - \Psi_n(\mu))/\tau_n(\mu) - (T_{1n} - \Psi_n(\mu))/\tau_n(\mu) \xrightarrow{P} 0 .$$

T_{1n} is Pitman regular implies that it satisfies condition C1 of

Appendix B ,that is

$$(T_{1n} - \Psi_n(\mu_0))/\tau_n(\mu_0) \xrightarrow{\mathcal{L}} N(0,1)$$

By fact 1 above we conclude that

$$(T_{2n} - \Psi_n(\mu_0))/\tau_n(\mu_0) \xrightarrow{\mathcal{L}} N(0,1) .$$

Hence T_{2n} satisfies condition C1 . Similar argument holds for condition C2. Therefore T_{2n} is Pitman regular.

Since the asymptotic relative efficiency as shown in Appendix B depends only on

$$R_n^2 = (\Psi_n'(\mu_0)/\tau_n(\mu_0))^2$$

and since R_n^2 is the same for both test statistics hence, we conclude

$$\text{ARE}(T_{1n} , T_{2n}) = 1 . \quad []$$

APPENDIX B

Pitman Asymptotic Relative Efficiency

Let $Y_1, Y_2, \ldots, Y_n, \ldots$ be a sequence of independent and identically distributed random variables with distribution P_θ, $\theta \in \Omega$, an interval of $\mathbb{R}^1$. Let $T_n = T_n(Y_1, \ldots, Y_n)$ be a test statistic for testing $H_0\colon \theta = \theta_0$ versus $H_1\colon \theta > \theta_0$.

The sequence $\{T_n\}$ is called Pitman-regular for testing the above hypotheses if there exist functions $\Psi_n(\theta)$ and $\tau_n(\theta)$ satisfying the conditions C1-C7 below.

Let Z be a standard normal random variable, and let

$$W_{0n}^{(0)} = \frac{T_n - \Psi_n(\theta_0)}{\tau_n(\theta_0)} \text{ under } P_{\theta_0} \quad \text{and} \quad W_{0n}^{(n)} = \frac{T_n - \Psi_n(\theta_n)}{\tau_n(\theta_n)} \text{ under } P_{\theta_n}$$

where $\theta_n = \theta_0 + k/\sqrt{n}$ for some $k > 0$.

C1. $W_{0n}^{(0)} \xrightarrow{\mathcal{L}} Z$ as $n \longrightarrow \infty$.

C2. $W_{1n}^{(n)} \xrightarrow{\mathcal{L}} Z$ as $n \longrightarrow \infty$.

C3. $\Psi_n(\theta)$ is differentiable with respect to θ in an open interval containing $[\theta_0, \theta_n]$.

C4. $\Psi_n'(\theta_0) > 0$.

C5. $\dfrac{\Psi_n'(\theta_0)}{\sqrt{n}\, \tau_n(\theta_0)} \longrightarrow c$ as $n \longrightarrow \infty$ for some $c > 0$.

C6. $\sup\limits_{\theta_0 \le \bar{\theta} \le \theta_n} \left| \dfrac{\Psi_n'(\bar{\theta})}{\Psi_n'(\theta_0)} - 1 \right| \longrightarrow 0$ as $n \longrightarrow \infty$

C7. $\dfrac{\tau_n(\theta_n)}{\tau_n(\theta_0)} \longrightarrow 1$ as $n \longrightarrow \infty$.

Let $0 < \alpha < 1$ and let $Z(\alpha)$ be such that

$$\Pr[Z > Z(\alpha)] = 1 - \Phi(Z(\alpha)) = \alpha .$$

Let $W_{0n} = \dfrac{T_n - \Psi_n(\theta_0)}{\tau_n(\theta_0)}$ with no distribution specified. Define a rejection region R_n as

$$R_n = \{ W_{0n} > Z(\alpha) \} .$$

The level of R_n is

$$\alpha_n = P_{\theta_0}[R_n] = P_{\theta_0}[W_{0n} > Z(\alpha)] = P[W_{0n}^{(0)} > Z(\alpha)] ,$$

and the power of R_n is

$$\Pi_n = P_{\theta_n}[R_n] = P_{\theta_n}[W_{0n} > Z(\alpha)] = P[W_{0n}^{(n)} > Z(\alpha)]$$

where $W_{0n}^{(0)}$ denotes W_{0n} under P_{θ_0} (as above) and $W_{0n}^{(n)}$ denotes W_{0n} under P_{θ_n} .

<u>Lemma B.1</u> $\alpha_n \longrightarrow \alpha$ as $n \longrightarrow \infty$.

<u>Proof:</u> By C1, $W_{0n}^{(0)} \xrightarrow{\mathcal{L}} Z$, so

$$\alpha_n = P[W_{0n}^{(0)} > Z(\alpha)] \longrightarrow P[Z > Z(\alpha)] = \alpha . \quad []$$

<u>Lemma B.2</u> $\Pi_n \longrightarrow \Pi$ as $n \longrightarrow \infty$ where $\Pi = P[Z > Z(\alpha) - kc]$.

<u>Proof:</u> $\Pi_n = P[W_{0n}^{(n)} > Z(\alpha)]$ and $\Pi = P[Z + kc > Z(\alpha)]$.

It suffices to show $W_{0n}^{(n)} \xrightarrow{\mathcal{L}} Z + kc$.

$$W_{0n}^{(n)} = \frac{T_n - \Psi_n(\theta_0)}{\tau_n(\theta_0)} = \frac{\tau_n(\theta_n)W_{1n}^{(n)} + \Psi_n(\theta_n) - \Psi_n(\theta_0)}{\tau_n(\theta_0)}$$

$$= \frac{\tau_n(\theta_n)}{\tau_n(\theta_0)} W_{1n}^{(n)} + \frac{\Psi_n(\theta_n) - \Psi_n(\theta_0)}{\tau_n(\theta_0)}$$

We can write

$$\Psi_n(\theta_n) = \Psi_n(\theta_0 + k/\sqrt{n}) = \Psi_n(\theta_0) + (k/\sqrt{n})\Psi_n'(\tilde{\theta}_n)$$

for some $\tilde{\theta}_n$ such that $\theta_0 \le \tilde{\theta}_n \le \theta_n$. Hence

$$\frac{\Psi_n(\theta_n) - \Psi_n(\theta_0)}{\tau_n(\theta_0)} = \frac{k\ \Psi_n'(\tilde{\theta}_n)}{\sqrt{n}\ \tau_n(\theta_0)} = k\ \frac{\Psi_n'(\theta_0)}{\sqrt{n}\ \tau_n(\theta_0)}\ \frac{\Psi_n'(\tilde{\theta}_n)}{\Psi_n'(\theta_0)} .$$

By C5 and C6, $k \dfrac{\Psi_n'(\tilde{\theta}_0)}{\sqrt{n}\ \tau_n(\theta_0)} \dfrac{\Psi_n'(\tilde{\theta}_n)}{\Psi_n'(\theta_0)} \longrightarrow kc$ as $n \longrightarrow \infty$.

By C7, $\dfrac{\tau_n(\theta_n)}{\tau_n(\theta_0)} \longrightarrow 1$ as $n \longrightarrow \infty$. Thus

$$W_{0n} = a_n W_{1n}^{(n)} + b_n$$

where $a_n \longrightarrow 1$ and $b_n \longrightarrow kc$.

By C2, $W_{1n}^{(n)} \xrightarrow{\mathcal{L}} Z$, and so

$$W_{0n}^{(n)} = a_n W_{1n}^{(n)} + b_n \xrightarrow{\mathcal{L}} Z + kc . \quad []$$

For each sample size n suppose we have two test statistics T_{1n} and T_{2n} . Suppose $\{T_{1n}\}$ and $\{T_{2n}\}$ are both Pitman regular for testing H_0: $\theta = \theta_0$ versus H_1: $\theta > \theta_0$. Let k_1 , c_1 , k_2 and c_2 be the constants referred to in the regularity conditions. The asymptotic relative efficiency of $\{T_{1n}\}$ relative to $\{T_{2n}\}$ is defined to be the ratio

$$\mathrm{ARE}(T_{1n},T_{2n}) = \frac{n_2}{n_1}$$

where n_1 and n_2 are sample sizes such that the two tests T_{1n_1} and T_{2n_2} have identical power with respect to identical alternatives.

The alternatives are identical if $\dfrac{k_1}{\sqrt{n_1}} = \dfrac{k_2}{\sqrt{n_2}}$ and the powers are identical if $k_1c_1 = k_2c_2$.

Therefore

$$\mathrm{ARE}(T_{1n},T_{2n}) = \frac{n_2}{n_1} = (\sqrt{n_2}/\sqrt{n_1})^2 = (k_2/k_1)^2 = (c_1/c_2)^2 .$$

Define $R_n(\theta_0) = \Psi_n'(\theta_0)/\tau_n(\theta_0)$. Then C5 says $\lim_{n \to \infty} R_n/\sqrt{n} = c$.

Hence

$$\mathrm{ARE}(T_{1n},T_{2n}) = \lim_{n \to \infty} \frac{R_{2n}^2(\theta_0)}{R_{1n}^2(\theta_0)} .$$

APPENDIX C

Proofs of Chapter 3

This appendix contains proofs of the different lemmas given in Chapter 3 along with some preliminary lemmas needed in the proofs.

Lemma C.1 Let ϵ be a random variable with finite moments of all orders. Then for all b and all $a > 0$, (i) $E[(1+a|\epsilon|)^b] < \infty$ (ii) $E[(1+a|\epsilon|)^b \ln(1+a|\epsilon|)] < \infty$ (iii) $E[|\epsilon|(1+a|\epsilon|)^b] < \infty$.

Proof Since $1+a|\epsilon| \geq 1$ for all ϵ, hence $(1+a|\epsilon|)^b$ is an increasing function of b. If $b > 0$, let $[b]$ = the smallest integer greater than or equal b. Then for all ϵ,

$$(1+a|\epsilon|)^b \leq (1+a|\epsilon|)^{[b]} .$$

But $(1+a|\epsilon|)^{[b]}$ is a polynomial of degree $[b]$ in $|\epsilon|$ and hence has a finite expectation. Therefore $E\{(1+a|\epsilon|)^b\}$ is finite. If $b < 0$, then $(1+a|\epsilon|)^b < 1$ and (i) is immediate.

Since $1+a|\epsilon| \geq 1$ for all ϵ, hence $\ln(1+a|\epsilon|) < 1+a|\epsilon|$. Therefore,

$$(1+a|\epsilon|)^b \ln(1+a|\epsilon|) < (1+a|\epsilon|)^{b+1}$$

and (ii) follows from (i). For (iii) note that

$|\epsilon|(1+a|\epsilon|)^b < |\epsilon|(1+a|\epsilon|)^{[b]}$ which is also a polynomial in $|\epsilon|$. []

Lemma C.2 For all n let $U_{n1}, U_{n2}, \ldots, U_{nn}$ be continuous and iid. with distribution function F_n. Let A_n be a measurable subset of $\mathbb{R}^1$. Set $P_n = \Pr\{U_{n1} \in A_n\}$. Suppose that $P_n \longrightarrow 0$ as $n \longrightarrow \infty$. Then,

$$S_n = (1/n) \sum_{i=1}^{n} I_{A_n}(U_{ni}) \xrightarrow{p} 0$$

where I_{A_n} denotes the indicator function of A_n.

Proof:

In Proposition 5.3.4 of Laha and Rohatgi (1979, p.319) let

$$X_{ni} = (1/n) I_{A_n}(U_{ni})$$

and

$$S_n = \sum_{i=1}^{n} X_{ni} \ .$$

According to the proposition, $S_n \xrightarrow{p} 0$ if the following three conditions hold for any $\delta > 0$:

i. $n \Pr\{ |X_n| \geq \delta \} \longrightarrow 0$

ii. $\sum_{i=1}^{n} E(X_{ni}) \longrightarrow 0$

iii. $\sum_{i=1}^{n} \mathrm{var}(X_{ni}) \longrightarrow 0$

By definition of X_{ni} ,

$$\Pr\{ |X_{ni}| \geq \delta\} = \Pr\{I_{A_n}(U_{ni}) \geq n\delta\} \ .$$

Note that the first condition follows because $\Pr\{I_{A_n}(U_{ni}) \geq n\delta\} = 0$ for $n > 1/\delta$.

Since $X_{n1}, X_{n2}, \ldots, X_{nn}$ are i.i.d and $P_n = E[I_{A_n}(U_{ni})]$, therefore

$$\sum_{i=1}^{n} E(X_{ni}) = n\ (P_n/n) = P_n \longrightarrow 0$$

and

$$\sum_{i=1}^{n} \mathrm{var}(X_{ni}) = n\ (P_n(1-P_n)/n^2) \longrightarrow 0$$

Therefore conditions (ii) and (iii) hold. []

Proof of Lemma 3.2.1

(i) Let $L_n = [1 - (1+k_1/\sqrt{n}\,)^{\lambda_*}]/\ \lambda_*\sigma$ so that (3.2.17) becomes $B_n = (L_n, 0)$. Define

$P_n = \Pr\{Y \in A_n\}$. More prcisely,

$$P_n = \Pr\{Y_{n1} \in A_n\} = \Pr\{\epsilon_{n1} \in B_n\} = \Pr\{\epsilon \in B_n\}$$
$$= \Pr\{L_n < \epsilon < 0\} = F_\epsilon(0) - F_\epsilon(L_n) \ .$$

Since $L_n \longrightarrow 0$ as $n \longrightarrow \infty$, hence $P_n \longrightarrow F_\epsilon(0) - F_\epsilon(0) = 0$.

(ii) Using the indicator function of the set A_n , write

$$\sqrt{n}\ [(1/n) \sum_{Y_{ni} \in A_n} h(Y_{ni}-\theta_0,\lambda_*)] = \sqrt{n}\ [(1/n) \sum_{i=1}^{n} h(Y_{ni}-\theta_0,\lambda_*) I_{A_n}(Y_{ni})] \ .$$

Since $\theta_0 < Y_{ni} < \theta_n$ for all $Y_{ni} \in A_n$ and $h(Y_{ni}-\theta_0,\lambda_*)$ is monotone increasing in Y_{ni} , therefore

$$0 < h(Y_{ni}-\theta_0,\lambda_*) < ((1+k_1/\sqrt{n}\ )^{\lambda_*} - 1)/\lambda_* \qquad \text{(C.1)}$$

for all $Y_{ni} \in A_n$.

Multiplying (C.1) by $(1/\sqrt{n}\ ) I_{A_n}(Y_i)$ and taking the sum over i we

get

$$0 < \sqrt{n}\ [(1/n) \sum_{i=1}^{n} h(Y_{ni},\mu_0,\lambda_*)\ I_{A_n}(Y_{ni})] <$$
$$\sqrt{n}\ [(1+k_1/\sqrt{n}\ )^{\lambda_*}-1]/\lambda_*\ (\sum_{i=1}^{n} I_{A_n}(Y_{ni})/n) \ . \qquad \text{(C.2)}$$

From part (i) above and Lemma C.2 we have

$$\sum_{i=1}^{n} I_{A_n}(Y_{ni})/n \xrightarrow{P} 0 \ . \qquad \text{(C.3)}$$

Let $x = 1/\sqrt{n}$ so that

$$\sqrt{n}\ [(1+k_1/\sqrt{n}\ )^{\lambda_*} - 1]/\lambda_* = [(1+k_1x)^{\lambda_*} - 1]/(\lambda_* x)$$

Apply L'Hopital's rule to obtain

$$\lim_{x\to 0} [(1+k_1x)^{\lambda_*} - 1]/(\lambda_* x) = \lim_{x\to 0} [k_1(1+k_1x)^{\lambda_*-1}] = k_1 < \infty \ . \qquad \text{(C.4)}$$

(C.2), (C.3) and (C.4) imply

$$\sqrt{n}\ [(1/n) \sum_{Y_{ni} \in A_n} h(Y_{ni}-\theta_0,\lambda_*)] \xrightarrow{P} 0 \ .$$

(iii) Since

$$(1/n) \sum_{Y_i \in A_n} h^2(Y_i-\theta_0,\lambda_*) = (1/n) \sum_{i=1}^{n} h^2(Y_i-\theta_0,\lambda_*) I_{A_n}(Y_i)$$

and since from (C.1) we have

$$0 < h^2(Y_i-\theta_0,\lambda_*) < [(1+k_1/\sqrt{n}\,)^{\lambda_*} - 1]^2/\lambda_*^2 ,$$

hence

$$0 < (1/n) \sum_{i=1}^{n} h^2(Y_i-\theta_0,\lambda_*) I_{A_n}(Y_i) < ([(1+k_1/\sqrt{n}\,)^{\lambda_*} - 1]^2/\lambda_*^2) \sum_{i=1}^{n} I_{A_n}(Y_i)/n .$$

By the argument given in part (ii) above it suffices to show that $[(1+k_1/\sqrt{n}\,)^{\lambda_*} - 1]^2/\lambda_*^2$ is bounded as $n \longrightarrow \infty$. This is true since for all $\lambda_* > 0$

$$0 < [(1+k_1/\sqrt{n}\,)^{\lambda_*} -1\,]^2/\lambda_*^2 \longrightarrow 0 \quad \text{as} \quad n \longrightarrow \infty .$$

To prove parts (iv), (v) and (vi), first note that by an argument similar to the one used in parts (ii) and (iii), the following is true. If $|g_n(\epsilon)| \leq M$ for all $\epsilon \in B_n$ and all $n \geq n_0$, then

$$(1/n) \sum_{\epsilon_i \in B_n} g_n(\epsilon_i) \xrightarrow{p} 0 .$$

(iv) $g_n(\epsilon) = \sqrt{n}\epsilon$. Since $|\epsilon| < [(1+k_1/\sqrt{n}\,)^{\lambda_*} -1]/(\lambda_*\sigma)$ for $\epsilon \in B_n$, hence

$$|\sqrt{n}\epsilon| = \sqrt{n}\;|\epsilon| \leq \sqrt{n}\;[(1+k_1/\sqrt{n}\,)^{\lambda_*} -1]/(\lambda_*\sigma) \longrightarrow k_1/\sigma .$$

(v) $g_n(\epsilon) = \epsilon^2$. From (iv)

$$0 < \epsilon^2 < [(1+k_1/\sqrt{n}\,)^{\lambda_*} -1]^2/(\lambda_*\sigma)^2 \longrightarrow 0$$

(vi) $g_n(\epsilon) = (1 + \lambda_*\sigma|\epsilon|)^{1-1/\lambda_*}$.

$$g_n(\epsilon) \leq (1+k_1/\sqrt{n}\,)^{\lambda_*-1} \leq (1+k_1/\sqrt{n}\,)^{\lambda_*} \leq (1+k_1)^{\lambda_*} . \qquad []$$

Proof of Lemma 3.2.2

(i) From (3.2.14) since

$$\left|(1/n) \sum_{\epsilon_i \in B_n^c} \epsilon_i R_1(\epsilon_i, \lambda_*, n)\right|$$

$$\le (1/n) \sum_{\epsilon_i \in B_n^c} |\epsilon_i| \sum_{j=1}^{\infty} ((k_1/\sqrt{n})^j/j!)$$

$$\left|\prod_{m=1}^{j-1} (\lambda_* - m)\right| (1 + \lambda_* \sigma |\epsilon_i|)^{1-(j/\lambda_*)}$$

$$\le (1/\lambda_*)(1/n)\left(\sum_{\epsilon_i \in B_n^c} |\epsilon_i| (1+\lambda_*\sigma|\epsilon_i|)^{1-1/\lambda_*} \right)$$

$$\left(\left|\left(\sum_{j=1}^{\infty} (k_1/\sqrt{n})^j/j!\right) \prod_{m=0}^{j-1} (\lambda_* - m)\right|\right) . \qquad (C.5)$$

From a result similar to Lemma 3.2.1 (vi)

$$(1/n) \sum_{i=1}^{n} |\epsilon_i| (1+\lambda_*\sigma|\epsilon_i|)^{1-1/\lambda_*} - (1/n) \sum_{\epsilon_i \in B_n^c} |\epsilon_i| (1+\lambda_*\sigma|\epsilon_i|)^{1-1/\lambda_*}$$

$$\xrightarrow{p} 0 .$$

Since by the WLLN and the result of Lemma C.1 (iii)

$$(1/n) \sum_{i=1}^{n} |\epsilon_i| (1+\lambda_*\sigma|\epsilon_i|)^{1-1/\lambda_*} \xrightarrow{p} E\{|\epsilon|(1+\lambda_*\sigma|\epsilon|)^{1-1/\lambda_*}\} < \infty ,$$

and since

$$\sum_{j=1}^{\infty} (k_1/\sqrt{n})^j/j! \prod_{m=0}^{j-1} (\lambda_* - m) = (1+k_1/\sqrt{n})^{\lambda_*} - 1 \longrightarrow 0 \quad \text{as} \quad n \longrightarrow \infty ,$$

therefore from (C.5) $(1/n) \sum_{\epsilon_i \in B_n^c} \epsilon_i R_1(\epsilon_i, \lambda_*, n) \xrightarrow{p} 0$.

(ii) Note

$$(1/n) \sum_{\epsilon_i \in B_n^c} R_1^2(\epsilon_i, \lambda_*, n) =$$

$$(1/n) \sum_{\epsilon_i \in B_n^c} \left[\sum_{j=1}^{\infty} (\mathrm{sign}(\epsilon_i))^{j-1} (1+\lambda_*\sigma|\epsilon_i|)^{1-j/\lambda_*} ((k_1/\sqrt{n})^j/j!) \prod_{m=1}^{j-1} (\lambda_* - m)\right]^2$$

$$\le (1/\lambda_*^2)(1/n)\left[\sum_{\epsilon_i \in B_n^c} (1+\lambda_*\sigma|\epsilon_i|)^{1-1/\lambda_*}\right]^2 \left[\sum_{j=1}^{\infty} ((k_1/\sqrt{n})^j/j!) \prod_{m=0}^{j-1} (\lambda_* - m)\right]^2 ,$$

because $(1+\lambda_*\sigma|\epsilon_i|)^{1-j/\lambda_*} \le (1+\lambda_*\sigma|\epsilon_i|)^{1-1/\lambda_*}$ for all $j \ge 1$.

From Lemma 3.2.1(vi)

$$(1/n) \sum_{\epsilon_i \in B_n^c} (1+\lambda_*\sigma|\epsilon_i|)^{1-1/\lambda_*} - (1/n) \sum_{i=1}^{n} (1+\lambda_*\sigma|\epsilon_i|)^{1-1/\lambda_*} \xrightarrow{p} 0 .$$

It follows from the WLLN and the result of Lemma C.1(i) that

$$(1/n) \sum_{i=1}^{n} (1+\lambda_*\sigma|\epsilon_i|)^{1-1/\lambda_*} \xrightarrow{p} E\{(1+\lambda_*\sigma|\epsilon_i|)^{1-1/\lambda_*}\} < \infty$$

As before, $\sum_{j=1}^{\infty} ((k_1/\sqrt{n})^j/j!) \prod_{m=0}^{j-1} (\lambda_*-m) \longrightarrow 0$ as $n \longrightarrow \infty$, and

therefore $(1/n) \sum_{\epsilon_i \in B_n^c} R_1^2(\epsilon_i, \lambda_*, n) \xrightarrow{p} 0$.

(iii) From (3.2.16) write

$$(\sqrt{n}/n) \sum_{\epsilon_i \in B_n^c} R_2(\epsilon_i, \lambda_*, n) = k_1^2 V/(\lambda_*\sqrt{n}) \tag{C.6}$$

where

$$V = (1/n) \sum_{\epsilon_i \in B_n^c} \sum_{j=2}^{\infty} (\operatorname{sign}(\epsilon_i))^{j-1} ((k_1/\sqrt{n})^{j-2}/j!) \prod_{m=0}^{j-1} (\lambda_*-m) (1+\lambda_*\sigma|\epsilon_i|)^{1-(j/\lambda_*)}$$

To show that the left hand side of (C.6) tends in probability to 0 , it suffices to show that $|V|$ is bounded in probability.

Note that

$$(1+\lambda_*\sigma|\epsilon_i|)^{1-j/\lambda_*} \le (1+\lambda_*\sigma|\epsilon_i|)^{1-2/\lambda_*} \quad \text{for all } j \ge 2 .$$

Therefore,

$$|V| \le [(1/n) \sum_{\epsilon_i \in B_n^c} (1+\lambda_*\sigma|\epsilon_i|)^{1-2/\lambda_*}] \, [\sum_{j=2}^{\infty} ((k_1/\sqrt{n})^{j-2}/j!) |\prod_{m=0}^{j-1} (\lambda_*-m)| .$$

By a result similar to Lemma 3.2.1(vi),

$(1/n) \sum_{\epsilon_i \in B_n^c} (1+\lambda_*\sigma|\epsilon_i|)^{1-2/\lambda_*} \xrightarrow{p} 0$, so

$$(1/n)\sum_{i=1}^{n}(1+\lambda_*\sigma|\epsilon_i|)^{1-2/\lambda_*} - (1/n)\sum_{\epsilon_i \in B_n^c}(1+\lambda_*\sigma|\epsilon_i|)^{1-2/\lambda_*} \xrightarrow{p} 0 .$$

By the WLLN and Lemma C.1(i)

$$(1/n)\sum_{i=1}^{n}(1+\lambda_*\sigma|\epsilon_i|)^{1-2/\lambda_*} \xrightarrow{p} E\,\{(1+\lambda_*\sigma|\epsilon_i|)^{1-2/\lambda_*}\} < \infty .$$

So, it remains to show that

$$Q = \sum_{j=2}^{\infty}(k_1/\sqrt{n}\,)^{j-2}/j!\ \left|\prod_{m=0}^{j-1}(\lambda_*-m)\right| < \infty .$$

Since we are assuming that λ_* is positive,

$$Q \le \sum_{j=2}^{\infty}(k_1/\sqrt{n}\,)^{j-2}/\ j!\ \prod_{m=0}^{j-1}(\lambda_*+m) .$$

Let $q = [\lambda_*]$. Then

$$Q \le \sum_{j=2}^{\infty}(k_1/\sqrt{n}\,)^{j-2}/\ j!\ \prod_{m=0}^{j-1}(q+m)$$

$$= \sum_{j=2}^{\infty}(k_1/\sqrt{n}\,)^{j-2}\binom{q+j-1}{q-1} .$$

Since we are proving an asymptotic result, we can suppose

$$\sqrt{n} > k_1([\lambda_*]+2)/3 = k_1(q+2)/3 . \qquad \text{(C.7)}$$

We now use this assumption and the ratio test for convergence of infinite series to show that the series bounding Q is convergent and hence Q is finite. Let f(j) denote the term indexed by j in the expansion of the above series. Then

$$\begin{aligned} f(j+1)/f(j) &= \frac{k_1}{\sqrt{n}}\ \frac{(q+j)!}{(q-1)!(j+1)!}\ \frac{(q-1)!\ j!}{(q+j-1)!} \\ &= (k_1/\sqrt{n}\,)\ (q+j)/(j+1) \\ &= (k_1/\sqrt{n}\,)\ (1 + (q-1)/(j+1)) \\ &\le (k_1/\sqrt{n}\,)\ (1 + (q-1)/3) \qquad \text{for all } j \ge 2 \\ &= k_1/\sqrt{n}\ (q+2)/3 \\ &< 1 \qquad \text{(from (C.7))}. \qquad \square \end{aligned}$$

<u>Lemma C.3</u> Given $b \geq 0$, there exist functions $g_m(\epsilon,b)$, $m = 0,1$, not depending on n , such that

(i) Under H_0, $(|Y-\theta_0|+1)^b \leq g_0(\epsilon,b)$

(ii) Under H_1, $(|Y-\theta_0|+1)^b \leq g_1(\epsilon,b)$ for all $n \geq k_1^2$.

Also, these functions are such that $(1/n)\sum_{i=1}^{n} g_m(\epsilon_i,b)$, $m = 0,1$, is bounded in probability.

<u>Proof</u>

(i) $(|Y-\theta_0| + 1) = (1 + \lambda_*\sigma|\epsilon|)^{1/\lambda_*}$,

$$(|Y-\theta_0| + 1)^b = (1 + \lambda_*\sigma|\epsilon|)^{b/\lambda_*}$$
$$\leq (1 + \lambda_*\sigma|\epsilon|)^{[b/\lambda_*]} = g_0(\epsilon,b) \ . \tag{C.8}$$

where as defined in this thesis, $[x]$ = (integral part of x) + 1 .

Let $G_0 = E\{(1+\lambda_*\sigma|\epsilon|)^{[k_*/\lambda_*]}\}$. By Lemma C.1, $G_0 < \infty$. By the WLLN

$$(1/n)\sum_{i=1}^{n} g_0(\epsilon_i,b) \xrightarrow{p} G_0 \ . \tag{C.9}$$

From Lemma A.1 we conclude that $(1/n)\sum_{i=1}^{n} g_0(\epsilon_i,b)$ is bounded in probability.

<u>Under H_1</u>

(ii) For $Y \leq \theta_0$

$$(|Y-\theta_0|+1)^b I_{(-\infty,\theta_0)}(Y) = (\theta_n-Y+1-k_1/\sqrt{n})^b I_{(-\infty,\theta_0)}(Y)$$
$$\leq (\theta_n-Y+1)^b I_{(-\infty,\theta_0)}(Y)$$
$$= (1-\lambda_*\sigma\epsilon)^{b/\lambda_*} I_{(-\infty,L_n)}(\epsilon) \leq (1-\lambda_*\sigma\epsilon)^{[b/\lambda_*]} I_{(-\infty,L_n)}(\epsilon)$$

where $L_n = (1 - (1+k_1/\sqrt{n})^{\lambda_*})/\lambda_*$.

For $\theta_0 < Y < \theta_n$ note that $|Y-\theta_0|+1 < 1+k_1/\sqrt{n}$ and hence $(|Y-\theta_0|+1)^b < (1+k_1/\sqrt{n})^b$. For $n > k_1^2$, this implies,

$$(|Y-\theta_0|+1)^b I_{(\theta_0,\theta_n)}(Y) < 2^b I_{(L_n,0)}(-\epsilon) .$$

For $Y \geq \theta_n$

$$(|Y-\theta_0|+1)^b I_{(\theta_n,\infty)}(Y) = (Y-\theta_n+1+k_1/\sqrt{n})^b I_{(\theta_n,\infty)}(Y)$$

$$= (Y-\theta_n+1)^b [1 + k_1/(\sqrt{n}(Y-\theta_n+1))]^b I_{(\theta_n,\infty)}(Y) .$$

Since $\sqrt{n}(Y-\theta_n+1)I_{(\theta_n,\infty)}(Y) > 1$, hence

$$(|Y-\theta_0|+1)^b I_{(\theta_n,\infty)}(Y) \leq (Y-\theta_n+1)^b (1+k_1)^b I_{(\theta_n,\infty)}(Y)$$

$$= (1+\lambda_*\sigma\epsilon)^{b/\lambda_*} (1+k_1)^b I_{(0\ ,\infty)}(\epsilon)$$

$$\leq (1+\lambda_*\sigma\epsilon)^{[b/\lambda_*]} (1+k_1)^b I_{(0\ ,\infty)}(\epsilon) .$$

Let $g_1(\epsilon,b) = (1-\lambda_*\sigma\epsilon)^{[b/\lambda_*]} I_{(-\infty,L_n)}(\epsilon) + 2^b I_{(L_n,0)}(\epsilon) +$

$$(1+\lambda_*\sigma\epsilon)^{[b/\lambda_*]} (1+k_1)^b I_{(0,\infty)}(\epsilon) . \tag{C.10}$$

Let $G_1 = E\{(1-\lambda_*\sigma\epsilon)^{[b/\lambda_*]} I_{(-\infty,L_n)}(\epsilon) + (1+\lambda_*\sigma\epsilon)^{[b/\lambda_*]} (1+k_1)^b I_{(0,\infty)}(\epsilon) \}$.

Then

$$(1/n) \sum_{i=1}^{n} g_1(\epsilon_i,k_*) \xrightarrow{p} G_1 < \infty. \tag{C.11}$$

From Lemma A.1 $(1/n) \sum_{i=1}^{n} g_1(\epsilon_i,b)$ is bounded in probability.

<u>Lemma C.4</u> Given $0< \delta_0 < \lambda_*$, under H_m $(m = 0,1)$, there exist functions $M_{mk}(\epsilon)$, $m = 0,1$ and $k = 1,2,3$, such that $(1/n) \sum_{i=1}^{n} M_{mk}(\epsilon_i)$

is bounded in probability for all m and k and such that for all n,

(i) $$(1/n)\sum_{i=1}^{n}\sup_{|\lambda-\lambda_*|\le\delta_0}\left|\frac{\partial h(Y_i-\theta_0,\lambda)}{\partial\lambda}\right| \le (1/n)\sum_{i=1}^{n} M_{m1}(\epsilon_i)\ .$$

(ii) $$(1/n)\sum_{i=1}^{n}\sup_{|\lambda-\lambda_*|\le\delta_0}\left|\frac{\partial h^2(Y_i-\theta_0,\lambda)}{\partial\lambda}\right| \le (1/n)\sum_{i=1}^{n} M_{m2}(\epsilon_i)\ .$$

(iii) $$(1/n)\sum_{i=1}^{n}\sup_{|\lambda-\lambda_*|\le\delta_0}\left|\frac{\partial^2 h(Y_i-\theta_0,\lambda)}{\partial\lambda^2}\right| \le (1/n)\sum_{i=1}^{n} M_{m3}(\epsilon_i)\ .$$

Proof Since

(i) $$\frac{\partial h(Y-\theta_0,\lambda)}{\partial\lambda} = \lambda^{-1}\mathrm{sign}(Y-\theta_0)[(|y-\theta_0|+1)^{\lambda}\ln(|y-\theta_0|+1) - |h(Y-\theta_0,\lambda)|], \qquad (C.12)$$

hence

$$\sup_{|\lambda-\lambda_*|\le\delta_0}\left|\frac{\partial h(Y_i-\theta_0,\lambda)}{\partial\lambda}\right| \le$$

$$(\lambda_*-\delta_0)^{-1}(|Y-\theta_0|+1)^{\lambda_*+\delta_0}\ln(|Y-\theta_0|+1) + (\lambda_*-\delta_0)^{-2}[(|Y-\theta_0|+1)^{\lambda_*+\delta_0} - 1]$$
$$\le (\lambda_*-\delta_0)^{-1}(|Y-\theta_0|+1)^{\lambda_*+\delta_0+1} + (\lambda_*-\delta_0)^{-2}(|Y-\theta_0|+1)^{\lambda_*+\delta_0} \qquad (C.13)$$

because $|Y-\theta_0|+1 \ge \ln(|Y-\theta_0|+1)$.

Under H_m the result follows from Lemma C.3 with

$$M_{m1}(\epsilon) = (\lambda_*-\delta_0)^{-1}g_m(\epsilon,\lambda_*+\delta_0+1) + (\lambda_*-\delta_0)^{-2}\, g_m(\epsilon,\lambda_*+\delta_0) \qquad (C.14)$$

(ii) Since $\dfrac{\partial h^2(Y-\theta_0,\lambda)}{\partial\lambda} = 2h(Y-\theta_0,\lambda)\dfrac{\partial h(Y-\theta_0,\lambda)}{\partial\lambda}$ and since

$$\sup_{|\lambda-\lambda_*|\le\delta_0}|h(Y-\theta_0,\lambda)| \le (\lambda_*-\delta_0)^{-1}(|Y-\theta_0|+1)^{(\lambda_*+\delta_0)},$$

therefore multiplying (C.13) by $2(\lambda_*-\delta_0)^{-1}(|Y-\theta_0|+1)^{(\lambda_*+\delta_0)}$ we get

$$\sup_{|\lambda-\lambda_*|\le\delta_0}\left|\frac{\partial h^2(Y_i-\theta_0,\lambda)}{\partial\lambda}\right| \le$$

$$(\lambda_*-\delta_0)^{-2}(|Y-\theta_0|+1)^{2(\lambda_*+\delta_0)+1} + (\lambda_*-\delta_0)^{-3}(|Y-\theta_0|+1)^{2(\lambda_*+\delta_0)}\ .$$

Under H_m the result follows from Lemma C.3 with

$$M_{02}(\epsilon) = 2(\lambda_*-\delta_0)^{-3} g_m(\epsilon,2(\lambda_*+\delta_0)) + 2/(\lambda_*-\delta_0)^{-2} g_m(\epsilon,2(\lambda_*+\delta_0)+1) . \tag{C.15}$$

(iii) It is easy to see that

$$\left|\frac{\partial^2 h(Y_i-\theta_0,\lambda)}{\partial\lambda^2}\right| = (2/\lambda_*)\left|\frac{\partial h(Y_i-\theta_0,\lambda)}{\partial\lambda}\right| + (1/\lambda_*^2)(|Y-\theta_0|+1)(\ln(|Y-\theta_0|+1))^2 \tag{C.16}$$

and hence under H_m the result follows with

$$M_{m3}(\epsilon) = (2/(\lambda_*-\delta_0))\, M_{m1}(\epsilon) + (1/(\lambda_*-\delta_0)^2)\, g_m(\epsilon,\lambda_*+\delta_0+2) . \tag{C.17}$$

Since the functions $M_{mk}(\epsilon)$ for all m and k are linear combinations of the functions $g_m(\epsilon)$ $m = 0,1$, hence from (C.9) and (C.10)

$(1/n) \sum_{i=1}^{n} M_{mk}(\epsilon_i)$ are bounded in probability for all m and k . []

<u>Proof of Lemma 3.3.1</u>

(i) Consider a Taylor expansion of $\bar{h}_n(\theta_0,\hat{\lambda}_n)$ about $\hat{\lambda}_n = \lambda_*$.

$$\bar{h}_n(\theta_0,\hat{\lambda}_n) = \bar{h}_n(\theta_0,\lambda_*) + (\hat{\lambda}_n - \lambda_*) \left[(1/n) \sum_{i=1}^{n} \frac{\partial h(Y_i-\theta_0,\tilde{\lambda}_n)}{\partial\lambda} \right]$$

where $\tilde{\lambda}_n$ is such that $|\tilde{\lambda}_n-\lambda_*| \le |\hat{\lambda}_n-\lambda_*|$. Hence

$$\bar{h}_n(\theta_0,\hat{\lambda}_n) - \bar{h}_n(\theta_0,\lambda_*) = (\hat{\lambda}_n-\lambda_*) \left[(1/n) \sum_{i=1}^{n} \frac{\partial h(Y_i-\theta_0,\tilde{\lambda}_n)}{\partial\lambda} \right]$$

From Lemma C.4 (i), since under H_m ($m = 0,1$)

$$(1/n) \sum_{i=1}^{n} \left|\frac{\partial h(Y_i-\theta_0,\tilde{\lambda}_n)}{\partial\lambda}\right| I_{(\lambda_*-\delta_0,\lambda_*+\delta_0)}(\tilde{\lambda}_n) \le (1/n) \sum_{i=1}^{n} \sup_{|\lambda-\lambda_*|\le\delta_0} \left|\frac{\partial h(Y_i-\theta_0,\lambda)}{\partial\lambda}\right|$$

$$\le (1/n) \sum_{i=1}^{n} M_{m1}(\epsilon_i) ,$$

we see that $(1/n) \sum_{i=1}^{n} \left|\frac{\partial h(Y_i-\theta_0,\tilde{\lambda}_n)}{\partial\lambda}\right| I_{(\lambda_*-\delta_0,\lambda_*+\delta_0)}(\tilde{\lambda}_n)$ is bounded in probability. Since $\Pr[|\tilde{\lambda}_n - \lambda_*| \le \delta_0] \longrightarrow 1$, hence

$I_{(\lambda_*-\delta_0,\lambda_*+\delta_0)}(\bar{\lambda}_n) \xrightarrow{p} 0$ and so by Lemma A.3 we conclude that

$(1/n)\left|\sum_{i=1}^{n}\frac{\partial h(Y_i-\theta_0,\bar{\lambda}_n)}{\partial\lambda}\right|$ is bounded in probability. (Note that this assumes measurability of $\bar{\lambda}_n$. If $\bar{\lambda}_n$ is not measurable, the argument might still go through using outer measure as in Huber (1967, p.).

Since $\hat{\lambda}_n \xrightarrow{p} \lambda_*$, therefore from Theorem A.1 in Appendix A with

$A_n = \hat{\lambda}_n-\lambda_*$ and $B_n = (1/n)\sum_{i=1}^{n}\frac{\partial h(Y_i-\theta_0,\bar{\lambda}_n)}{\partial\lambda}$, we conclude

$$\bar{h}_n(\theta_0,\hat{\lambda}_n) - \bar{h}_n(\theta_0,\lambda_*) \xrightarrow{p} 0 .$$

(ii) From a similar argument using a Taylor expansion of

$(1/n)\sum_{i=1}^{n} h^2(Y-\theta_0,\hat{\lambda}_n)$ about $\hat{\lambda}_n=\lambda_*$ and the functions $M_{m2}(\epsilon)$ in Lemma C.4(ii), it can be shown that

$$(1/n)\sum_{i=1}^{n} h^2(Y-\theta_0,\hat{\lambda}_n) - (1/n)\sum_{i=1}^{n} h^2(Y-\theta_0,\lambda_*) \xrightarrow{p} 0 . \quad []$$

Proof Lemma 3.3.2 From Lemma C.4(iii)

$$(1/n)\sum_{i=1}^{n}\left|\frac{\partial^2 h(Y_i-\theta_0,\lambda)}{\partial\lambda^2}\right| I_{(\lambda_*-\delta_0,\lambda_*+\delta_0)}(\bar{\lambda}_n) \le (1/n)\sum_{i=1}^{n}\sup_{|\lambda-\lambda_*|\le\delta_0}\left|\frac{\partial^2 h(Y_i-\theta_0,\lambda)}{\partial\lambda^2}\right|$$

$$\le (1/n)\sum_{i=1}^{n} M_{m3}(\epsilon_i) .$$

Therefore $(1/n)\sum_{i=1}^{n}\left|\frac{\partial^2 h(Y_i-\theta_0,\bar{\lambda}_n)}{\partial\lambda^2}\right| I_{(\lambda_*-\delta_0,\lambda_*+\delta_0)}(\bar{\lambda}_n)$ is bounded in probability. The proof proceeds as in Lemma 3.3.1. []

Proof of Lemma 3.3.3 Since

$$\frac{\partial h(Y-\theta_0,\lambda_*)}{\partial\lambda} = (1/\lambda_*)\mathrm{sign}(Y-\theta_0)[(|Y-\theta_0|+1)^{\lambda_*}\ln(|Y-\theta_0|+1) - |h(Y-\theta_0,\lambda_*)|] ,$$

therefore under H_0

$$\frac{\partial h(Y-\theta_0,\lambda_*)}{\partial\lambda} = (1/\lambda_*^2)\,\mathrm{sign}(\epsilon)\,[(1+\lambda_*\sigma|\epsilon|)^{\lambda_*}\ln(1+\lambda_*\sigma|\epsilon|) - \lambda_*\sigma|\epsilon|] \ . \tag{C.18}$$

Under symmetric distributions of ϵ , $E[\frac{\partial h(Y-\theta_0,\lambda_*)}{\partial\lambda}] = 0$ because (C.18) is an odd function of ϵ . Therefore by the WLLN

$$\frac{\partial\bar{h}_n(\theta_0,\lambda_*)}{\partial\lambda} \xrightarrow{P} 0 \ .$$

Under H_1 from a Taylor expansion about θ_n we can write

$$\frac{\partial h(Y-\theta_0,\lambda_*)}{\partial\lambda} = \frac{\partial h(Y-\theta_n,\lambda_*)}{\partial\lambda} + (\theta_n-\theta_0)\,\frac{\partial^2 h(Y-\tilde{\theta}_n,\lambda_*)}{\partial\lambda\,\partial\theta} \ , \quad \theta_0\le\tilde{\theta}_n\le\theta_n \ .$$

As for H_0, $(1/n)\sum_{i=1}^{n}\frac{\partial h(Y-\theta_n,\lambda_*)}{\partial\lambda} \xrightarrow{P} 0$. So, it suffices to show that $(1/n)\sum_{i=1}^{n}\frac{\partial^2 h(Y-\tilde{\theta}_n,\lambda_*)}{\partial\lambda\,\partial\theta}$ is bounded in probability.

$$\frac{\partial^2 h(Y-\theta,\lambda_*)}{\partial\lambda\,\partial\theta} = -(|Y-\theta|+1)^{\lambda-1}\ln(|Y-\theta|+1) \ .$$

Therefore

$$\left|\frac{\partial^2 h(Y-\theta,\lambda_*)}{\partial\lambda\,\partial\theta}\right| \le (|Y-\theta|+1)^{\lambda}$$

$$\left|\frac{\partial^2 h(Y-\tilde{\theta}_n,\lambda_*)}{\partial\lambda\,\partial\theta}\right| \le (|Y-\tilde{\theta}_n|+1)^{\lambda_*} \ .$$

From the triangular inequality we have

$$(|Y-\tilde{\theta}_n|+1)^{\lambda_*} \le (|Y-\theta_0|+1)^{\lambda_*} + (|Y-\theta_n|+1)^{\lambda_*} \ .$$

Under H_1 $(|Y-\theta_n|+1)^{\lambda_*} = \lambda_*\sigma|\epsilon|+1$. Therefore it suffices to show that there exists some function $M(\epsilon)$ with $E[M(\epsilon)] < \infty$ and such that for all n , $(|Y-\theta_0|+1)^{\lambda_*} \le M(\epsilon)$.

$$|Y-\theta_0|+1 = |Y-\theta_n+k/\sqrt{n}|+1 \le |Y-\theta_n|+k/\sqrt{n}+1 \le |Y-\theta_n|+k+1 \ .$$

Therefore

$$(|Y-\theta_0|+1)^{\lambda_*} \le (|Y-\theta_n|+1+k)^{\lambda_*} = [(1+\lambda_*\sigma|\epsilon|)^{1/\lambda_*} + k]^{\lambda_*}$$
$$\le \{(1+\lambda_*\sigma|\epsilon|)^{1/\lambda_*} + k\}^{[\lambda_*]} .$$

Let $m = [\lambda_*]$, then

$$\{(1+\lambda_*\sigma|\epsilon|)^{1/\lambda_*} + k\}^{[\lambda_*]} = \sum_{j=1}^{m} \binom{m}{j} (1+\lambda_*\sigma|\epsilon|)^{j/\lambda_*} k^{m-j}.$$

Let

$$M(\epsilon) = \sum_{j=1}^{m} \binom{m}{j} (1+\lambda_*\sigma|\epsilon|)^{[j/\lambda_*]} k^{m-j} ,$$

then $M(\epsilon)$ is a polynomial in $|\epsilon|$ with a finite expectation. []

APPENDIX D

Proofs of Chapter 4

Proof of Lemma 4.3.1

From a Taylor expansion of $\bar{h}_n(\theta_0, \hat{\underline{\lambda}}_n)$ about $\hat{\lambda}_{1n} = \lambda_{1*}$ and $\hat{\lambda}_{2n} = \lambda_{2*}$

$$\bar{h}_n(\theta_0, \hat{\underline{\lambda}}_n) - \bar{h}_n(\theta_0, \underline{\lambda}_*) = (\hat{\lambda}_{1n} - \lambda_{1*})(1/n) \sum_{i=1}^{n} \frac{\partial h(Y_i - \theta_0, \tilde{\underline{\lambda}}_n)}{\partial \lambda_1} I_{(-\infty, \theta_0)}(Y_i)$$

$$+ (\hat{\lambda}_{2n} - \lambda_{2*})\ (1/n) \sum_{i=1}^{n} \frac{\partial h(Y_i - \theta_0, \tilde{\underline{\lambda}}_n)}{\partial \lambda_2} I_{(\theta_0, \infty)}(Y_i) \qquad \text{(D.1)}$$

where $\tilde{\underline{\lambda}}_n$ is such that $|\tilde{\lambda}_{1n} - \lambda_{1*}| \leq |\hat{\lambda}_{1n} - \lambda_{1*}|$ and $|\tilde{\lambda}_{2n} - \lambda_{2*}| \leq |\hat{\lambda}_{2n} - \lambda_{2*}|$. The result follows if the LHS of (D.1) tends in probability to zero. Consistency of the maximum likelihood estimator of $\underline{\lambda}$ (follows from Appendix E) implies, $\hat{\lambda}_{1n} \xrightarrow{p} \lambda_{1*}$ and $\hat{\lambda}_{2n} \xrightarrow{p} \lambda_{2*}$. Hence by Theorem A.1 of Appendix A it suffices to show that there exist M_{km} such that under H_m, m=1,2 ,

$$\Pr\left\{(1/n) \sum_{i=1}^{n} \left|\frac{\partial h(Y_i - \theta_0, \tilde{\underline{\lambda}}_n)}{\partial \lambda_k}\right| \leq M_{km}\right\} \xrightarrow{p} 1 \quad \text{as} \quad n \longrightarrow \infty \quad k = 1,2.$$

To show that there exists such M_{km} , note that $\left|\frac{\partial h(Y - \theta_0, \underline{\lambda})}{\partial \lambda_k}\right|$, k = 1,2

is the same as the first derivative with respect to λ under the John-Draper transformations. Therefore we define the function M_{10} from (C.12) upon replacing λ_* by λ_{1*} . Similarly, we define the function M_{20} from a replacement of λ_* in (C.12) by λ_{2*} and so forth. []

Proof of Lemma 4.3.2 Note that

$$\left|\frac{\partial^2 h(Y - \theta_0, \underline{\lambda})}{\partial \lambda_1^2}\right| = 2/\lambda_1 \left|\frac{\partial h(Y - \theta_0, \underline{\lambda})}{\partial \lambda_1}\right| + 1/\lambda_1 (\theta_0 - Y + 1)^{\lambda_1} \ln(\theta_0 - Y + 1) .$$

This is the same as (C.16) in Appendix C when λ there is replaced by λ_1. Therefore the result follows from (C.17) and (C.18) under H_0 and H_1 if λ_* in these two equations is replaced by λ_{1*}.

A similar argument holds for $\left|\dfrac{\partial^2 h(Y-\theta_0,\underline{\lambda})}{\partial\lambda_2^2}\right|$. []

<u>Proof of Lemma 4.3.3</u>

Since

$$\frac{\partial h(Y-\theta_0,\underline{\lambda}_*)}{\partial\lambda_1} = (-1/\lambda_{1*})[(\theta_0-Y+1)^{\lambda_{1*}}\ln(\theta_0-Y+1) + h(Y-\theta_0,\underline{\lambda}_*)]I_{(-\infty,\theta_0)}(Y) ,$$

hence under H_0

$$\frac{\partial h(Y-\theta_0,\underline{\lambda}_*)}{\partial\lambda_1} = [(-1/\lambda_{1*}^2)\ (1-\lambda_{1*}\sigma\epsilon)\ln(1-\lambda_{1*}\sigma\epsilon) - \sigma\epsilon/\lambda_{1*}]\ I_{(-\infty,0)}(\epsilon) .$$

Therefore

$$(1/n)\sum_{i=1}^{n}\frac{\partial h(Y_i-\theta_0,\underline{\lambda}_*)}{\partial\lambda_1} = (1/n)\sum_{i=1}^{n}[(-1/\lambda_{1*}^2)\ (1-\lambda_{1*}\sigma\epsilon_i)^{\lambda_{1*}}\ln(1-\lambda_{1*}\sigma\epsilon_i) - \sigma\epsilon_i/\lambda_{1*}]\ I_{(-\infty,0)}(\epsilon_i) .$$

Note that

$$(1/n)\sum_{i=1}^{n}(1-\lambda_{1*}\sigma\epsilon_i)^{\lambda_{1*}}\ln(1-\lambda_{1*}\sigma\epsilon_i)I_{(-\infty,0)}(\epsilon_i) \le 1/n\sum_{i=1}^{n}(1-\lambda_{1*}\sigma\epsilon_i)^{\lambda_{1*}+1}I_{(-\infty,0)}(\epsilon_i)$$
$$\le (1/n)\sum_{i=1}^{n}(1-\lambda_{1*}\sigma\epsilon_i)^{[\lambda_{1*}]+1}I_{(-\infty,0)}(\epsilon_i) .$$

But $(1-\lambda_{1*}\sigma\epsilon_i)^{[\lambda_{1*}]+1}I_{(-\infty,0)}(\epsilon_i)$ is a polynomial of a normal variable and hence has a finite expectation. Also, $(1/n)\sum_{i=1}^{n}\epsilon_i I_{(-\infty,0)}(\epsilon_i)$ has a finite expectation. Therefore by the WLLN, under H_0 for $Y \le \theta_0$

$$(1/n)\sum_{i=1}^{n}\frac{\partial h(Y-\theta_0,\underline{\lambda}_*)}{\partial\lambda_1} \xrightarrow{P} E\{\ [(-1/\lambda_{1*}^2)(1-\lambda_{1*}\sigma\epsilon_i)^{\lambda_{1*}}\ln(1-\lambda_{1*}\sigma\epsilon_i) - \sigma\epsilon_i/\lambda_{1*}]I_{(-\infty,0)}(\epsilon_i)\ \} = \ < \infty .$$

From a similar arguemnt for $Y \geq \theta_0$ it can be shown that

$$(1/n) \sum_{i=1}^{n} \frac{\partial h(Y_i-\theta_0,\underline{\lambda}_*)}{\partial \lambda_2} \xrightarrow{p} E\{ [(1/\lambda_{2*}^2)(1+\lambda_{2*}\sigma\epsilon_i)^{\lambda_{2*}} \ln(1+\lambda_{2*}\sigma\epsilon_i) -$$

$$\sigma\epsilon_i/\lambda_{2*}]I_{(0,\infty)}(\epsilon_i) \} < \infty .$$

Therefore under H_0

$$\frac{\partial \bar{h}(\theta_0,\underline{\lambda}_*)}{\partial \underline{\lambda}} \xrightarrow{p} E[S(\epsilon,\underline{\lambda}_*)] \quad \text{where}$$

$$E[S(\epsilon,\underline{\lambda}_*)] = \begin{bmatrix} E\{ [-\lambda_{1*}^{-2}(1-\lambda_{1*}\sigma\epsilon)\ln(1-\lambda_{1*}\sigma\epsilon) - \lambda_{1*}^{-1}\sigma\epsilon]I_{(-\infty,0)}(\epsilon) \} \\ \\ E\{ [\lambda_{2*}(1+\lambda_{2*}\sigma\epsilon)\ln(1+\lambda_{2*}\sigma\epsilon) - \lambda_{2*}^{-1}\sigma\epsilon]I_{(0,\infty)}(\epsilon) \} \end{bmatrix}$$

Under H_1 as we did in proving Lemma 3.3.3. we write

$$\frac{\partial h(Y-\theta_0,\underline{\lambda}_*)}{\partial \underline{\lambda}} = \frac{\partial h(Y-\theta_n,\underline{\lambda}_*)}{\partial \underline{\lambda}} + (\theta_n-\theta_0)\frac{\partial^2 h(Y-\theta_0,\underline{\lambda}_*)}{\partial \underline{\lambda}\ \partial \theta}, \quad \theta_0 \leq \bar{\theta}_n \leq \theta_n .$$

As for H_0, $(1/n) \sum_{i=1}^{n} \frac{\partial h(Y_i-\theta_0,\underline{\lambda}_*)}{\partial \underline{\lambda}} \xrightarrow{p} E[S(\epsilon,\underline{\lambda}_*)]$.

Since $\theta_n-\theta_0 = k_1/\sqrt{n} \longrightarrow 0$ as $n \longrightarrow \infty$, so it suffices to show

that the functions $\frac{\partial^2 h(Y-\theta_0,\underline{\lambda}_*)}{\partial \lambda_1 \partial \theta}$ and $\frac{\partial^2 h(Y-\theta_0,\underline{\lambda}_*)}{\partial \lambda_2 \partial \theta}$ are both bounded in

probability. This follows the same steps given in the proof of Lemma 3.3.3. []

APPENDIX E

Consistency and Asymptotic Normality of $\hat{\underline{\lambda}}_n$ the MLE of $\underline{\lambda}$

Let $h(y-\theta_0,\lambda_1,\lambda_2)$ denote the two-domain transformations. Since we are assuming that λ_{1*} and λ_{2*} are positive, the parameter space is defined as

$$\Omega = \{\ \omega = (\lambda_1,\lambda_2,\sigma)^t:\ \lambda_1 > 0\ ,\ \lambda_2 > 0\ ,\ \sigma > 0\ \}\ .$$

Any $\omega \in \Omega$ is an element of an open set I contained in Ω of the form

$$I = \{\omega : a_1 < \lambda_1 < b_1\ ,\ a_2 < \lambda_2 < b_2\ ,\ a_3 < \sigma < b_3\ \} \tag{E.1}$$

for some positive numbers $a_i < b_i$, $i = 1,2,3$. Without loss of generality for $i=1,2$ we assume that $b_i/a_i \le 2$. Under the model $h(y-\theta_0,\lambda_1,\lambda_2) = \sigma\epsilon$ the pdf of Y with ϵ assumed to have a standard normal distribution is given by,

$$f_Y(y,\lambda_1,\lambda_2,\sigma) = 1/\sigma\ f_\epsilon(h(y-\theta_0,\lambda_1,\lambda_2)/\sigma)\ J(\lambda_1,\lambda_2) \tag{E.2}$$

where,

$$1/\sigma\ f_\epsilon(h(y-\theta_0,\lambda_1,\lambda_2)/\sigma) = (2\pi\sigma^2)^{-1/2}\exp\{-1/2\sigma^2h^2(y-\theta_0,\lambda_1,\lambda_2)\}$$

$$=(2\pi\sigma^2)^{-1/2}\exp\{-1/2\sigma^2[(\frac{1-(\theta_0-y+1)^{\lambda_1}}{\lambda_1})^2 I_{(-\infty,\theta_0)}(y) + (\frac{(y-\theta_0+1)^{\lambda_2}-1}{\lambda_2})^2 I_{(\theta_0,\infty)}(y)\ ]\} \tag{E.3}$$

and,

$$J(\lambda_1,\lambda_2) = \frac{\partial h(y-\theta_0,\lambda_1,\lambda_2)}{\partial y}$$

$$= (\theta_0-y+1)^{\lambda_1-1} I_{(-\infty,\theta_0)}(y) + (y-\theta_0+1)^{\lambda_2-1} I_{(\theta_0,\infty)}(y) \tag{E.4}$$

For a given y the loglikelihood as a function of ω is given by,

$$L(\lambda_1,\lambda_2,\sigma) = -1/2\ \ln(\sigma^2) - 1/2\sigma^2\ \{[(\frac{1-(\theta_0-y+1)^{\lambda_1}}{\lambda_1})^2\ I_{(-\infty,\theta_0)}(y)$$

$$+ (\frac{(y-\theta_0+1)^{\lambda_2}-1}{\lambda_2})^2\ I_{(\theta_0,\infty)}(y)\ ]\} + \ln\ (J(\lambda_1,\lambda_2)) \qquad (E.5)$$

<u>lemma E.1</u> Let $L(\lambda_1,\lambda_2,\sigma)$ be as defined in (E.5) and let $\frac{\partial^k f(y,\lambda_1,\lambda_2,\sigma)}{\partial\lambda_1^r\ \partial\lambda_2^s\ \partial\sigma^t}$ denote the k^{th} partial derivative of $f_Y(y,\lambda_1,\lambda_2,\sigma)$ differentiated r-times, s-times and t-times with respect to λ_1, λ_2 and σ respectively where, $k = 1,2,3$ and $r,s,t = 0,1,2,3$ are such that $r+s+t = k$. There exist functions $M_{rst}(y)$ such that,

$$\left|\frac{\partial^k f(y,\lambda_1,\lambda_2,\sigma)}{\partial\lambda_1^r\ \partial\lambda_2^s\ \partial\sigma^t}\right| \le M_{rst}(y) \qquad \text{for all } \omega \in I$$

and

$$\int M_{rst}(y)\ dy \ \le\ \infty \qquad \text{for all } k,\ r,\ s \text{ and } t$$

<u>Proof</u>

It is more convenient to give the proof in terms of the loglikelihood by noting that for any three times differentiable

function $g(\gamma,\nu)$

$$\left|\frac{\partial g(\gamma,\nu)}{\partial\gamma}\right| = \left|\frac{\partial \ln(g(\gamma,\nu))}{\partial\gamma}\right|\ \left|g(\gamma,\nu)\right| \qquad (E.6)$$

$$\left|\frac{\partial^2 g(\gamma,\nu)}{\partial\gamma\ \partial\nu}\right| \le \left|\frac{\partial^2 \ln(g(\gamma))}{\partial\gamma\ \partial\nu}\right|\left|g(\gamma,\nu)\right| + \left|\frac{\partial \ln(g(\gamma,\nu))}{\partial\gamma}\right|\left|\frac{\partial \ln(g(\gamma,\nu))}{\partial\nu}\right|\left|g(\gamma,\nu)\right| \qquad (E.7)$$

and

$$\left|\frac{\partial^3 g(\gamma,\nu)}{\partial\gamma^2\,\partial\nu}\right| \le \left|\frac{\partial^3 \ln g(\gamma,\nu)}{\partial\gamma^2\partial\nu}\right| \; |g(\gamma,\nu)| + 2\left|\frac{\partial^2 \ln(g(\gamma,\nu))}{\partial\gamma\;\partial\nu}\right|$$
$$\left|\frac{\partial \ln(g(\gamma,\nu))}{\partial\gamma}\right| \; |g(\gamma,\nu)| + \left|\frac{\partial^2 \mathrm{LN}(g(\gamma,\nu))}{\partial\gamma^2}\right| \; \left|\frac{\partial \ln(g(\gamma,\nu))}{\partial\nu}\right|$$
$$|g(\gamma,\nu)| + \left|\frac{\partial \ln(g(\gamma,\nu))}{\partial\gamma}\right|^2 \; \left|\frac{\partial \ln(g(\gamma,\nu))}{\partial\nu}\right| \; |g(\gamma,\nu)| \qquad \text{(E.8)}$$

We first start to look for some function $f^*(y)$ that dominates $f_Y(y,\lambda_1,\lambda_2,\sigma)$ over I that is, $f_Y(y,\lambda_1,\lambda_2,\sigma) \le f^*(y)$ for all $\omega \in I$ then we find functions $G_{rst}(y)$ such that, $\left|\frac{\partial^k L(\lambda_1,\lambda_2,\sigma)}{\partial\lambda_1^r\partial\lambda_2^s\;\partial\sigma^t}\right| \le G_{rst}(y)$ for all $\omega \in I$ and such that, $\int G_{rst}(y)\; f^*(y)\; dy < \infty$. The results will then follow with $M_{rst}(y)$ taken to be some linear combination of the functions $G_{ijl}(y)$ for some i,j,l determined from equations (E.7) and (E.8) and from which partial derivative is considered.

We give below some upper bounds over I for the different terms included in the functions $f_Y(y,\lambda_1,\lambda_2,\sigma)$ and $\frac{\partial^k L(\lambda_1,\lambda_2,\sigma)}{\partial\lambda_1^r\partial\lambda_2^s\;\partial\sigma^t}$ so that if every term is replaced by its upper bound we get $f^*(y)$ and $G_{rst}(y)$

$$1/\lambda_1 < 1/a_1 \;,\quad 1/\lambda_2 < 1/a_2 \;,\quad 1/\sigma < 1/a_3 \;,\quad -1/\sigma < -1/b_3$$
$$(\theta_0-y+1)^{\lambda_1} < (\theta_0-y+1)^{b_1} \;,\quad (y-\theta_0+1)^{\lambda_2} < (y-\theta_0+1)^{b_2}$$
$$-\left(\frac{1-(\theta_0-y+1)^{\lambda_1}}{\lambda_1}\right)^2 \le -\left(\frac{1-(\theta_0-y+1)^{a_1}}{b_1}\right)^2$$
$$-\left(\frac{(y-\theta_0+1)^{\lambda_2}-1}{\lambda_2}\right)^2 \le -\left(\frac{(y-\theta_0+1)^{a_2}-1}{b_2}\right)^2$$

1. <u>An integrable upper bound for $f_Y(y,\lambda_1,\lambda_2,\sigma)$</u>

From (E.3)

$$f_\epsilon(h(y-\theta_0,\lambda_1,\lambda_2)/\sigma) \le (2\pi)^{-1/2}\exp\{-(1/2b_3^2)$$
$$[(a_1/b_1)^2\left(\frac{1-(\theta_0-y+1)^{\lambda_1}}{\lambda_1}\right)^2 I_{(-\infty,\theta_0)}(Y) + (a_2/b_2)^2\left(\frac{(y-\theta_0+1)^{\lambda_2}-1}{\lambda_2}\right)^2 I_{(\theta_0,\infty)}(Y)\;]\}\;.$$

Let $\tau = \max(\frac{b_1 b_3}{a_1}, \frac{b_2 b_3}{a_2})$ then,

$$f_\epsilon(h(y-\theta_0,\lambda_1,\lambda_2)/\sigma) \le (2\pi)^{-1/2} \exp\{ - 1/\tau^2 h^2(y-\theta_0,a_1,a_2) \}$$

hence

$$f_\epsilon(h(y-\theta_0,\lambda_1,\lambda_2)/\sigma) \le \tau (1/\tau f_\epsilon(h(y-\theta_0,a_1,a_2)/\tau) \qquad (E.9)$$

From (E.4)

$$J(\lambda_1,\lambda_2) \le J(b_1,b_2) \qquad (E.10)$$

let,

$$f^*(y) = \tau/a_3 (1/\tau f_\epsilon(h(y-\theta_0,a_1,a_2)/\tau) J(b_1,b_2)$$

then it follows from (E.2) , (E.9) and (E.10) that

$$f_Y(y,\lambda_1,\lambda_2,\sigma) \le f^*(y) \qquad \text{for all } \omega \in I \qquad (E.11)$$

It remains to show that, $\int_{-\infty}^{\infty} f^*(y)\, dy < \infty$. Note that $f^*(y)$ can be written as,

$$f^*(y) = f_Y(y,a_1,a_2,\tau) [\tau/a_3 J(b_1-a_2+1 ,b_2-a_2+1)] . \qquad (E.12)$$

Let

$$X = \frac{1-(\theta_0-Y+1)^{a_1}}{a_1} I_{(-\infty,\theta_0)}(Y) + \frac{(Y-\theta_0+1)^{a_2}-1}{a_2} I_{(\theta_0,\infty)}(Y) \qquad (E.13)$$

From (E.12) observe that $f_X(x,a_1,a_2,\tau)$ is a normal density with mean zero and variance τ^2 . In proving the integrability of $f^*(y)$ and later each of the functions $G_{rst}(y)f^*(y)$ we express each of these functions in terms of X and then show that the resulting expression is some function of a normal random variable with a finite expectation.

From (E.4),

$$J(b_1-a_1+1,b_2-a_2+1) = (\theta_0-y+1)^{b_1-a_1} I_{(-\infty,\theta_0)}(y) + (y-\theta_0+1)^{b_2-a_2} I_{(\theta_0,\infty)}(y)$$

from (E.13) note that $(\theta_0-y+1) = (1-a_1x)^{1/a_1} I_{(-\infty,0)}(x)$ and

$(y-\theta_0+1) = (1+a_2x)^{1/a_2} I_{(0,\infty)}(x)$. Hence

$$J(b_1-a_1+1, b_2-a_2+1) = (1-a_1x)^{b_1/a_1-1} I_{(-\infty,0)}(x) + (1+a_2x)^{b_2/a_2-1} I_{(0,\infty)}(x)$$

Therefore under the assumption that $b_i/a_i \leq 2$ we get,

$$\int_{-\infty}^{\infty} f^*(y)\, dy \leq \tau/(a_3 \sqrt{2\pi\tau^2}) \left\{ \int_{-\infty}^{0} (1-a_1x) \exp[-x^2/2\tau^2]\, dx + \int_{0}^{\infty} (1+a_2x) \exp[-x^2/2\tau^2]\, dx \right\}$$

$$< \infty \tag{E.14}$$

2. An integrable upper bound for $\left|\frac{\partial L(\lambda_1,\lambda_2,\sigma)}{\partial\lambda_1}\right|$

From (E.5) note that,

$$\left|\frac{\partial L(\lambda_1,\lambda_2,\sigma)}{\partial\lambda_1}\right| = \Big\{\ln(\theta_0-Y+1) + (1/\lambda_1\sigma^2)\ [(\frac{(\theta_0-y+1)^{\lambda_1}-1}{\lambda_1})$$

$$(\theta_0-Y+1)^{\lambda_1} \ln(\theta_0-Y+1) + (\frac{1-(\theta_0-y+1)^{\lambda_1}}{\lambda_1})^2\Big\}\, I_{(-\infty,\theta_0)}(Y) \quad . \tag{E.15}$$

Let

$$G_{100}(y) = \Big\{\ln(\theta_0-Y+1) + (1/a_1a_3^2)\ [(\frac{1-(\theta_0-y+1)^{b_1}}{a_1})^2 -$$

$$(\frac{1-(\theta_0-y+1)^{b_1}}{a_1})\ (\theta_0-Y+1)^{b_1} \ln(\theta_0-Y+1)]\Big\}\, I_{(-\infty,\theta_0)}(Y) \tag{E.16}$$

which implies that,

$$\left|\frac{\partial L(\lambda_1,\lambda_2,\sigma)}{\partial\lambda_1}\right| \leq G_{100}(y) \ . \tag{E.17}$$

Let, $M_{100}(y) = G_{100}(y)\, f^*(y)$ then from (E.6) ,(E.11) and (E.17)

$$\left|\frac{\partial f_Y(y,\lambda_1,\lambda_2,\sigma)}{\partial\lambda_1}\right| \leq M_{100}(y) \qquad \text{for all } \omega \in I \tag{E.18}$$

We now show that $\int_{-\infty}^{\infty} G_{100}(y)\, f^*(y)\, dy < \infty$

$$\int_{-\infty}^{\infty} G_{100}(y)\, f^*(y)\, dy = \int_{-\infty}^{\infty} \{\ln(\theta_0-Y+1) + (1/a_1a_3^2)$$

$$[(\frac{1-(\theta_0-y+1)^{b_1}}{a_1})^2 - (\frac{1-(\theta_0-y+1)^{b_1}}{a_1})\,(\theta_0-Y+1)^{b_1}\,\ln(\theta_0-Y+1)]\}$$

$$[\tau/a_3\, J(\, b_1-a_1+1\, , b_2-a_2+1\,)]\, f_Y(y,\lambda_1,\lambda_2,\sigma)\, dy$$

Note that $G_{100}(y)$ is defined over $(-\infty,\theta_0)$ hence , the integral over the positive co-domain will vanish. Hence,

$$\int_{-\infty}^{\infty} G_{100}(y)\, f^*(y)\, dy = \int_{-\infty}^{0} \{\ln(\theta_0-Y+1) + (1/a_1a_3^2)$$

$$[(\frac{1-(\theta_0-y+1)^{b_1}}{a_1})^2 - (\frac{1-(\theta_o-y+1)^{b_1}}{a_1})\,(\theta_0-Y+1)^{b_1}\,\ln(\theta_0-Y+1)]\}$$

$$[\tau/a_3(\theta_0-Y+1)^{b_1-a_1}](\theta_0-Y+1)^{a_1-1}\exp\{-1/2\tau^2(\frac{1-(\theta_0-y+1)^{a_1}}{a_1})^2\}/\sqrt{2\pi\tau^2}\, dy$$

then from (E.13),

$$\int_{-\infty}^{\infty} G_{100}(y)\, f^*(y)\, dy = \int_{-\infty}^{0} \{\ln(1-a_1x)/a_1 +$$

$$(\tau/a_1a_3^3)[(\frac{(1-a_1x)^{b_1/a_1}-1}{a_1})^2 + (\frac{(1-a_1x)^{b_1/a_1}-1}{a_1})\,(1-a_1x)^{b_1/a_1}$$

$$\ln(1-a_1x)/a_1]\}\,(1-a_1x)^{b_1/a_1-1}\quad \exp\{-x^2/2\tau^2\}/\sqrt{2\pi\tau^2}\, dx$$

Since $a_1 > 0$ and $X < 0$ imply $(1-a_1x) > 1$ and $\ln(1-a_1x) < (1-a_1x)$ then by the assumption that $b_1/a_1 \leq 2$

$$\int_{-\infty}^{\infty} G_{100}(y)\, f^*(y)\, dy \leq \int_{-\infty}^{0} (1-a_1x)/a_1 + (\tau/a_1^3a_3^3)$$

$$[((1-a_1x)^2-1)(1-a_1x)^3 + (1-a_1x)^2-1)^2]\,(1-a_1x)$$

$$\exp\{-x^2/2\tau^2\}/\sqrt{2\pi\tau^2}\quad dx \tag{E.19}$$

The right hand side of (E.19) is the expectation of a polynomial of degree 6 defined over the negative region of a normal random variable and hence is finite. Therefore,

$$\int_{-\infty}^{\infty} M_{100}(y)\,dy = \int_{-\infty}^{\infty} G_{100}(y)\, f^*(y)\,dy < \infty \qquad \text{for all } \omega \in I. \tag{E.20}$$

3. An integrable upper bound for $\left|\dfrac{\partial L(\lambda_1, \lambda_2, \sigma)}{\partial \lambda_2}\right|$

From (E.5)

$$\left|\frac{\partial L(\lambda_1,\lambda_2,\sigma)}{\partial \lambda_2}\right| = \Big\{\ln(y-\theta_0+1) + \Big[\Big(\frac{(y-\theta_0+1)^{\lambda_2}-1}{\lambda_2}\Big)(y-\theta_0+1)^{\lambda_2}/(\lambda_2\sigma^2)$$

$$\ln(y-\theta_0+1) + \Big(\frac{(y-\theta_0+1)^{\lambda_2}-1}{\lambda_2}\Big)^2\Big]\Big\}\, I_{(\theta_0,\infty)}(y)$$

Let

$$G_{010}(y) = \ln(y-\theta_0+1) + \Big\{(1/a_2a_3^2)\ \Big[\Big(\frac{(y-\theta_0+1)^{b_2}-1}{a_2}\Big)$$

$$(y-\theta_0+1)^{b_2}\ln(y-\theta_0+1) + \Big(\frac{(y-\theta_0+1)^{b_2}-1}{a_2}\Big)^2\Big]\Big\}\, I_{(\theta_0,\infty)}(y) \tag{E.22}$$

which implies that

$$\left|\frac{\partial L(\lambda_1,\lambda_2,\sigma)}{\partial \lambda_2}\right| \le G_{010}(y)\ . \tag{E.23}$$

Let, $M_{010}(y) = G_{010}(y)\, f^*(y)$ then from (E.6) ,(E.11) and (E.23)

$$\left|\frac{\partial f_Y(y,\lambda_1,\lambda_2,\sigma)}{\partial \lambda_2}\right| \le M_{010}(y) \qquad \text{for all } \omega \in I\ . \tag{E.24}$$

From a similar argument like that given in 1 , it can be shown that,

$$\int_{-\infty}^{\infty} M_{010}(y)\,dy = \int_{0}^{\infty} G_{010}(y)\, f^*(y)\,dy < \infty \qquad \text{for all } \omega \in I$$

4. An integrable upper bound for $\left|\dfrac{\partial L(\lambda_1, \lambda_2, \sigma)}{\partial \sigma}\right|$

From (E.5)

$$\left|\frac{\partial L(\lambda_1,\lambda_2,\sigma)}{\partial \sigma}\right| = 1/\sigma + 1/\sigma^3 \ \{[(\frac{1-(\theta_0-y+1)^{\lambda_1}}{\lambda_1})^2 \ I_{(-\infty,\theta_0)}(y) \ +$$

$$(\frac{(y-\theta_0+1)^{\lambda_2}-1}{\lambda_2})^2 I_{(\theta_0,\infty)}(y)]\} \tag{E.25}$$

Let,

$$G_{001}(y) = 1/a_3 + 1/a_3^3 \ \ [(\frac{1-(\theta_0-y+1)^{b_1}}{a_1})^2 +$$

$$(\frac{(y-\theta_0+1)^{b_2}-1}{a_2})^2 I_{(\theta_0,\infty)}(y) \ ] \ \} \tag{E.26}$$

and let, $M_{001}(y) = G_{001}(y) \ f^*(y)$. Then

$$\left|\frac{\partial L(\lambda_1,\lambda_2,\sigma)}{\partial \sigma}\right| \le G_{001}(y) \tag{E.27}$$

By (E.6) , (E.11) and (E.27),

$$\left|\frac{\partial f(y,\lambda_1,\lambda_2,\sigma)}{\partial \sigma}\right| \le M_{001}(y) \tag{E.28}$$

We next show that, $\int_{-\infty}^{\infty} M_{001}(y)\, dy = \int_{-\infty}^{\infty} G_{001}(y) f^*(y)\, dy < \infty$.

$$\int_{-\infty}^{\infty} G_{001}(y)f^*(y)\ dy = 1/a_3 \int_{-\infty}^{\infty} f^*(y)dy + (1/\sqrt{2\pi\tau^2}) \ \tau/a_3^4$$

$$\{ \int_{-\infty}^{0} (\frac{1-(\theta_0-y+1)^{b_1}}{a_1})^2 \ (\theta_0-y+1)^{b_1-1} \ \exp[-(\frac{1-(\theta_0-y+1)^{a_1}}{a_1})^2/2\tau^2]dy \ +$$

$$\int_0^{\infty} (\frac{(y-\theta_0+1)^{b_2}-1}{a_2})^2 \ (y-\theta_0+1)^{b_2-1} \ \exp\ [-(\frac{(y-\theta_0+1)^{a_2}-1}{a_2})^2/2\tau^2 \ ] \ dy \ \}$$

under the variable transformation defined in (E.13)

$$\int_{-\infty}^{\infty} G_{001}(y)f^*(y)\ dy = 1/a_3 \int_{-\infty}^{\infty} f^*(y)\ dy \ +$$

$$\tau/a_3^4\{ \int_{-\infty}^{0} (1-(1-a_1x)^{b_1/a_1})^2/a_1^2 \ (1-a_1x)^{b_1/a_1-1} \ \exp\{-x^2/2\}\ dx$$

$$+\int_0^{\infty} ((1+a_2x)^{b_2/a_2}-1)^2/a_2^2 \ (1+a_2x)^{b_2/a_2-1} \ \exp\{-x^2/2\}\ dx \ \}/\sqrt{2\pi\tau^2}$$

By the assumption that $b_1/a_1 \le 2$ and $b_2/a_2 \le 2$,

$$\int_{-\infty}^{\infty} G_{001}(y) f^*(y)\, dy \le 1/a_3 \int_{-\infty}^{\infty} f^*(y)\, dy + \tau/a_3^4 \left\{ \int_{-\infty}^{0} \left(\frac{1-(1-a_1x)^2}{a_1}\right)^2 \right.$$

$$\left. (1-a_1x) \exp\{-x^2/2\}\, dX + {}_0\int^{\infty} \left(\frac{(1+a_2x)^2-1}{a_2}\right)^2 (1+a_2x) \exp\{-x^2/2\}\, dx \right\} / \sqrt{2\pi\tau^2} \tag{E.29}$$

The right hand side of (E.29) is the expectation of a polynomial of degree 5 of a normal random variable and hence it is finite.

5. An integrable upper bound for $\left|\frac{\partial^2 L(\lambda_1,\lambda_2,\sigma)}{\partial\sigma^2}\right|$

From (E.25)

$$\left|\frac{\partial^2 L(\lambda_1,\lambda_2,\sigma)}{\partial\sigma^2}\right| = 3/\sigma \left(\left|\frac{\partial L(\lambda_1,\lambda_2,\sigma)}{\partial\sigma}\right| - 1/\sigma\right) + 1/\sigma^2 .$$

Let $G_{002}(y) = 3/a_3\ (G_{001}(y) - 1/a_3\ ) + 1/a_3^2$ then it follows that,

$$\left|\frac{\partial^2 L(\lambda_1,\lambda_2,\sigma)}{\partial\sigma^2}\right| f_Y(y,\lambda_1,\lambda_2,\sigma) \le G_{002}(y)\ f^*(y) .$$

Let

$$M_{002}(y) = 3/a_3\ (M_{001}(y) - 1/a_3 f^*(y)) + 1/a_3^2\ f^*(y) + G_{001}^2(y) f^*(y) \tag{E.30}$$

From part 3 above, it easy to see that

$$\left|\frac{\partial L(\lambda_1,\lambda_2,\sigma)}{\partial\sigma}\right|^2 f_Y(y,\lambda_1,\lambda_2,\sigma) \le G_{001}^2(y)\ f^*(y)$$

and that $\int_{-\infty}^{\infty} G_{001}^2(y)\ f^*(y) < \infty$. (E.7) and (E.30) imply ,

$$\left|\frac{\partial^2 f(Y,\lambda_1,\lambda_2,\sigma)}{\partial\sigma^2}\right| f(y,\lambda_1,\lambda_2,\sigma) \le M_{002}(y)$$

Since $M_{001}(y)$, $G_{001}^2(y)f^*(y)$ and $f^*(y)$ are integrable then, it follows from (E.30) that $M_{002}(y)$ is integrable.

6. An integrable upper bound for $\left|\dfrac{\partial^2 L(\lambda_1,\lambda_2,\sigma)}{\partial\lambda_1^2}\right|$

From (15) $\left|\dfrac{\partial^2 L(\lambda_1,\lambda_2,\sigma)}{\partial\lambda_1^2}\right| = 1/\sigma^2\{4/\lambda_1^3((\theta_0-Y+1)^{\lambda_1}-1)(\theta_0-Y+1)^{\lambda_1}$

$\ln((\theta_0-Y+1) + 1/\lambda_1^2\ [(\theta_0-y+1)^{\lambda_1}(\ln(\theta_0-y+1))^2+ 2(\theta_0-y+1)^{2\lambda_1}$

$(\ln(\theta_0-y+1))^2 + 3/\lambda_1^4\ ((\theta_0-y+1)^{\lambda_1} - 1)^2\}$ (E.31)

Let,

$$G_{200}(y) = 1/a_3^2\ \{4/a_1^3\ ((\theta_0-y+1)^{b_1} - 1)\ (\theta_0-y+1)^{b_1}\ \ln((\theta_0-y+1)$$

$$+ 1/a_1^2\ [(\theta_0-y+1)^{b_1}(\ln(\theta_0-y+1))^2+ 2(\theta_0-y+1)^{2b_1}(\ln(\theta_0-y+1))^2]$$

$$+ 3/a_1^4\ ((\theta_0-y+1)^{b_1} - 1)^2\} \qquad \text{(E.32)}$$

then,

$$\left|\frac{\partial^2 L(\lambda_1,\lambda_2,\sigma)}{\partial\lambda_1^2}\right| \leq G_{200}(y) \qquad \text{for all } \omega \in I \qquad \text{(E.33)}$$

Using the variable transformation given by (E.13) and the assumption that $a_i/b_i \leq 2$ for $i=1,2$ it can be shown that,

$$\int_{-\infty}^{\infty} G_{200}(y)\ f^*(y)\ dy \leq \int_{-\infty}^{0} (\tau/a_1^4a_3^3)\ \{4((1-a_1x)^2 - 1)\ (1-a_1x)^3 +$$

$$(1-a_1x)^4 + 2\ (1-a_1x)^6 + 3\ ((1-a_1x)^2 - 1)^2\}$$

$$(1-a_1x)\ \exp\{ -x^2/2 \}\ \}dx/\sqrt{2\pi\tau^2} < \infty \qquad \text{(E.34)}$$

From (E.15)it can be shown that,

$$\left|\frac{\partial L(\lambda_1,\lambda_2,\sigma)}{\partial\lambda_1}\right|^2 \leq G_{100}^2(y) \qquad \text{(E.35)}$$

and

$$\int_{-\infty}^{\infty} G_{100}^2(y)\ f^*(y)\ dy \leq \int_{-\infty}^{0} (\tau/a_1^3a_3^3)[\ a_1^2(1-a_1x) + ((1-a_1x)^2-1)$$

$$(1-a_1x)^3 + ((1-a_1x)^2-1)^2]^2\ (1-a_1x)\ \exp\{-x^2/2\tau^2\}\ dx/\sqrt{2\pi\tau^2}$$

$$< \infty \qquad \text{(E.36)}$$

Let

$$M_{200}(y) = G_{200}(y)\ f^*(y) + G_{100}^2(y)\ f^*(y) \tag{E.37}$$

then by (E.7) , (E.33) and (E.35) $\left|\dfrac{\partial^2 f(y,\lambda_1,\lambda_2,\sigma)}{\partial\lambda_1^2}\right| \le M_{200}(y)$ and from

(E.34) and (E.36) $\int_{-\infty}^{\infty} M_{200}(y)\ dy < \infty$.

From a similar argument it can be shown that the absolute value of the second partial derivative of $f(y,\lambda_1,\lambda_2,\sigma)$ differentiated twice with respect to λ_2 is dominated by the integrable function $M_{020}(y)$ where $M_{020}(y)$ is defined as $M_{200}(y)$ given by (E.37) with a_2 and b_2 replacing a_1 and b_1 respectively.

For the third partial derivatives of $f(y,\lambda_1,\lambda_2,\sigma)$ we show for one case only that there exist integrable functions $M_{rst}(y)$ such that

$$\left|\frac{\partial^k f(y,\lambda_1,\lambda_2,\sigma)}{\partial\lambda_1^r\ \partial\lambda_2^s\ \partial\sigma^t}\right| \le M_{rst}(y) \qquad \text{for all } \omega \in I$$

The rest of the cases can be treated similarly, however note that whenever both r and s are different from zero the resulting derivative will be zero due to the multiplication of the two indicator functions $I_{(\theta_0,\infty)}(y)$ and $I_{(\theta_0,\infty)}(y)$.

7. An integrable upper bound for $\left|\dfrac{\partial^3 f(y,\lambda_1,\lambda_2,\sigma)}{\partial\lambda_1^2\ \partial\sigma}\right|$

From (E.31) note that,

$$\left|\frac{\partial^3 f(y,\lambda_1,\lambda_2,\sigma)}{\partial\lambda_1^2\ \partial\sigma}\right| = 2/\sigma\ \left|\frac{\partial^2 L(\lambda_1,\lambda_2,\sigma)}{\partial\lambda_1^2}\right|$$

Let $G_{201}(y) = 1/a_3\ G_{200}(y)$. since $\left|\dfrac{\partial^2 L(\lambda_1,\lambda_2,\sigma)}{\partial\lambda_1^2}\right| \le G_{200}(y)$ and

$1/\sigma < 1/a_3$ hence

$$\left|\frac{\partial^3 L(\lambda_1,\lambda_2,\sigma)}{\partial\lambda_1^2\ \partial\sigma}\right| < G_{201}(y) \ . \tag{E.38}$$

From (E.36)

$$\int_{-\infty}^{\infty} G_{201}(y)\ f^*(y)\ dy = 1/a_3 \int_{-\infty}^{\infty} G_{200}(y)\ f^*(y)\ dy < \infty \tag{E.39}$$

From (E.15) note that,

$$\left|\frac{\partial^2 L(\lambda_1,\lambda_2,\sigma)}{\partial\lambda_1\ \partial\sigma}\right| = 2/\sigma\ \left(\left|\frac{\partial L(\lambda_1,\lambda_2,\sigma)}{\partial\lambda_1}\right| - \ln(\theta_0-y+1)\right)$$

$$\leq 2/\sigma\ \left|\frac{\partial L(\lambda_1,\lambda_2,\sigma)}{\partial\lambda_1}\right| \ .$$

From (E.17) we conclude that,

$$\left|\frac{\partial^2 L(\lambda_1,\lambda_2,\sigma)}{\partial\lambda_1\ \partial\sigma}\right| \leq 2/a_3\ G_{100}(y) \tag{E.40}$$

From (E.19)

$$G_{100}(y) \leq \{(1-a_1x)/a_1 + 1/a_1^3[((1-a_1x)^2-1)(1-a_1x)^3+ ((1-a_1x)^2-1)^2]\}I_{(-\infty,0)}(x) \tag{E.41}$$

where x is given by (E.13) . From (E.29)

$$G_{001}(y) \leq 1/a_3 + 1/a_3^3\ \{(\frac{1-(1-a_1x)^2}{a_1})^2\ I_{(-\infty,0)}(x) + (\frac{(1+a_2x)^2-1}{a_2})^2\ I_{(0\ ,\infty)}(x)\ \} \tag{E.42}$$

From (E.11) , (E.27) and (E.40) ,

$$\left|\frac{\partial^2 L(\lambda_1,\lambda_2,\sigma)}{\partial\lambda_1\ \partial\sigma}\right|\ \left|\frac{\partial L(\lambda_1,\lambda_2,\sigma)}{\partial\sigma}\right|\ f(y,\lambda_1,\lambda_2,\sigma)$$

$$\leq 2/a_3\ G_{100}(y)\ G_{001}(y)\ f^*(y) \tag{E.43}$$

Note that the integration of the product of the right hand sides of (E.14) (E.41) and (E.42) under the normality of x represents the expectation of a polynomial of degree 10 of a normal random variable. Hence,

$$\int_{-\infty}^{\infty} 2/a_3 \; G_{100}(y) \; G_{001}(y) \; f^*(y) \; dy < \infty \tag{E.44}$$

From (E.17) since $\left|\dfrac{\partial L(\lambda_1,\lambda_2,\sigma)}{\partial\lambda_1}\right| \leq G_{100}(y)$. Hence,

$$\left|\frac{\partial L(\lambda_1,\lambda_2,\sigma)}{\partial\lambda_1}\right|^2 \leq G_{100}^2(y) \tag{E.45}$$

(E.11) , (E.27) and (E.45) imply that,

$$\left|\frac{\partial L(\lambda_1,\lambda_2,\sigma)}{\partial\lambda_1}\right|^2 \left|\frac{\partial L(\lambda_1,\lambda_2,\sigma)}{\partial\sigma}\right| f(y,\lambda_1,\lambda_2,\sigma) \leq G_{100}^2(y) \; G_{001}(y) \; f^*(y) \tag{E.46}$$

From (E.41)

$$G_{100}^2(y) \leq \{(1-a_1x)/a_1 + 1/a_1^3 \; [((1-a_1x)^2-1)(1-a_1x)^3 + ((1-a_1x)^2-1)^2]\}^2 \; . \tag{E.47}$$

Note that the integration of the product of the right hand sides of (E.14) (E.46) and (E.47) under the normality of x represents the expectation of a polynomial of degree 15 of a normal random variable. Hence,

$$\int_{-\infty}^{\infty} G_{100}^2(y) \; G_{001}(y) \; f^*(y) \; dy < \infty \tag{E.48}$$

From (E.11) , (E.27) and (E.33),

$$\left|\frac{\partial^2 L(\lambda_1,\lambda_2,\sigma)}{\partial\lambda_1^2}\right| \left|\frac{\partial L(\lambda_1,\lambda_2,\sigma)}{\partial\sigma}\right| f(y,\lambda_1,\lambda_2,\sigma) \leq G_{200}(y) \; G_{001}(y) \; f^*(y) \tag{E.49}$$

From (E.34) it can be seen that,

$$G_{200}(y) \leq 1/a_1^4 \; \{4((1-a_1x)^2 - 1) \; (1-a_1x)^3 + (1-a_1x)^4 + 2 \; (1-a_1x)^6 \; ((1-a_1x)^2 - 1)^2\} \tag{E.50}$$

Note that the integration of the product of the right hand sides of (E.14), (E.42) and (50) under the normality of x represents the expectation of a polynomial of a normal random variable. Hence,

$$\int_{-\infty}^{\infty} G_{200}(y)\ G_{001}(y)\ f^{*}(y)\ dy < \infty . \tag{E.51}$$

Let,

$$M_{201}(y) = \{ G_{201}(y) + [4/a_3\ G_{100}(y) + G_{100}^2(y) + G_{200}(y)]\ G_{001}(y) \} f^{*}(y)$$

then from (E.8), (E.38), (E.43), (E.46) and (E.49) we conclude that,

$$\left| \frac{\partial^3 f(y,\lambda_1,\lambda_2,\sigma)}{\partial\lambda_1^2\ \partial\sigma} \right| \le M_{201}(y)$$

and from (E.39), (E.44), (E.48) and (E.51) we conclude that,

$$\int_{-\infty}^{\infty} M_{201}(y)\ dy < \infty .$$

This concludes Lemma E.1 . []

Lemma E.2 For some $\omega \in I$ where I is as defined in (E.1) let $U(\omega)$ denote the score vector of ω and let $I(\omega)$ denote the information matrix of ω that is $I(\omega) = \text{var}(U(\omega))$ then,

i. $E_{\omega}(U(\omega)) = 0$ ii. $I(\omega) = E \left\{ \dfrac{-\partial^2 L(\lambda_1,\lambda_2,\sigma)}{\partial\omega\ \partial\omega^t} \right\}$

Proof

By THeorem 10.3 of K.T. Smith (1971, p.330) and by Lemma E.1, the first and second derivatives of $f_Y(y,\lambda_1,\lambda_2,\sigma)$ can be obtained under the integral sign. The results are then immediate from Lemma 2.6.1 of Lehmann (1983, p.118).

Lemma E.3 Let $\dfrac{\partial L}{\partial\lambda_1}$, $\dfrac{\partial L}{\partial\lambda_2}$ and $\dfrac{\partial L}{\partial\sigma}$ denote the first partial derivatives of the loglikelihood defined in (E.5) then these partial derivatives are affinely independent with probability 1 .

Proof

From (E.5) we get,

$$\frac{\partial L}{\partial \lambda_1} = \{(1/\lambda_1 \sigma^2)\ [\ (\frac{1-(\theta_0-y+1)^{\lambda_1}}{\lambda_1})\ (\theta_0-y+1)^{\lambda_1}\ln(\theta_0-y+1)$$

$$+ (\frac{1-(\theta_0-y+1)^{\lambda_1}}{\lambda_1})^2] + \ln(\theta_0-y+1)\}\ I_{(-\infty,\theta_0)}$$

$$\frac{\partial L}{\partial \lambda_2} = \{1/(\lambda_2 \sigma^2)\ [\ -\ (\frac{(y-\theta_0+1)^{\lambda_2}-1}{\lambda_2})\ (y-\theta_0+1)^{\lambda_2}\ln(y-\theta_0+1)$$

$$+ (\frac{(y-\theta_0+1)^{\lambda_2}-1}{\lambda_2})^2] + \ln(y-\theta_0+1)\}\ I_{(\theta_0,\infty)}(y)$$

$$\frac{\partial L}{\partial \sigma} = -1/\sigma + 1/\sigma^3\ \{[(\frac{1-(\theta_0-y+1)^{\lambda_1}}{\lambda_1})^2 I_{(-\infty,\theta_0)}(Y) +$$

$$(\frac{(y-\theta_0+1)^{\lambda_2}-1}{\lambda_2}) I_{(\theta_0,\infty)}(Y)]\}\ .$$

Under the variable change given in (E.13) with λ_1 and λ_2 in place of a_1 and a_2 respectively, the above derivatives can be written in terms of x as

$$\frac{\partial L}{\partial \lambda_1} = \{(1/\lambda_1 \sigma^2)\ [x(1-\lambda_1 x)\ \ln(1-\lambda_1 x)/\lambda_1 + x^2]$$

$$+ \ln(1-\lambda_1 x)/\lambda_1\}\ I_{(-\infty,\ 0)}(x) \tag{E.52}$$

$$\frac{\partial L}{\partial \lambda_2} = \{(1/\lambda_2 \sigma^2)\ [-x(1+\lambda_2 x)\ \ln(1+\lambda_2 x)/\lambda_2 + x^2]$$

$$+ \ln(1+\lambda_2 x)/\lambda_2\ \} I_{(0,\infty)}(x) \tag{E.53}$$

$$\frac{\partial L}{\partial \sigma} = -1/\sigma + 1/\sigma^3\ x^2 \tag{E.54}$$

To show that the above derivatives are affinely independent with probability 1 we show that, for any real numbers a_0 , a_1 a_2 and a_3

$$\Pr\ \{a_0 + a_1\frac{\partial L}{\partial \lambda_1} + a_2\frac{\partial L}{\partial \lambda_2} + a_3\frac{\partial L}{\partial \sigma} = 0\} = 0$$

unless $a_0 = a_1 = a_2 = a_3 = 0$.

$$\Pr\{a_0 + a_1\frac{\partial L}{\partial \lambda_1} + a_2\frac{\partial L}{\partial \lambda_2} + a_3\frac{\partial L}{\partial \sigma} = 0\} =$$
$$\Pr\{a_0 + a_1\frac{\partial L}{\partial \lambda_1} + a_3\frac{\partial L}{\partial \sigma} = 0 \text{ and } x < 0\} +$$
$$\Pr\{a_0 + a_2\frac{\partial L}{\partial \lambda_2} + a_3\frac{\partial L}{\partial \sigma} = 0 \text{ and } x > 0\}$$

On the domain $\{ x: x < 0 \}$ and from (E.52) and (E.54) let,

$$g(x) = a_0 + a_1 \frac{\partial L}{\partial \lambda_1} + a_3 \frac{\partial L}{\partial \sigma}$$
$$= a_0 + a_1 \{(1/\lambda_1\sigma^2) [x (1-\lambda_1 x) \ln(1-\lambda_1 x)/\lambda_1 + x^2]$$
$$+ \ln(1-\lambda_1 x)/\lambda_1\} + a_3 \{-1/\sigma + 1/\sigma^3 x^2\}$$

hence,

$$g(x) = (a_0 - a_3/\sigma) + (a_1/\lambda_1\sigma^2 + a_3/\sigma^3) x^2$$
$$+ (a_1/\lambda_1^2\sigma^2) x (1-\lambda_1 x) \ln(1-\lambda_1 x)/\lambda_1 + a_1/\lambda_1 \ln(1-\lambda_1 x)$$

Note that $g(x)$ is an analytic function on the domain $\{ x: x < 0 \}$ hence if $g(x)$ is not identically 0, then the set $\{x: g(x) = 0\}$ is countable and hence has Lebesgue measure 0 . Since x has a density with respect to Lebsegue measure hence,

$$\Pr \{g(x) = 0\} = 0$$

Now suppose $g(x)$ is identically 0 . Expand $g(x)$ in a power series in a neighborhood of $x = 0$. Its coefficients must all be 0.

$$\ln(1-\lambda_1 x) = -\lambda_1 x - \lambda_1^2/2\ x^2 - \lambda_1^3/3\ x^3 - \ldots$$
$$x(1-\lambda_1 x)\ln(1-\lambda_1 x) = -\lambda_1 x^2 + (\lambda_1^2-\lambda_1^2/2)x^3 + (\lambda_1^3/2-\lambda_1^3/3)x^4 + \ldots$$

hence,

$$g(x) = (a_0 - a_3/\sigma) + a_1/\lambda_1(-\lambda_1)x + [a_1/\lambda_1\sigma^2 + a_3/\sigma^3 +$$
$$(a_1/\lambda_1^2\sigma^2)(-\lambda_1) + a_1/\lambda_1(-\lambda_1^2/2)]\ x^2 + \ldots]$$

Since

$$a_1/\lambda_1(-\lambda_1) = 0 \quad \text{imply} \quad a_1 = 0$$
$$[a_1/\lambda_1\sigma^2 + a_3/\sigma^3 + (a_1/\lambda_1^2\sigma^2)(-\lambda_1) + a_1/\lambda_1(-\lambda_1^2/2)] = 0 \quad \text{imply}$$
$$a_3 = 0 \quad a_3 = 0 \quad \text{and} \quad (a_0 - a_3/\sigma) = 0 \quad \text{imply} \quad a_0 = 0$$

From a similar argument on the domain $\{ x: x > 0 \}$ we get $a_2 = 0$. Therefore,

$$Pr\{ g(x) = 0 \} = 0 \quad \text{unless} \quad a_0 = a_1 = a_2 = a_3. \qquad []$$

Lemma E.4

Let the pdf of the random variable Y (assume $\theta_0 = 0$) be given as

$$f_Y(y,\lambda_1,\lambda_2,\sigma) = (2\pi\sigma^2)^{-1/2} \exp\{-1/2\sigma^2 [(\frac{1-(1-y)^{\lambda_1}}{\lambda_1})^2 I_{(-\infty,0)}(y) + (\frac{(y+1)^{\lambda_2}-1}{\lambda_2})^2 I_{(0,\infty)}(y)]\} [(1-y)^{\lambda_1-1} I_{(-\infty,0)}(y) + (y+1)^{\lambda_2-1} I_{(0,\infty)}(y)]$$

and let Ω denote the parameter space defined in Section 4.1 . If ω' and ω'' are any two points in Ω such that $f_Y(y,\omega') = f_Y(y,\omega'')$ then $\omega' = \omega''$.

Proof

Over the domain $\{y: y > 0\}$ suppose that $f_Y(y,\omega') = f_Y(y,\omega'')$. Then

$$\sigma'^{-1} \exp\{-1/2\sigma'^2 (\frac{(1+y)^{\lambda_2'} - 1}{\lambda_2'})^2\}(1+y)^{\lambda_2'-1} =$$

$$\sigma''^{-1} \exp\{-1/2\sigma''^2 (\frac{(1+y)^{\lambda_2''} -1}{\lambda_2''})^2 \} (1+y)^{\lambda_2''-1}$$

that is,

$$-\ln(\sigma'/\sigma'') + (\lambda_2'- \lambda_2'') \ln(1+y) - 1/2\sigma'^2[(\frac{(1+y)^{\lambda_2'} - 1}{\lambda_2'})^2 - (\frac{(1+y)^{\lambda_2''} -1}{\lambda_2''})^2]$$

$$= 0 \qquad \text{(E.55)}$$

Expand $\ln(1+y)$, $(1+y)^{\lambda_2'}$ and $(1+y)^{\lambda_2''}$ as a power series. For $|y| < 1$ we get,

$$-\ln(\sigma'/\sigma'') + (\lambda_2'- \lambda_2'') [y - y^2/2 + y^3/3 + \ldots] - 2/\sigma'^2 [\lambda_2' y + \lambda_2'(\lambda_2'-1)/2 y^2+\ldots]^2 - 2/\sigma''^2 [\lambda_2'' y + \lambda_2'' (\lambda_2''-1)/2 y^2+\ldots]^2 = 0 .$$

For these series to be 0 , the coefficients of y^r must be zeros for $r = 0,1,2,\ldots$. For $r = 0$, $-\ln(\sigma'/\sigma'') = 0$ iff $\sigma' = \sigma''$.
For $r = 1$, $(\lambda_2' - \lambda_2'')\, y = 0$ iff $\lambda_2' = \lambda_2''$.

From a similar argument over the set $\{ Y: Y < 0 \}$ we can show that $\lambda_1' = \lambda_1''$. Therefore $\omega' = \omega''$

[]

<u>Theorem E.1</u>

Let $\hat{\underline{\lambda}}_n$ denote the maximum likelihood estimator of $\underline{\lambda}$ under the two-domain family of transformations. Let $I(\omega)$ be as defined in Lemma E.2 above and let $I^{\lambda\lambda}$ denote the upper (2x2) block diagonal matrix of $I^{-1}(\omega)$ then,

i. $\hat{\underline{\lambda}}_n$ is a consistent estimator of $\underline{\lambda}_*$.

ii. $\sqrt{n}\,(\hat{\underline{\lambda}}_n - \underline{\lambda}_*) \xrightarrow{\mathcal{L}} N(0, I^{\lambda\lambda})$.

<u>Proof</u>

The proof follows from Theorem 6.4.1 of Lehmann (1983, p.429) if we show that the regularity conditions stated in the theorem hold.
Condition (A_0): The distributions P_θ of the observations are distinct follows from Lemma E.4. Conditions (A_1) and (A_2): Under the model $h(y-\theta_0, \lambda_1, \lambda_2) = \sigma\epsilon$ with ϵ assumed to have a standard normal distribution we get the support of the distribution of Y the whole real line and the observations $Y_1, Y_2, \ldots, Y_n$ are iid with pdf with respect to Lebesegue measure.
Condition(A): As we claimed before every point in Ω can be made a point of an open rectangle contained in Ω . In particular this is true for the true parameter point ω_* . Also since $f_Y(y, \lambda_1, \lambda_2, \sigma)$ is differentiable with respect to λ_1 , λ_2 and σ to any order for all Y and all ω hence, all the third partial derivatives exist for ω in an

open subset of Ω containing ω_* .

Condition (B): Follows from Lemma E.2

Condition (C): Follows from Lemma E.3

condition (D): follows from Lemma E.1 .

[]

Lemma E.5

Let $\hat{\omega}_n$ denote the maximum likelihood estimator of ω where, $\omega = (\lambda_1, \lambda_2, \sigma)^t$ and let $U_1(\omega)$ and $I(\omega)$ denote, respectively the score vector and the information matrix of ω for one observation. Then, $\sqrt{n}(\hat{\omega}_n - \omega_*) - I^{-1}(\omega_*)\ (1/\sqrt{n}\, U_n(\omega_*)) \xrightarrow{P} 0$

Proof

In the proof ω_1 , ω_2 and ω_3 will be used to mean λ_1 , λ_2 and σ respectively. Also if $L_n(\omega) = \sum_{i=1}^{n} \ln(\ f_{Y_i}(y_i, \omega)\ )$ then $L_n^r(\omega)$, $L_n^{rs}(\omega)$ and $L_n^{rst}(\omega)$ are used to denote, respectively the first partial derivative of the loglikelihood with respect to the r^{th} component of ω , the second partial derivative with respect to the r and s^{th} components and the third partial derivative with respect to the r , s and t^{th} components of ω . Consider a Taylor expansion of $L_n^r(\hat{\omega}_n)$ about ω_* . Since by definition $L_n^r(\hat{\omega}_n) = 0$, then we get

$$L_n^r(\omega_*) + \sum_{s=1}^{3} (\hat{\omega}_{ns} - \omega_{*s})\ L_n^{rs}(\omega_*) + 1/2 \sum_{s=1}^{3} \sum_{t=1}^{3} (\hat{\omega}_{nt} - \omega_{*t}) (\hat{\omega}_{ns} - \omega_{*s}) L_n^{rst}(\tilde{\omega}_n) = 0 \qquad \text{(E.56)}$$

where $\tilde{\omega}_n$ is such that $|\tilde{\omega}_{ns} - \omega_{*s}| \le |\hat{\omega}_{ns} - \omega_{*s}|$ for s =1,2,3. Note that according to our notation $U(\omega_*) = (L_n^1(\omega_*), L_n^2(\omega_*), L_n^3(\omega_*))^t$. Let $L_n^{**}(\omega)$ be the 3x3 matrix with (r,s) entry $L_n^{rs}(\omega)$ and let $L_n^{**t}(\omega)$ be the 3x3 matrix with (r,s) entry $L_n^{rst}(\omega)$. Now equation (E.56) can be written as

$$U_n(\omega_*) + L_n^{**}(\omega)(\hat{\omega}_n-\omega_*) + 1/2\sum_{t=1}^{3}(\hat{\omega}_{nt}-\omega_{*t})L_n^{**t}(\bar{\omega}_n)(\hat{\omega}_n-\omega_*) = 0 \tag{E.57}$$

From (E.57) we can write

$$1/\sqrt{n}\, U_n(\omega_*) + H_n \sqrt{n}\,(\dot{\omega}_n - \omega_*) = 0 \tag{E.58}$$

where

$$H_n = (L_n^{**}(\omega)/n + 1/2 \sum_{t=1}^{3}(\dot{\omega}_{nt} - \omega_{*t})\, L_n^{**t}(\bar{\omega}_n)/n) \tag{E.59}$$

For a matrix $A = \{a_{ij}\}$ let $|A|$ denote the maximum of the absolute values of the elements A that is

$$|A| = \max_{i,j} \{|a_{ij}|\}.$$

Then it follows that

$$\sum_{t=1}^{3}(\dot{\omega}_{nt} - \omega_{*t})\, L_n^{**t}(\bar{\omega}_n)/n \le \sum_{t=1}^{3}(\hat{\omega}_{nt} - \omega_{*t})\, |L_n^{**t}(\bar{\omega}_n)|\, /n$$

Since by Theorem (E.1) $\hat{\omega}_n \xrightarrow{P} \omega_*$ then there exist an open rectangle V and n_0 such that for some $\delta > 0$ and for all $n > n_0$

$$Pr_{\omega_*}\{\hat{\omega}_n \in V\} > 1 - \delta.$$

Since $|\bar{\omega}_{ns} - \omega_{*s}| \le |\hat{\omega}_{ns} - \omega_{*s}|$ for $s = 1,2,3$. Hence $\bar{\omega}_n \in V$.

Therefore with probability exceeding $1-\delta$ we have

$$|L_n^{**t}(\bar{\omega}_n)| \le \sup_{\omega \in V} |L_n^{**t}(\omega)| \le \sum_{r=1}^{3}\sum_{s=1}^{3} \sup_{\omega \in V} |L_n^{rst}(\omega)|$$

$$\le \sum_{i=1}^{n}\sum_{r=1}^{3}\sum_{s=1}^{3} \sup_{\omega \in V} \left|\frac{\partial^3 \ln(f(y_i,\omega))}{\partial\omega_r\, \partial\omega_s\, \partial\omega_t}\right| .$$

By the results of Lemma E.1 we have

$$E_{\omega_*}\{\sum_{r=1}^{3}\sum_{s=1}^{3} \sup_{\omega \in V}\left|\frac{\partial^3 \ln(f(y_i,\omega))}{\partial\omega_r\, \partial\omega_s\, \partial\omega_t}\right|\} = C_* < \infty.$$

Hence by the WLLN

$$|L_n^{**t}(\bar{\omega}_n)|/n \le \sum_{i=1}^{n}\sum_{r=1}^{3}\sum_{s=1}^{3} \sup_{\omega \in V}\left|\frac{\partial^3 \ln(f(y_i,\omega))}{\partial\omega_r\, \partial\omega_s\, \partial\omega_t}\right|/n \xrightarrow{P} C_* < \infty$$

Since $\hat{\omega}_{nt} - \omega_{*t} \xrightarrow{p} 0$. Therefore,

$$\sum_{t=1}^{3} (\hat{\omega}_{nt} - \omega_{*t})\, L_n^{**t}(\tilde{\omega}_n)/n \xrightarrow{p} 0 \tag{E.60}$$

Let $L_n^{**}(\omega_*) = \sum_{i=1}^{n} W_i$ where W_i is the 3x3 random matrix with (r,s) entry $\dfrac{\partial^2 \ln(f(y_i, \omega_*))}{\partial\omega_r \partial\omega_s}$. Then by Lemma E.2 we have $E(W_i) = -I(\omega_*)$ and by the WLLN for the vector case (a matrix can be regarded as a double-indexed vector) we get

$$(1/n)\, L_n^{**}(\omega_*) \xrightarrow{p} -I(\omega_*) \tag{E.61}$$

From (E.59) , (E.60) and (E.61) we get,

$$H_n \xrightarrow{p} -I(\omega_*) \tag{E.62}$$

Now the result follows from (E.58) and (E.62).

[]

APPENDIX F

Asymptotic Relative Efficiency and Simulation Programs

A. Asymptotic Relative Efficiency

1. Transformed normal model using John-Draper

```
@ INITIALIZE THE MODEL PARAMETER @
PITEF=ZEROS(8,1);     II=0;     DO WHILE II < 4;     II=II+1;
LAMDA= 1/4|1/3|1/2|1;     L = LAMDA[II,.];     ARE = 0 ; JJ = 0;
DO WHILE JJ < 7;     JJ = JJ + 1;     SIGMA = .25| .5| 1| 2| 3| 4| 5;
SIG = SIGMA [JJ,.];     C1 = 1/(SQRT(2*PI));     A1=SIG*L ;     A2 = 2;
A3 = 1-1/L;     A4 = 1/L;
@ EVALUATE EXPECTATIONS @
PROC MYF(X,A);
RETP((A[.,1]+A[.,2].*X)^(A[.,4]).*EXP(-X^A[.,3]./A[.,3]))*C1;     ENDP;
LB=0;     UB=10;     A=1~A1~A2~A3;     Y = INTSIMP(&MYF,LB,UB,A,1E-8);
PROC VF(X,A);
RETP(((A[.,1]+A[.,2].*X)^A[.,5]-A[.,1])^A[.,3].*EXP(-X^A[.,3]./A[.,3]))
*C1;     ENDP;     LB=0;     UB=10;     A=1~A1~A2~A3~C4;
ZZ=INTSIMP(&VF,LB,UB,A,1E-8);     ZZ=2*ZZ;     E2=(2*Y)^2;
AR= SIG^2/(ZZ*E2);     ARE=ARE|AR;
ENDO;
PITEF=PITEF~ARE;
ENDO;
FORMAT/RZ 9,5;
PITEF;
END;
```

2. Transformed contaminated normal models using John-Draper

```
@ INITIALIZE THE MODEL PARAMETER @

PITEF=ZEROS(8,1);    P=.2;    V1=1;    V2=25;    VY=(1-P)*V1+P*V2;

SDY=SQRT(VY);    VV=SQRT(V2);    II = 0;

DO WHILE II < 4;    II=II+1;

LAMDA=1/4|1/3|1/2|1;    L = LAMDA[II,.];    ARE=0;    JJ=0;

DO WHILE JJ < 7;    JJ=JJ+1;    SIGMA = .25| .5| 1| 2| 3| 4| 5;

SIG = SIGMA[JJ,.]*SDY;    C1=1/(SQRT(2*PI));    C2=SIG*1;    C3=1-1/L;

C4=1/L;    PROC MYF1(X,A);

RETP(A[.,2].*(A[.,1]+A[.,3].*X)^(A[.,5]).*EXP(-(X.*A[.,6])^A[.,4]./

A[.,4]).*A[.,6]);    ENDP;

LB=0;    UB=15;    A=1~C1~C2~2~C3~SDY;    Y1=

INTSIMP(&MYF1,LB,UB,A,1E-8);

PROC MYF2(X,A);

RETP(A[.,2].*(A[.,1]+A[.,3].*X)^(A[.,5]).*EXP(-(X.*A[.,6]./A[.,7])^

A[.,4]./A[.,4]).*A[.,6]./A[.,7]);    ENDP;

LB=0;    UB=15;    A=1~C1~C2~2~C3~SDY~VV;

Y2=INTSIMP(&MYF2,LB,UB,A,1E-8);    PROC VF1(X,A);

RETP(A[.,2].*((A[.,1]+A[.,3].*X)^A[.,6]-A[.,1])^A[.,4].*EXP(-(X.*A[.,7]

)^A[.,4]./A[.,4]).*A[.,7]);    ENDP;    LB=0;    UB=15;

A=1~C1~C2~2~C3~C4~SDY;    Z1=INTSIMP(&VF1,LB,UB,A,1E-8);

PROC VF2(X,A);

RETP(A[.,2].*((A[.,1]+A[.,3].*X)^A[.,6]-A[.,1])^A[.,4]

.*EXP(-(X.*A[.,7]/A[.,8])^A[.,4]./A[.,4]).*A[.,7]./A[.,8]);    ENDP;

LB=0;    UB=15;    A=1~C1~C2~2~C3~C4~SDY~VV;

Z2=INTSIMP(&VF2,LB,UB,A,1E-8);    ZZ=2*((1-P)*Z1+P*Z2);
```

```
E2=(2*((1-P)*Y1+P*Y2))^2;    AR= SIG^2/(ZZ*E2);    ARE=ARE|AR;    ENDO;
PITEF=PITEF~ARE;    ENDO;    FORMAT/RZ 9,5;    PITEF;    END;
```

3. Transformed Student' t model using John-Draper

```
@ INITIALIZE THE MODEL PARAMETER @
VV=10|20|30;    JJ=0;    DO WHILE  JJ < 3;    JJ=JJ+1;    K=VV[jj,1];
B=SQRT(K/(K-2));
@EVALUATION OF THE GAMMA FUNCTION INVOLVED IN THE CONSTANT@
F1=K/2;    IF (F1-FLOOR(F1)) > 0;    NUM=((K-1)/2)!;    DEN=1;
DO WHILE F1 > 1 ;    F1=F1-1;    DEN=DEN*F1;    ENDO;
R=NUM/(DEN*SQRT(PI));    ELSE;    DEN=(K/2-1)!;    F2=(K+1)/2;
NUM=1;    DO WHILE F2 > 1;    F2=F2-1;    NUM=NUM*F2;    ENDO;
R=NUM*SQRT(PI)/DEN;    ENDIF;    C4=-.5*(K+1);    CC=1/SQRT(K*PI);
PITEF=ZEROS(8,1);    II=0;    DO WHILE II < 4;    II=II+1;
LAMDA = .25| 1/3| .5| 1;    L=LAMDA[II,.];    ARE=0;    KK=0;
DO WHILE KK < 7;    KK=KK+1;    SIGMA = .25| .5| 1| 2| 3| 4| 5;
SIG=SIGMA[KK,1]*B;    C1=L*SIG;    C2=1/L;    C3=1-1/L;
PROC  MYF(X,A);
RETP((A[.,1]+A[.,2].*X)^A[.,3].*(A[.,1]+(X.*A[.,7])^A[.,6]/A[.,4])^
A[.,5].*A[.,7]);    ENDP;    LB=0;    UB=50;    A=1~C1~C3~K~C4~2~C5;
Y= INTSIMP(&MYF,LB,UB,A,1E-8);    PROC VF(X,A);
RETP(((A[.,1]+A[.,2].*A)^A[.,6]-A[.,1])^A[.,7].*(A[.,1]+(X.*A[.,8])^
A[.,7]/A[.,4])^A[.,5].*A[.,8]);    ENDP;    LB=0;    UB=50;
A=1~C1~C3~K~C4~C2~2~B;    ZZ=INTSIMP(&VF,LB,UB,A,1E-8);
ZZ=2*CC*R*ZZ;    E2=(2*CC*R*Y)^2;    AR= SIG2/(ZZ*E2);    ARE=ARE|AR;
ENDO;  PITEF=PITEF~ARE;  ENDO;    FORMAT/RZ 9,5;    PITEF;    ENDO; END;
```

4. Transformed normal model using the two-domain family

```
LAMDA= 1/4| 1/3| 1/2| 3/4| 1;
III=0;   DO WHILE III < 5;   III=III+1;   L2=LAMDA[III,.];
PIT1=ZEROS(9,1);   PIT2=ZEROS(9,1);   II=0;   DO WHILE II < 5;
II=II+1;   L1=LAMDA[II,.];   ARE1=0; ARE2=0;   FORMAT/RZ 10,6;   J=0;
DO WHILE J < 8;   J=J+1;   SIGMA= .1| .25| .5| 1| 2| 3| 4| 5;
SIG=SIGMA[J,.];   C1=SQRT(2*PI);   A1=1;   A21=L1*SIG;
A22=L2*SIG; A3=2;   A41=1-1/11;   A42=1-1/12;   A51=1/L1;
A52=1/L2;
A=A1~A21~A3~A41~A42~A22~A51~A52;   LB=0;   UB=10;
PROC MM1(X,A);
RETP(((A[.,1]+A[.,2].*X)^A[.,4]+(A[.,1]+A[.,6].*X)^A[.,5]).*
EXP(-X^A[.,3]./A[.,3]));   ENDP;
GG=INTSIMP(&MM1,LB,UB,A,1E-8)/C1;
PROC MY2(X,A);
RETP((((A[.,1]+A[.,2].*X)^A[.,7]-A[.,1])-((A[.,1]+A[.,6].*X)^
A[.,8]-A[.,1])).*EXP(-X^A[.,3]./A[.,3]));   ENDP;
Y1=INTSIMP(&MY2,LB,UB,A,1E-8)/C1;   CLEAR X;
PROC MY3(X,A);
RETP((((A[.,1]+A[.,2].*X)^A[.,7]-A[.,1])^A[.,3]+((A[.,1]+A[.,6].*X)^
A[.,8]-A[.,1])^A[.,3]).*EXP(-X^A[.,3]./A[.,3]));   ENDP;
Y2=INTSIMP(&MY3,LB,UB,A,1E-8)/C1;
AR1= SIG^2/((Y2-(Y1^2))*(GG^2));   ARE1=ARE1|AR1;   LB=-10;   UB=0;
PROC FF1(X,A);
RETP(X.*(A[.,1]-A[.,2].*X).*LN(A[.,1]-A[.,2].*X).*EXP(-X^A[.,3]./
A[.,3]));   ENDP; Z1=INTSIMP(&FF1,LB,UB,A,1E-8)/C1;   CLEAR X;
```

```
PROC FF3(X,A);
RETP(X.*(A[.,1]-A[.,2].*X).*(LN(A[.,1]-A[.,2].*X))^A[.,3]
.*EXP(-X^A[.,3]./A[.,3]));     ENDP;
Z3=INTSIMP(&FF3,LB,UB,A,1E-8)/C1;     CLEAR X;
PROC FF5(X,A);
RETP(((A[.,1]-A[.,2].*X).*LN(A[.,1]-A[.,2].*X)+A[.,2].*X)^A[.,3].*EXP(-
X^A[.,3]./A[.,3]));     ENDP;
Z5=INTSIMP(&FF5,LB,UB,A,1E-8)/C1;     CLEAR X;
PROC SS5(X,A);
RETP(((A[.,1]-A[.,2].*X).*LN(A[.,1]-A[.,2].*X)+A[.,2].*X)
.*EXP(-X^A[.,3]./A[.,3]));     ENDP;
S1=-INTSIMP(&SS5,LB,UB,A,1E-8)/(C1*L1^2);
I11=1/(SIG*L1^3)*(-Z3+2*Z1+L1*SIG)+Z5/(L1^4*SIG^2);
I13=2/(SIG*L1)^2*Z1+1/(SIG*L1);
A=A1~A22~A3;     LB=0;     UB=10;
PROC FF10(X,A); RETP(X.*(A[.,1]+A[.,2].*X).*LN(A[.,1]+A[.,2].*X)
.*EXP(-X^A[.,3]./A[.,3]));     ENDP;
Z10=INTSIMP(&FF10,LB,UB,A,1E-8)/C1;     CLEAR X;
PROC FF12(X,A);
RETP(X.*(A[.,1]+A[.,2].*X).*(LN(A[.,1]+A[.,2].*X))^A[.,3]
.*EXP(-X^A[.,3]./A[.,3]));     ENDP;
Z12=INTSIMP(&FF12,LB,UB,A,1E-8)/C1;     CLEAR X;
PROC FF14(X,A);
RETP(((A[.,1]+A[.,2].*X).*LN(A[.,1]+A[.,2].*X)-A[.,2].*X)^A[.,3].*EXP(-
X^A[.,3]./A[.,3]));     ENDP;
Z14=INTSIMP(&FF14,LB,UB,A,1E-8)/C1;
```

```
PROC SS14(X,A);
RETP(((A[.,1]+A[.,2].*X).*LN(A[.,1]+A[.,2].*X)-A[.,2].*X)
.*EXP(-X^A[.,3]./A[.,3]));    ENDP;
S2=INTSIMP(&SS14,LB,UB,A,1E-8)/(C1*L2^2);
I22=1/(SIG*L2^3)*(Z12-2*Z10+SIG*L2)+Z14/(L2^4*SIG^2);
I23=-2/(SIG*L2)^2*Z10+1/(SIG*L2);
I33=2/SIG^2;    S=S1|S2;    I=(I11~0~I13)|(0~I22~I23)|(I13~I23~I33);
INVI=INV(I);    ILL= INVI[1:2,1:2];    IS=ILL*S;    CLEAR X,A;
A4=SIG;    A51=IS[1,1]/SIG^2;    A52=IS[2,1]/SIG^2;    A61=SIG/L1^2;
A62=SIG/L2^2;    A71=SIG^2/L1;    A72=SIG^2/L2;    A81=IS[1,1]/L1;
A82=IS[2,1]/L2;    LB=0;    UB=10;
A=A1~A22~A3~A4~A52~A62~A72~A82~A21~A51~A61~A71~A81;    PROC UU1(X,A);
RETP((A[.,4].*X-A[.,5].*(A[.,6].*X.*(A[.,1]+A[.,2].*X).*LN(A[.,1]+A[.,2
].*X)-A[.,7].*X^A[.,3])+A[.,8].*LN(A[.,1]+A[.,2].*X))^A[.,3].*EXP(-X^
A[.,3]./A[.,3]));    ENDP;
U1=INTSIMP(&UU1,LB,UB,A,1E-8)/C1;    LB=-10;    UB=0;    PROC UU3(X,A);
RETP((A[.,4].*X+a[.,10].*(a[.,11].*x.*(a[.,1]-a[.,9].*x).*ln(a[.,1]-
a[.,9].*x)+a[.,12].*x^a[.,3])+a[.,13].*ln(a[.,1]-a[.,9].*X))^a[.,3]
.*EXP(-X^A[.,3]./A[.,3]));    ENDP;
U3=INTSIMP(&UU3,LB,UB,A,1E-8)/C1;    AR2=(U1+U3)/SIG^2;
ARE2=ARE2|AR2;    ENDO;    PIT1=PIT1~ARE1;    PIT2=PIT2~ARE2;    ENDO;
FORMAT/RZ 8,4;    PRINT"  LAMDA2 = " L2;
PRINT"  EFFICIENCY OF ORIGINAL TO KNOWN TRANSF. ";  LAMDA';
SSS=0|SIGMA;    SSS~PIT1;
PRINT" EFF. OF KNOWN TRANF. TO UNKNOWN TRANSF. " ;    LAMDA';
SSS~PIT2;    PRINT " EFF. OF ORIG TO UNKNOWN TRANSF. ";
LAMDA';    SSS~(PIT1.*PIT2);    ENDO;   END;
```

B. SIMULATION

1. Program for evaluating the MLE of λ under tHe John-Draper family and power of different test statistics from symmetric models

```
@ INITIALIZE  THE MODEL PARAMETERS @
SIGMA2= ;   AA=0;    DO WHILE AA < 3;    AA=AA+1;    NS= 10| 20| 50;
N=NS[AA,1];   NNK= 250| 250| 100;     NK=NNK[AA,1];     L=   ;
SIG=SQRT(SIGMA2);     UU=-.2;   DO WHILE UU < 1.4;    UU=UU+.2;
MU=SIG*UU;    S=9831815;     IF MU==0;    NSIM=30000;    ELSE; NSIM=5000;
ENDIF;    KSIM=NSIM/NK;     LAM1=0; LAM2=0;     ISIM=0;     OF01=0;
OF05=0;  TF01=0;  TF05=0;  K2=0;  TR11S01=0;  TR11S05=0;  TR21S01=0;
TR21S05=0;    TTR11S01=0;  TTR11S05=0;    TTR21S01=0;  TTR21S05=0;
DO WHILE ISIM < KSIM;


GET THE DATA FROM THE GENERATE THE DATA PART BELOW


J=0;   DO WHILE J<NK;     J=J+1;    Y=YY[.,J]; Y=SORTC(Y,1);
@ TEST AND SIG LEVEL IN ORIGINAL SCALE @
YB=MEANC(Y);    TO=SQRT(N)*YB/STDC(Y);     SLO=CDFTC(TO,(N-1));
IF SLO<.01;     OF01=OF01+1;     ENDIF;    IF SLO<.05; OF05=OF05+1;
ENDIF;
R=.1*N;  R1=R+1;    NR=N-R;     YTR=Y[R1:NR,1];    MTR=MEANC(YTR);
VTR=(SUMC((YTR-MTR)^2)+R*(YTR[1,1]-MTR)^2+R*(YTR[(NR-R),1]-MTR)^2)/
(NR-R1);      TTRM=MTR*SQRT((NR-R)/VTR); SLTR11S=CDFTC(TTRM,(NR-R1));
IF SLTR11S<.01; TR11S01=TR11S01+1; ENDIF;
IF SLTR11S<.05; TR11S05=TR11S05+1; ENDIF;
R=.2*N;     R1=R+1;     NR=N-R;     YTR=Y[R1:NR,1];     MTR=MEANC(YTR);
```

```
VTR=(SUMC((YTR-MTR)^2)+R*(YTR[1,1]-MTR)^2+R*(YTR[(NR-R),1]-MTR)^2)/
(NR-R1);    TTRM=MTR*SQRT((NR-R)/VTR); SLTR21S=CDFTC(TTRM,(NR-R1));
IF SLTR21S<.01; TR21S01=TR21S01+1; ENDIF;
IF SLTR21S<.05; TR21S05=TR21S05+1; ENDIF;
@ ESTIMATION OF LAMDA USING MODULE 10 (NLSYS) @
Z=ABS(Y)+1; SGN=ABS(Y)./Y; LNZ=LN(Z);
CONVTOL=0;   PRNTIT=0;   PRNTOUT=0;   FNAME=&F;    GRADNAME=&GRAD1;
JCO=0;   X0=.5;    VF=ZEROS(1,1);    PROC F(X);    LOCAL X1;
X1=X[1,1];
VF[1,1]=-N*(((Z^X1-1)'( Z^X1.*LNZ-(Z^X1-1)/X1))/(Z^X1-1)'(Z^X1-1))
+SUMC(LNZ);    RETP( VF );    ENDP;
X1=NLSYS(FNAME,X0,JCO,CONVTOL,PRNTIT,PRNTOUT);
LAM1=LAM1+(X1-L); LAM2=LAM2+(X1-L)^2;
@ TRANSFORMED MODEL USING THE MLE X1 @
TV=SGN.* (Z^X1-1)/X1;     TT=SQRT(N)*MEANC(TV)/STDC(TV);
SLT=CDFTC(TT,(N-1));    IF SLT<.01;    TF01=TF01+1;    ENDIF;
IF SLT<.05;    TF05=TF05+1;    ENDIF;   TV=SORTC(TV,1);
R=.1*N;   R1=R+1;   NR=N-R;    TVTR=TV[R1:NR,1];    MTR=MEANC(TVTR);
VTR=(SUMC((TVTR-MTR)^2)+R*(TVTR[1,1]-MTR)^2+R*(TVTR[(NR-R),1]-MTR)^2)
/(NR-R1);    TTRM=MTR*SQRT((NR-R)/VTR);SLTR11S=CDFTC(TTRM,(NR-R1));
IF SLTR11S<.01;    TTR11S01=TTR11S01+1;    ENDIF;
IF SLTR11S<.05;    TTR11S05=TTR11S05+1;    ENDIF;
R=.2*N;R1=R+1;NR=N-R;TVTR=TV[R1:NR,1]; MTR=MEANC(TVTR);
VTR=(SUMC((TVTR-MTR)^2)+R*(TVTR[1,1]-MTR)^2+R*(TVTR[(NR-R),1]-MTR)^2)
/(NR-R1);    TTRM=MTR*SQRT((NR-R)/VTR);SLTR21S=CDFTC(TTRM,(NR-R1));
IF SLTR21S<.01; TTR21S01=TTR21S01+1; ENDIF;   IF SLTR21S<.05;
TTR21S05=TTR21S05+1; ENDIF;    ENDO;   ISIM=ISIM+1;   ENDO;
```

```
PRINT "      NUMBER OF SAMPLES  WITH SL < .05 ";
PRINT " ORIG   TRAN   OR.1TR   OR.2TR   T.1TR  T.2TR   ";
 OF05~TF05~TR11S05~TR21S05~TTR11S05~TTR21S05;
PRINT "      NUMBER OF SAMPLES  WITH SL < .01 ";
OF01~TF01~TR11S01~TR21S01~TTR11S01~TTR21S01;
PRINT "BIAS = " LAM1/NSIM;
PRINT " MSE = " LAM2/NSIM;    ENDO;    ENDO;    END;
```

● GENERATE THE DATA ●

<u>TRANSFORMED NORMAL DATA</u>

```
A=RNDnS(NK,N,S)';    SGN=ABS(A)./A;  A=ABS(A);
YY=MU+SGN.*(( 1+SIG*L*A)^(1/L)-1);
```

<u>STUDENT'S T DATA WITH NU D.F.</u>

```
NU= ;   VAR = NU/(NU-2);   SIG=SQRT(VAR);    A=RNDNS(NU+1,N*NK,S);
A1=A[1,.];   A2=A[2:NU+1,.];    A3=SUMC(A2^2)/NU;   T=A1'./SQRT(A3);
J=0;    DO WHILE J<NK;    JY=0;    K1=0;    Y=MU+T[J*N+1:(J+1)*N,1];
```

<u>CONTAMINATED NORMAL DATA WITH CONTAMINATION RATIO=P , VARIANCE = SIGMA2</u>

```
SIGMA2=  ;  P= ;    VAR= (1-P) + P*SIGMA2;    SIG=SQRT(VAR);
A=RNDnS(nK,N,S)';    B=RNDUS(NK,N,S)';    J=0;    DO WHILE J < NK;
J=J+1;    JY=0;    K1=0;    Y1=A[.,J];    U1=B[.,J];    NN=0;
DO WHILE NN < N;    NN=NN+1;   IF U1[NN,1]< P;
Y1[NN,1]=SQRT(SIGMA2)*Y1[NN,1];   ENDIF;    ENDO;
Y1=ABS(Y1)./Y1.*((1+L*ABS(y1))^(1/L)-1);    Y=MU+Y1; Y=SORTC(Y,1);
```

2. Program for evaluating the MLE of λ_1 and λ_2 of the two-domain family and power of different test statistics from skewed models

```
SIG=1;    N=20;    NK=250;
```

```
L1=   ;    L2=   ;    AAA=L1|L2;   @ INITIALIZE  THE MODEL PARAMETERS @
UU=-.2;    U1=1;    U2=.2;    DO WHILE UU < U1;    UU=UU+U2;
MU=SIG*UU;
S=9831815;    IF MU==0;    NSIM=30000;    ELSE;    NSIM=5000;    ENDIF;
KSIM=NSIM/NK;    LAM1=0;    LAM2=0;    LAM11=0;    LAM21=0;
ISIM=0;   OF05=0;   TF05=0;   LHSTF05=0 ;   LHST1T05=0;   TTR105=0;
TTR205=0;   LHST2T05=0;   CN1=0;    CN2=0;   Fr1=0; F2=0;   F3=0;
F4=0;   F5=0;   F6=0;   F7=0;   F8=0;   F9=0;   F10=0;   F11=0;
F12=0;   F13=0;   F14=0;   F15=0;   F16=0;   F17=0;   F18=0;
DO WHILE ISIM<KSIM;

GET DATA FROM GENERATE DATA PART

@ TEST AND SIG LEVEL IN ORIGINAL SCALE @
YB=MEANC(Y);   TO=SQRT(N)*YB/STDC(Y);    SLO=CDFTC(TO,(N-1));
IF SLO<.05;    OF05=OF05+1;    ENDIF;
@ ESTIMATION OF LAMDA USING MODULE 10 (NLSYS) @
Y11=Y.*(Y.< 0);    K1=SUMC(Y11./Y);
Y21=Y.*(Y.> 0);    K2=SUMC(Y21./Y) ;
IF K1>=17;   CN1=CN1+1;   LHSTF05=LHSTF05+1;   LHST1T05=LHST1T05+1;
LHST2T05=LHST2T05+1;    GOTO ST;
ELSEIF K2 >=17;   TF05=TF05+1;   TTR105=TTR105+1; TTR205=TTR205+1;
CN2=CN2+1;   GOTO ST;   ENDIF;
Y1=1-Y11;    Y2=Y21+1;    LNY1=LN(Y1);    LNY2=LN(Y2);
CONVTOL=0;    PRNTIT=0;    PRNTOUT=0;    FNAME=&F;   GRADNAME=&GRAD1;
JC0=0;X0=.4|.4;   VF=ZEROS(2,1);    PROC F(X);    LOCAL X1,X2;
X1=X[1,1];    X2=X[2,1];
VF[1,1]=N*((1-Y1^X1)'((Y1^X1).*LNY1+(1-Y1^X1)/X1)/(X1^2))/((1-Y1^X1)'
```

```
    (1-Y1^X1)/(X1^2)+(Y2^X2-1)'(Y2^X2-1)/(X2^2))+SUMC(LNY1);
vf[2,1]=-N*((Y2^X2-1)'((Y2^X2).*LNY2-(Y2^X2-1)/X2)/(X2^2))/((1-Y1^X1)'
    (1-Y1^X1)/(X1^2)+(Y2^X2-1)'(Y2^X2-1)/(X2^2))+SUMC(LNY2);
RETP( VF );   ENDP;   X1=NLSYS(FNAME,X0,JC0,CONVTOL,PRNTIT,PRNTOUT);
L1=X1[1,1];    L2=X1[2,1];    LAM1=LAM1+L1-AAA[1,1];
LAM2=LAM2+L2-AAA[2,1];
LAM11=LAM11 + (L1-AAA[1,1])^2;    LAM21=LAM21+(L2-AAA[2,1])^2;
Y1L=Y1^L1;   Y2L=Y2^L2;   Y1T=(1-Y1L)/L1;   Y2T=(Y2L-1)/L2;
TV=(Y1T+Y2T);
@ TRANSFORMED MODEL USING THE MLE X1 X2 @
TT=SQRT(N)*MEANC(TV)/STDC(TV);    SLT=CDFTC(TT,(N-1));
LHSSL=1-CDFTC(TT,(N-1));    IF SLT<.05;   TF05=TF05+1;    ENDIF;
TV=SORTC(TV,1);    IF MU > 0;   GOTO NEXTST;    ENDIF;
IF LHSSL<.05;    LHSTF05=LHSTF05+1;    ENDIF;
IF TT<-4;  Fr1=Fr1+1;  ELSEIF TT<-3.5;  F2=F2+1;  ELSEIF TT<-3;
F3=F3+1;  ELSEIF TT<-2.5;  F4=F4+1;  ELSEIF TT<-2;  F5=F5+1;  ELSEIF
TT<-1.5;   F6=F6+1;  ELSEIF TT<-1;  F7=F7+1;  ELSEIF TT<-.5; F8=F8+1;
ELSEIF TT < 0;  F9=F9+1;  ELSEIF TT<.5;  F10=F10+1;  ELSEIF TT<1;
F11=F11+1;  ELSEIF TT< 1.5;  F12=F12+1;  ELSEIF TT< 2;  F13=F13+1;
ELSEIF TT< 2.5;  F14=F14+1;  ELSEIF TT< 3;  F15=F15+1;  ELSEIF TT<3.5;
F16=F16+1;  ELSEIF TT< 4;  F17=F17+1; ELSE;  F18=F18+1;  ENDIF;
SSA=SEQA(-4,.5,18);   NEXTST:;
R=.1*N;R1=R+1;NR=N-R;TVTR=TV[R1:NR,1]; MTR=MEANC(TVTR);
VTR=(SUMC((TVTR-MTR)^2)+R*(TVTR[1,1]-MTR)^2+R*(TVTR[(NR-R),1]-MTR)^2)
/(NR-R1);    TTRM=MTR*SQRT((NR-R)/VTR);SLTR11S=CDFTC(TTRM,(NR-R1));
IF SLTR11S<.05;   TTR105=TTR105+1;   ENDIF;
LHSSL=1-CDFTC(TTRM,(NR-R1));   IF LHSSL<.05;   LHST1T05=LHST1T05+1;
```

```
ENDIF;     R=.2*N;R1=R+1;NR=N-R;TVTR=TV[R1:NR,1]; MTR=MEANC(TVTR);
VTR=(SUMC((TVTR-MTR)^2)+R*(TVTR[1,1]-MTR)^2+R*(TVTR[(NR-R),1]-MTR)^2)
/(NR-R1);    TTRM=MTR*SQRT((NR-R)/VTR);     SLTR21S=CDFTC(TTRM,(NR-R1));
IF SLTR21S<.05;     TTR205=TTR205+1;     ENDIF;
LHSSL=1-CDFTC(TTRM,(NR-R1));   IF LHSSL<.05;    LHST2T05=LHST2T05+1;
ENDIF;   ST:;    ENDO;    stp:;    isim=isim+1;    ENDO;
PRINT "       NUMBER OF SAMPLES  WITH SL < .05 ";
PRINT "      ORIG     TRAN          T.1TR    T.2TR    ";
 (OF05~TF05~TTR105~TTR205)/NSIM;
PRINT " NUMBER OF SAMPLES NOT REJECTED BY SIGN TEST = " CN1;
PRINT " NUMBER OF SAMPLES REJECTED BY SIGN TEST = " CN2;
PRINT "BIAS1 MSE1= " ;(LAM1~LAM11)/NSIM;
PRINT "BIAS2 MSE2= " ;(LAM2~LAM21)/NSIM;
IF MU > 0;    GOTO NST;  ENDIF;
PRINT " SIG LEVEL UNDER LHS. TEST " ; LHSTF05~LHST1T05~LHST2T05;
PRINT " FREQUENCY DIST OF THE TRANS. T-STAT. ";
FREQ=(Fr1|F2|F3|F4|F5|F6|F7|F8|F9|F10|F11|F12|F13|F14|F15|F16|F17|F18)/
NSIM;   CUM=ZEROS(18,1);   CUMF=0;K=0;
DO WHILE K<18; K=K+1; CUMF=CUMF+FREQ[K,1]; CUM[K,1]=CUMF;ENDO;
SSA~FREQ~CUM;   NST:;   ENDO;    ENDO;    END;
@ GENERATE THE DATA @
```

<u>TRANSFORMED NORMAL DATA</u>

```
A=RNDnS(nK,N,S)';   J=0;
DO WHILE J< NK; J=J+1; JY=0; KK1=0;
EPS=A[.,J];
L1=AAA[1,1]; L2=AAA[2,1];
EPS1=MU+1-(1-L1*SIG*EPS.*(EPS.<=0))^(1/11);
```

```
EPS2=MU-1+(1+L2*SIG*EPS.*(EPS.>0))^(1/12);

Y=EPS1+EPS2;
```

EXTREME VALUE DATA

```
MED=LN(-LN(.5)); SIG=B*PI/SQRT(6);

A=LN(-LN(1-RNDUS(NK,N,S)'));

J=0;

DO WHILE J<NK; J=J+1;

Y=MU+A[.,J]-MED;
```

GAMA WITH 5 D.F. DATA

```
A=-LN(1-RNDUS(NK,N,S)')-LN(1-RNDUS(NK,N,S)')-LN(1-RNDUS(NK,N,S)')

-LN(1-RNDUS(NK,N,S)')-LN(1-RNDUS(NK,N,S)');

Y=-MU+A[.,J]-MED;
```