A STUDY OF NON-CENTRAL SKEW $T$ DISTRIBUTIONS AND THEIR APPLICATIONS IN DATA ANALYSIS AND CHANGE POINT DETECTION

Abeer M. Hasan

A Dissertation
Submitted to the Graduate College of Bowling Green State University in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2013

Committee:
Arjun K. Gupta, Co-advisor
Wei Ning, Advisor

Mark Earley,
Graduate Faculty Representative

Junfeng Shang.
ABSTRACT

Arjun K. Gupta, Co-advisor
Wei Ning, Advisor

Over the past three decades there has been a growing interest in searching for distribution families that are suitable to analyze skewed data with excess kurtosis. The search started by numerous papers on the skew normal distribution. Multivariate $t$ distributions started to catch attention shortly after the development of the multivariate skew normal distribution. Many researchers proposed alternative methods to generalize the univariate $t$ distribution to the multivariate case. Recently, skew $t$ distribution started to become popular in research. Skew $t$ distributions provide more flexibility and better ability to accommodate long-tailed data than skew normal distributions.

In this dissertation, a new non-central skew $t$ distribution is studied and its theoretical properties are explored. Applications of the proposed non-central skew $t$ distribution in data analysis and model comparisons are studied. An extension of our distribution to the multivariate case is presented and properties of the multivariate non-central skew $t$ distribution are discussed. We also discuss the distribution of quadratic forms of the non-central skew $t$ distribution. In the last chapter, the change point problem of the non-central skew $t$ distribution is discussed under different settings. An information based approach is applied to detect the location of the change point in the non-central skew $t$ distribution. The power of this approach is illustrated via simulation studies. Finally, the change point approach is used to detect the location of the change point in the weekly return rates of three Latin American countries using the non-central skew $t$ distribution.
To my family ... 
To my advisors ... 
To every individual who supported me during my career ...
I dedicate this work.
ACKNOWLEDGMENTS

My graduate school experience was an isolating stage of my life until I met Professor Wei Ning and Professor Arjun Gupta. Having them as my teachers, mentors and advisors was the best thing that happened to me over the past four years. They were willing to guide me whenever I knocked their door for help. They were my family away from home, my counselors and great role models.

I would like to express my gratitude and to convey my deepest respect and appreciation to my mentors Professor Arjun Gupta and Professor Wei Ning for their continuous support and guidance throughout my four years in Bowling Green State University. Their expertise, talent, understanding, support and guidance added considerably to my graduate experience. I would like to thank them for the joyful discussions we had while working on this dissertation. It was a great pleasure for me to work with them and an honor to be their student.

I am grateful to Dr. John Carson, Dr. Reinaldo B. Arellano-Valle, Dr. Luis M. Castro and Dr. Rosangela H. Loschi for providing the data sets that has been used in Chapter 3 and Chapter 4 of this dissertation. Thanks to Dr. Wen-Jang Huang, Dr. Nan-Cheng Su and Dr. Hui-Yi Teng for sharing their personal copy of a paper on the skew t distribution and for providing access to the volcano elevations data and the fiber glass data that have been studied in Chapter 3 of this dissertation.

I am thankful to the members of my dissertation committee, Dr. Junfeng Shang and Dr. Mark Earley for the time they took to review my dissertation. My deepest gratitude goes to the professors who inspired me during graduate school. Thanks to all my colleagues who gave me company during graduate school, especially Adamou, Junvie, Ramadha, Sayonita, Swarup, Wenren, and Ying-ju. Finally, I am mostly grateful to my family for their support, love and patience that provided me with the strength to overcome all the difficult times.
# TABLE OF CONTENTS

## CHAPTER 1: LITERATURE REVIEW

1.1 INTRODUCTION .................................................. 1

1.2 MOTIVATION ..................................................... 1

1.3 THE SKEW NORMAL DISTRIBUTION ............................ 5

1.4 THE MULTIVARIATE SKEW NORMAL DISTRIBUTION ........... 10

1.5 STUDENT’S t DISTRIBUTION ................................... 11

   1.5.1 HISTORICAL NOTES ..................................... 11

   1.5.2 CONSTRUCTION ........................................... 12

   1.5.3 MULTIVARIATE AND NON-CENTRAL t DISTRIBUTIONS ... 15

1.6 LITERATURE REVIEW OF SKEW t DISTRIBUTIONS .......... 16

## CHAPTER 2: A NON-CENTRAL SKEW t DISTRIBUTION

2.1 INTRODUCTION .................................................. 18

2.2 A UNIVARIATE NON-CENTRAL SKEW t DISTRIBUTION .......... 19

   2.2.1 INTRODUCTION .......................................... 19

   2.2.2 PROBABILITY DENSITY FUNCTION OF THE St∗ DISTRIBUTION 20

2.3 PROPERTIES OF THE St∗ DISTRIBUTION ........................ 26

2.4 NUMERICAL RESULTS AND GRAPHICAL ILLUSTRATIONS ....... 34

   2.4.1 INTRODUCTION .......................................... 34

   2.4.2 GRAPHICAL ILLUSTRATION OF THE St∗ DENSITY ........ 34
2.4.3 GENERATING RANDOM SAMPLES FROM THE $St'_r$ DISTRIBUTION

2.5 A MULTIVARIATE NON-CENTRAL SKEW $t$ DISTRIBUTION
   2.5.1 INTRODUCTION
   2.5.2 A MULTIVARIATE NON-CENTRAL SKEW $t$ DISTRIBUTION
   2.5.3 PROPERTIES OF THE $MSt'_k$ DISTRIBUTION

2.6 QUADRATIC FORMS OF THE $St'_r$ DISTRIBUTION

2.7 GENERALIZED NON-CENTRAL SKEW $t$ DISTRIBUTION
   2.7.1 INTRODUCTION
   2.7.2 LITERATURE RESULTS ON THE GENERALIZED SKEW $t$ DISTRIBUTIONS
   2.7.3 GENERALIZED NON-CENTRAL SKEW $t$ DISTRIBUTION

CHAPTER 3: APPLICATIONS TO DATA ANALYSIS AND MODEL COMPARISONS

3.1 INTRODUCTION

3.2 SIMULATION STUDY

3.3 VOLCANO HEIGHTS DATA

3.4 FIBER GLASS DATA

3.5 ENVIRONMENTAL DATA
   3.5.1 THE CONCENTRATION OF ARSENIC IN SOIL SAMPLES
   3.5.2 THE CONCENTRATION OF CADMIUM IN SOIL SAMPLES
   3.5.3 THE CONCENTRATION OF PLUMBUM (LEAD) IN SOIL SAMPLES

3.6 CONCLUSIONS AND RECOMMENDATIONS

CHAPTER 4: AN INFORMATION APPROACH FOR THE CHANGE POINT PROBLEM UNDER THE NON-CENTRAL SKEW $t$ DISTRIBUTION

4.1 INTRODUCTION
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.2 LITERATURE REVIEW OF THE CHANGE POINT PROBLEM</td>
<td>77</td>
</tr>
<tr>
<td>4.3 METHODOLOGY</td>
<td>78</td>
</tr>
<tr>
<td>4.4 SIMULATION STUDY</td>
<td>80</td>
</tr>
<tr>
<td>4.4.1 THE CHANGE OCCURS IN THE SHAPE PARAMETER</td>
<td>80</td>
</tr>
<tr>
<td>4.4.2 THE CHANGE OCCURS IN THE LOCATION PARAMETER ONLY</td>
<td>84</td>
</tr>
<tr>
<td>4.4.3 THE LOCATION AND SCALE PARAMETERS CHANGE SIMULTANEOUSLY</td>
<td>88</td>
</tr>
<tr>
<td>4.5 APPLICATIONS: CHANGE POINT ANALYSIS OF EMERGING MARKETS’ WEEKLY RETURNS</td>
<td>90</td>
</tr>
<tr>
<td>4.5.1 CHILE’S WEEKLY RETURNS DATA</td>
<td>90</td>
</tr>
<tr>
<td>4.5.2 BRAZIL’S WEEKLY RETURNS DATA</td>
<td>96</td>
</tr>
<tr>
<td>4.5.3 ARGENTINA’S WEEKLY RETURNS DATA</td>
<td>97</td>
</tr>
<tr>
<td>4.6 CONCLUDING REMARKS</td>
<td>102</td>
</tr>
<tr>
<td>APPENDIX: SAMPLE OF PROGRAMMING CODE USING R</td>
<td>103</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>113</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

1.1 Density curve of the Student $t$ distribution for different degrees of freedom versus the standard normal density. ................................................................. 3

1.2 Density curve of $SN(0,1,\lambda)$ for different values of $\lambda$ ................................................................. 7

1.3 Density of the ESN distribution with $\gamma = 0$ coincides with the SN density ................................................................. 8

1.4 Density of $ESN$ versus the $SN$ density when $\gamma \neq 0$ ................................................................. 9

2.1 The density of $St_1'(0,1,0)$ coincides with $Cauchy(0,1)$ density. ................................................................. 35

2.2 Graphs of the p.d.f. of $St'_r(0,1,-1)$ for several values of $r$. ................................................................. 35

2.3 Graphs of the p.d.f. of $St'_3(0,1,\alpha)$ for several values of $\alpha$. ................................................................. 36

2.4 Graphs of the p.d.f. of $St'_3(\mu,1,1)$ for several values of $\mu$. ................................................................. 36

2.5 Graphs of the p.d.f. of $St'_3(0,\sigma,1)$ for selected values of $\sigma$. ................................................................. 37

2.6 Graphs of the p.d.f. of $St'_3(0,1,1)$ as $r$ increases versus the $SN(1)$ density. ................................................................. 38

2.7 Graphs of the p.d.f. of $St'_3(2,2,-1)$ as $r$ increases versus the density of $SN(2,2,-1)$. ................................................................. 38

2.8 Graphs of the p.d.f. of $St'_3(0,1,\alpha)$ as $\alpha$ increases they approach the density of $|t_3|$. ................................................................. 39

2.9 Histograms of simulated $St'_r$ data versus theoretical density curve. ................................................................. 40

3.1 Histogram and Model Comparison for the Simulated Inverse Gaussian Data ................................................................. 57

3.2 Histogram and Model Comparison for the Simulated Log Normal Data ................................................................. 58

3.3 Histogram and Model Comparison for the Simulated Log Normal Data ................................................................. 59
3.4 Histogram of the Volcano Elevations Data .................................................................... 60
3.5 Normal Q-Q Plots of the Volcano Elevations Data ....................................................... 60
3.6 Graph of the Histogram and Fitted Density Curves for the Volcano Data ................. 61
3.7 Histogram of the fiber glass data .................................................................................. 62
3.8 Normal Q-Q Plots of the Fiber Glass Data ................................................................... 62
3.9 Fitted density curves to the fiber glass data ................................................................. 63
3.10 Histogram of the Concentration of As in mg/kg of soil ............................................. 64
3.11 Normal Q-Q Plots of the Concentration of As in mg/kg of soil ................................. 65
3.12 Fitted density curves to the Arsenic data .................................................................... 66
3.13 illustration of the tails of the fitted density curves to the Arsenic data ....................... 66
3.14 Histogram of the Concentration of Cd in mg/kg of soil ............................................ 68
3.15 Normal Q-Q Plots of the Concentration of Cd in mg/kg of soil ................................. 68
3.16 Fitted density curves to the Cadmium data ............................................................... 69
3.17 Illustration of the tail of the fitted density curves to the Cadmium data ...................... 69
3.18 Histogram of the concentration of Lead in soil samples measured in 10 mg/kg of soil ......................................................................................................................... 70
3.19 Normal Q-Q Plots of the Concentration of Pb in mg/kg of soil ................................. 71
3.20 Fitted density curves to the concentration of Lead in soil samples measured in 10 mg/kg of soil ......................................................................................................................... 72
3.21 Illustration of the tails of the fitted density curves to the concentration of Lead in soil samples measured in 10 mg/kg of soil ................................................................. 72

4.1 Graphs of the densities included in the power comparisons of the MSICE of change point as the shape parameter changes: The p.d.f. of $St'_3(0,1,\alpha)$ for $\alpha = 1, -1, -5$ are displayed ........................................................................... 84
4.2 Graphs of the densities included in the power comparisons of the MSICE of change point as the location parameter changes: The p.d.f. of $St'_3(\mu,1,1)$ for $\mu = 0, 2, -2$ are displayed ........................................................................... 87
4.3 Histogram for Chile’s Weekly Return Rates ........................................... 91
4.4 Normal Q-Q Plots for Chile’s Weekly Return Rates ............................... 92
4.5 Probability Histogram of the Standardized Rates of Chile’s Weekly Returns
4.6 SIC(k) Values for Chile’s Standardized Return Rates at each Location with a
Vertical Line at the First Detected Change Point \( \hat{k} = 149 \) ......................... 93
4.7 SIC(k) Values for Chile’s Standardized Return Rates for the First Segment of
the Data ............................................................................................ 93
4.8 SIC(k) Values for Chile’s Standardized Return Rates for the Second Segment
of the Data ........................................................................................ 94
with three vertical lines that mark the detected change points ..................... 95
4.10 Histogram of the standardized weekly returns of the Brazilian market from
4.11 Normal Q-Q Plots for Brazil’s’s Weekly Return Rates ............................. 97
4.12 SIC Values for the the Brazilian Returns at Each Week with Vertical Lines to
Mark the Detected Change Point ....................................................... 98
with a Vertical Line that Marks the Detected Change Point ....................... 98
........................................................................................................ 99
4.15 Normal QQ-Plot of Argentina’s Weekly Return Rates ............................. 100
4.16 SIC Values for the the Argentina’s Returns at Each Week with Vertical Lines
to Mark the Detected Change Point ..................................................... 101
with a Vertical Line that Marks the Detected Change Point ....................... 101


LIST OF TABLES

3.1 Simulated Inverse Gaussian Data ........................................... 56
3.2 Model Comparison for the Simulated Log Normal Data .......................... 58
3.3 Summary of The Volcano Heights Data .................................. 60
3.4 Models Fitted to the Volcano Data ........................................ 61
3.5 Summary of The Fiber Glass Data ..................................... 62
3.6 Models Fitted to the Fiberglass Data .................................. 63
3.7 Summary of The Arsenic Concentration Data .................................. 65
3.8 Models Fitted to the As Data ........................................ 66
3.9 Summary of The Cadmium Concentration Data .................................. 67
3.10 Summary of the Models Fitted to the Cd Data .......................... 68
3.11 Summary of The Plumbum Concentration Data .................................. 71
3.12 Summary of the Models Fitted to the Pb Data .................................. 71

4.1 Power of the MSICE change point detection as the shape parameter $\alpha$ changed from 1 to -1, $\mu = 0, \sigma = 1, r = 3$ and sample size $n = 40$ ........................................ 83
4.2 Power of the MSICE change point detection as the shape parameter $\alpha$ changed from 1 to -1, $\mu = 0, \sigma = 1, r = 3$ and sample size $n = 60$ ........................................ 83
4.3 Power of the MSICE change point detection as the shape parameter $\alpha$ changed from 5 to -1, $\mu = 0, \sigma = 1, r = 3$ and sample size $n = 40$ ........................................ 84
4.4 Power of the MSICE change point detection as the location parameter $\mu$ changed from 0 to 2, $\sigma = 1, \alpha = 1, r = 3$ and sample size $n = 40$ ........................................ 86
4.5 Power of the MSICE change point detection as the location parameter $\mu$
changed from 0 to -2, $\sigma = 1, \alpha = 1, r = 3$ and sample size $n = 40$ . . . . . . 87

4.6 Power of the MSICE change point detection as the location parameter $\mu$
changed from 0 to -2, $\sigma = 1, \alpha = 1, r = 3$ and sample size $n = 60$ . . . . . . 87

4.7 Power of the MSICE change point detection as the location parameter $\mu$
changed from 0 to -2, $\sigma$ changes from 1 to 2, and $\alpha = 1, r = 3$ and sample
size $n = 40$ . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 89

4.8 Power of the MSICE change point detection as the location parameter $\mu$
changed from 0 to -2, $\sigma$ changes from 1 to 2, and $\alpha = 1, r = 3$ and sample size
$n = 60$ . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 89

4.9 Summary of Chile’s Weekly Returns Data . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 91

4.10 Summary of Brazil’s Weekly Returns Data . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 96

4.11 Summary of Argentina’s Weekly Returns Data . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 99
Over the past three decades, there has been an increasing interest in studying distribution families that can model skewed data with long tails. The skew normal distribution has received special interest from many statisticians since 1985. However, skew normal is not an appropriate model for long-tailed data. This serious drawback of the skew normal distribution motivated the search for more flexible alternatives. In this dissertation, we present a non-central skewed version of the Student’s $t$ distribution and we study its applications in model comparison and data analysis. A finite series representation of the probability density function of our model is presented. The theoretical properties of our model show that it is a natural generalization of Student’s $t$ distribution. Unlike the skew normal distribution, our model uses the degrees of freedom as a tuning parameter to adjust the heaviness of the tail of the data. This makes our distribution a great candidate for fitting data with tails longer or heavier than those of the normal distribution.

We show that the normal, Cauchy and the skew normal distributions are special cases of our distribution. Simulation studies are conducted to prove the effectiveness of our model in fitting long-tailed distributions with asymmetry. Data analyses are used to illustrate some of the applications of our model in fitting data arising from environmental, financial and geographical fields of study. Model comparison analyses have been used to illustrate the superior performance of the our model in fitting skewed data with long tails. In conclusion, our model outperforms the normal and the skew normal models and provides a more general alternative that can be used to analyze a wide range of data sets regardless of their asymmetry or long tail behavior.

To further study the non-central skew $t$ distribution, we provide an extension to the multivariate case and study the distribution of quadratic forms of the non-central skew $t$ variable. We also discuss several change point problems related to the non-central skew $t$
distribution. An information based approach to detect the location of the change point of
the non-central skew $t$ model is applied. We study the performance of this approach through
a series of simulation studies. Finally, this approach is used to detect the location of the
change point in data sets that describe the weekly return rates of three Latin American
countries.
CHAPTER 1

LITERATURE REVIEW

1.1 INTRODUCTION

The normal distribution is a classical model that has been used to analyze data in many applications. However, it is not the best option if the data is asymmetric. The skew normal distribution was first introduced by Roberts (1966) as an example of a weighted model, but he did not use the term “skew normal” then. Azzalini (1985) formalized the skew normal distribution as a generalization of the normal distribution that can be used to model asymmetric data. His work inspired many statisticians to study different versions of the skew normal distribution and consider skewed versions of other distributions. Unfortunately, the skew normal model is not optimal for modeling long-tailed data which occur in many applications. Return rates of some stocks are a typical example of skewed data with long-tails for example. Other examples will be presented in chapters 3 and 4 of this dissertation.

1.2 MOTIVATION

The classical Theorems in linear regression have been developed under the assumption that the residuals are at least approximately normally distributed. However, more recent appli-
cations have resulted in data sets which are far from being normally distributed. In fact, in the context of financial data analysis and risk management, using a normal distribution to model asymmetric data with long tails can be a risky and questionable practice because it fails to predict extreme events that tend to coincide with disasters that one has to be prepared for.

Insurance losses and financial returns tend to be asymmetric. Skewed distributions are also important in many disciplines such as biology, meteorology, astronomy and wherever an asymmetric models with longer tails than normal is required. Many real data sets show that symmetric distributions such as the normal distribution and the $t$ distribution do not provide sufficiently good models. There is a strong need for a wide range of asymmetric alternatives that are flexible in their tail behavior.

In addition to the requirement of symmetry, the normal distribution assumes that the tails of the data decay exponentially. This assumption is often violated by many data sets. Using the normal model to approximate data with tails longer than normal is also a dangerous practice in many situations as it rules out a wide range of possibilities.

Student’s $t$ distribution is an alternative to the normal model. It is symmetric and bell shaped just like its normal distribution. However, it has an additional parameter, namely, the degrees of freedom. The degrees of freedom can be viewed as a tuning parameter that can be used to control the heaviness of the tail of the $t$ distribution. By varying the degrees of freedom of the Student’s $t$ distribution, one can obtain a variety of density curves with tails as long as those of the Cauchy distributions or as short as those of the normal model. A graphical illustration of the effect of the degrees of freedom on the thickness of the tail of the Student’s $t$ density curve is provided in figure (1.1). The figure also illustrates that the standard normal distribution is the limiting distribution of the $t$ distribution as the degrees
Figure 1.1: Density curve of the Student $t$ distribution for different degrees of freedom versus the standard normal density.

of freedom tend to infinity.

The requirement of symmetry in the normal model has been relaxed by Azzalini (1985) when he proposed a skewed version of the normal distribution. His Definition was extended and generalized by many other statisticians over the next few decades. Azzalini managed to generalize the univariate skew normal distribution to the multivariate setting in a paper co-authored with Dalla Valle in 1996.

Genton (2004) uses the multivariate skew normal distribution to model weekly returns of UK stocks during the period of 1978-1995 and compares the performance of the skew-normal distribution to the normal distribution. He shows that indeed the returns are skewed, i.e., the null hypothesis of symmetry is rejected. He also proves that the skew-normal model outperforms the standard multivariate normal model. However, for many applications the skew normal model fails to provide a good fit for skewed data due to the excessive kurtosis of the data.
The term “long-tailed distribution” has been used in the finance and insurance business for many years. A distribution is said to be long-tailed if a larger share of population rests within its tail than would under a normal distribution. In other words, a long-tailed distribution includes many values unusually far from the mean than the normal distribution. Furthermore, a long-tailed distribution tends to have higher kurtosis than the normal distribution.

Common examples of long-tailed distributions include the Cauchy distribution, the log-normal distribution and Student’s t distribution. Long-tailed data sets are very common in financial data analysis. For example, real asset returns do not follow the normal distribution. In fact, daily returns show longer tails than normal (i.e., their kurtosis exceeds 3). In other words, extreme events take place far too often than what a normal model would predict. Here are some specific examples of data that have longer tails than normal:

- The daily returns of the Dow Jones industrial average have a standard deviation, $\sigma = 0.012$ or 1.2%. In October 1987, there was a 21% decline in the Dow Jones industrial average. This decline was 20 standard deviations away from the mean.

- On 24 February 2003, the price of natural gas changed by 42% in one day, the new price was 12 standard deviation above the mean.

According to the empirical rule of the normal distribution, 99.7% of all the observations occur within three standard deviations from the mean. In other words, according to the normal model the probability of a random variable taking a value that exceeds 12 standard deviations from the mean is numerically zero. Despite its enhanced ability in fitting asymmetric data, the skew normal distribution remains inadequate for fitting long-tailed data. Several examples of data where the skew normal model is inappropriate will be discussed in chapters 3 and 4 of this dissertation.
Incidences of a violation of the normality assumption have been shown to be quite common. Investment returns have been known to violate assumptions of normality (see for example: Fama (1965), Myers et al. (1984), Stein et al. (1991)). Non-normal data has been shown to exist in biological laboratory data, psychological data, and RNA concentrations in medical data. The need for a skew distribution that can be used to fit long-tailed data motivated researchers to investigate skew versions of the \( t \) distribution.

1.3 THE SKEW NORMAL DISTRIBUTION

In this section we introduce the univariate skew normal distribution, denoted by \( SN \). The skew normal distribution extends the normal distribution to a possibly skewed distribution by incorporating a shape parameter. The skew normal distribution was formalized by Azzalini (1985). Azzalini’s Definition and its location-scale extension will be presented in the following. The notation \( \phi(\cdot) \) and \( \Phi(\cdot) \) will be used to denote the p.d.f. and c.d.f. of the standard normal distribution respectively.

**Definition 1.3.1. (Univariate Skew Normal Distribution):** A random variable \( Z \) is said to have the standard skew normal distribution with shape parameter \( \lambda \in \mathbb{R} \), denoted by: \( Z \sim SN(\lambda) \) if its p.d.f. is given by:

\[
 f(z; \lambda) = 2\phi(z)\Phi(\lambda z), \quad -\infty < z < \infty. \tag{1.1}
\]

A location-scale version of the skew normal distribution is defined as: A random variable \( X \) is said to have the skew normal distribution with parameters \( \mu, \sigma \) and \( \lambda \), denoted by: 

\[
 X \sim SN(\mu, \sigma, \lambda) 
\]
\( X \sim SN(\mu, \sigma, \lambda) \), if its p.d.f. is given by:

\[
f(x; \mu, \sigma, \lambda) = \frac{2}{\sigma} \phi \left( \frac{x - \mu}{\sigma} \right) \Phi \left( \lambda \frac{x - \mu}{\sigma} \right), \quad -\infty < x < \infty.
\]

(1.2)

where \( \mu, \lambda \in \mathbb{R} \) and \( \sigma > 0 \), are the location, shape and scale parameters respectively.

Figure (1.2) shows the graph of the density curves for \( SN(0, 1, \lambda) \) for selected values of the shape parameter (\( \lambda \)). Note that varying the value of the shape parameter, \( \lambda \), for a fixed location and scale parameters causes a change in location and the shape of the density curve. The curve labeled \( SN(0) \) is simply the standard normal density curve. Also note that that positive shape parameters correspond to density curves that are skewed to the right and negative ones correspond to densities that are skewed to the left.

Properties of the skew normal distribution have been studied throughly in the literature. The moment generating function and the first few moments of the skew normal distribution are given in Azzalini (1985). Henze (1986) presented a general expression of the odd moments of the skew normal distribution in a closed form. The even moments coincide with the normal ones because \( Z^2 \sim \chi^2_1 \). Theorem (1.3.2) lists some of the theoretical properties of the skew normal distribution.

**Theorem 1.3.2. (Properties of the Skew Normal Distribution):** Let \( SN(\lambda) \), where \( \lambda \in \mathbb{R} \), denote the standard skew normal distribution with skewness parameter \( \lambda \). The following are basic properties of \( SN(\lambda) \):

(i) If \( Z \sim SN(0) \), then \( Z \sim N(0, 1) \), i.e., the standard normal is a special case of the skew normal distribution.

(ii) If \( Z \sim SN(\lambda) \), then \( -Z \sim SN(-\lambda) \).
(iii) As $\lambda \to \infty, SN(\lambda)$ converges to the half-normal distribution.

(iv) If $Z \sim SN(\lambda)$, then $Z^2 \sim \chi^2_1$.

(v) For a fixed $\lambda$, the density $SN(\lambda)$ is strongly unimodal, i.e., $\log(f(z;a))$ is a concave function of $z$.

Proof. The readers are referred to Azzalini (1985) or Genton (2004) for the proofs and further discussion of the properties.

![Skew Normal Density](image)

Figure 1.2: Density curve of $SN(0,1,\lambda)$ for different values of $\lambda$

The following Definition was proposed by Azzalini (1985) as a two variable extension of his original Definition. We will use the term “Extended Skew Normal” to refer to this distribution from now on. Extended skew normal incorporates two shape parameters, $\lambda$ and $\gamma$, for enhanced control over the shape of the density curve.

**Definition 1.3.3.** *(Extended Skew Normal):* A real valued random variable $Z$ is said to have the extended skew normal distribution with parameters $\lambda, \gamma \in \mathbb{R}$, denoted by $Z_{\lambda,\gamma} \sim$
ESN(\lambda, \gamma), if its p.d.f. is given by:

\[ \phi_{SN}(z; \lambda, \gamma) = \frac{\phi(z) \Phi(\lambda z + \gamma)}{\Phi(\sqrt{\gamma^2 + \lambda^2})}. \] (1.3)

Figure 1.3: Density of the ESN distribution with \( \gamma = 0 \) coincides with the SN density

A graphical illustration of the relation between the original Definition of the skew normal random variable and the extended skew normal variable is presented below. The ESN density reduces to the SN density if \( \gamma = 0 \). Figure (1.3) shows that the graph of the ESN(\mu, \sigma, \lambda, \gamma) density with a given choices of the parameter (\mu, \sigma, \lambda) and \( \gamma = 0 \) coincides with the graph of SN density with the same choice of (\mu, \sigma, \lambda).
Figure 1.4: Density of ESN versus the SN density when $\gamma \neq 0$

Figure 1.4 provides a comparison between the ESN density with $\gamma \neq 0$ and the SN density with the same choice of $\mu, \sigma, \lambda$ values. The graphs indicates that the ESN density with $\gamma \neq 0$ differs in both the shape and location from the SN density that has the same parameter values for $\mu, \sigma, \alpha$. 
1.4 THE MULTIVARIATE SKEW NORMAL DISTRIBUTION

Azzalini (1985) introduced a multivariate extension of the univariate skew normal distribution. Unfortunately, his definition was not satisfactory because the marginals are not skew normal. Azzalini and Dalla-Valle (1996) constructed a multivariate distribution with skew normal marginals and refer to it as “multivariate skew normal distribution”. The readers are referred to Azzalini and Dalla Valle (1996) for further details. For the \( k \)-dimensional extension of Definition (1.3.1), we consider a multivariate random variable \( Z \) such that each component is skew normal. Then, it is natural to define the joint distribution of \( Z \) to be a multivariate skew normal variable.

**Definition 1.4.4. (Multivariate Skew Normal Distribution):** A random vector \( Z = (Z_1, Z_2, \ldots, Z_k)^T \) is said to have the \( k \)-dimensional skew normal distribution if its density is given by:

\[
f_k(z; \lambda) = 2 \phi_k(z; \Omega) \Phi(\lambda^T z); \quad z \in \mathbb{R}^k.
\] (1.4)

where \( \phi_k(z; \Omega) \) denotes the density of the \( k \)-dimensional multivariate normal distribution with standardized marginals and correlation matrix \( \Omega \). We denote such a random vector by \( Z \sim SN_k(\Omega, \lambda) \). The vector \( \lambda \) is also called shape vector, even though the components of \( \lambda \) do not necessarily coincide with the shape parameters of the marginals.

In order to compute the moment generating function of \( Z \), we need the following lemma.

**Lemma 1.4.5.** If \( Z \sim N(0, 1) \), and let \( \Phi \) be the c.d.f. of \( Z \), then

\[
E_Z\{\Phi(a + bZ)\} = \Phi \left[ \frac{a}{\sqrt{1 + b^2}} \right]
\] (1.5)
for any scalars $a, b \in \mathbb{R}$.

The multivariate version of lemma (1.4.5) is provided below.

**Lemma 1.4.6.** If $U \sim N_k(0, \Omega)$, where $\Omega$ is a $k \times k$ positive definite variance-covariance matrix, then

$$E_U \{ \Phi(a + b^T U) \} = \Phi \left( \frac{a}{(1 + b^T \Omega b)^{1/2}} \right)$$

(1.6)

for all $a \in \mathbb{R}$ and $b \in \mathbb{R}^k$.

**Proof.** See Zacks (1981).

\[ \square \]

1.5 **STUDENT’S $t$ DISTRIBUTION**

1.5.1 **HISTORICAL NOTES**

Student’s $t$ distribution was first derived as a posterior distribution in 1876 by Helmert and Lroth. In the English literature, a derivation of the $t$ distribution was published in 1908 by William Sealy Gosset while he worked at the Guinness Brewery in Dublin, Ireland. One version of the origin of the pseudonym “Student” is that Gosset’s employer forbade members of its staff from publishing scientific papers, so he had to hide his identity. Another version is that Guinness did not want their competitors to know that they were using the $t$-test to test the quality of raw material.

The $t$-test and the associated theories became well-known through the work of Ronald A. Fisher, who called the distribution “Student’s distribution”. This distribution is familiar to most statisticians due to its applications to small sample statistics in elementary statistics. It is parametrized by one parameter, $r$, that is commonly referred to as “degrees of freedom” associated with the distribution. As the degrees of freedom tends to infinity, Student’s $t$ distribution converges to the normal distribution. For finite values of the degrees
of freedom, \( r \), the tails of the density function decay as an inverse power of order \( r + 1 \). Therefore, the \( t \) distribution is said to have longer tails relative to the normal case. A recent paper by Ferguson and Platen (2006) suggests, for example, that the distribution \( t_4 \) is an accurate representation of index returns in a global setting, and propose models to underpin this idea. The fact that returns tend to have positive excess kurtosis has been known for over four decades.

### 1.5.2 CONSTRUCTION

Student’s \( t \) distribution can be derived from the normal distribution. This derivation is useful in studying the \( t \) distribution and in generating random samples for simulation studies. Let \( Z_0, Z_1, \ldots, Z_r \) be a series of independent standard normal random variables. Define

\[
X = Z^2_1 + Z^2_2 + \ldots + Z^2_r, \tag{1.7}
\]

The density function of \( X \) can be easily derived using the moment generating function. The random variable \( X \) has been well studied in the literature and it has the chi-squared distribution with \( r \) degrees of freedom. The notation \( \chi^2_r \) is commonly used to denote this distribution. Student’s \( t \) distribution with the parameter \( r \) is defined to be the random variable \( T \) defined as follows:

\[
T = \frac{Z_0 \sqrt{X/r}}. \tag{1.8}
\]

The notation \( t_r \) will be used to denote this random variable throughout this dissertation.

The \( p.d.f. \) of Student’s \( t_r \) distribution is given by the following form:

\[
f_r(t) = \frac{\Gamma \left( \frac{r+1}{2} \right)}{\sqrt{r\pi} \Gamma \left( \frac{r}{2} \right)} \cdot \frac{1}{\left(1 + t^2/r\right)^{(r+1)/2}} : -\infty < t < \infty. \tag{1.9}
\]
where $\Gamma$ denotes the Gamma function, which is defined for all positive real numbers by the integral equation:

$$\Gamma(a) = \int_0^\infty x^{a-1}e^{-x}\,dx, \quad \text{where } a > 0. \quad (1.10)$$

The number $r$, which sometimes is regarded as an integer, is called the “degrees of freedom” of the distribution. It is evident that a sample from this distribution can easily be obtained by using $r+1$ samples from the standard normal distribution, provided that $r$ is an integer. In general, the degrees of freedom of the $t$ distribution are not restricted to integers. In fact, equation (1.9) can be used to define the p.d.f. of the $t$ distribution for any positive real value of the parameter $r$. General non-integer low values of $r$ are of interest in financial analysis for short time scales. Gencay, et al. (2001) suggests that very short term returns exhibit power law decay in the density curve. For a $t_r$ distribution the decay of the p.d.f. is of the order of $O(t^{-r-1})$ and the decay of the c.d.f. is of the order of $O(t^{-r})$.

In order for the $k^{th}$ moment of Student’s $t$ distribution to be defined, its degrees of freedom must be greater than $k$. All the odd moments are equal to zero, provided that $r$ is large enough so that they are defined, due to the symmetry of the density. The mean, variance, skewness, and kurtosis of the $t$ distribution are listed in order as follows:

$$\mu = 0 : \ r > 1, \quad (1.11)$$

$$\sigma^2 = \frac{r}{r-2} : \ r > 2, \quad (1.12)$$

$$\gamma_1 = 0 : \ r > 3 \quad (1.13)$$

$$\gamma_2 = \frac{6}{r-4} : \ r > 4. \quad (1.14)$$
In general, the absolute moments \( E(|t|^k) \) can be calculated as follows:

\[
E(|t|^k) = \frac{2\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{r\pi} \Gamma\left(\frac{r}{2}\right)} \int_0^\infty \frac{t^k}{(1 + t^2/r)^{(r+1)/2}} \, dt
\]

(1.15)

\[
= \frac{r^{(k/2)} \Gamma\left(\frac{k+1}{2}\right) \Gamma\left(\frac{r-k}{2}\right)}{\sqrt{\pi} \Gamma\left(\frac{r}{2}\right)}, \text{ where } r > k.
\]

(1.16)

Student’s \( t_r \) random variable does not possess a moment generating function (If it does, then all the moments would exist). Furthermore, there is no simple form for the characteristic function of the \( t \) distribution. Sutradhar (1986) discussed the characteristic function of multivariate Student’s \( t \) distribution. One of the nice properties of the \( t \) distribution is that it is an infinitely divisible distribution. Epstein (1977) presented a theoretical proof of this interesting property. Other authors have also discussed the infinite divisibility of the \( t \) distribution under certain restrictions.

Student’s \( t \) distribution can be generalized to a three parameter location-scale family by introducing a location parameter \( \mu \) and a scale parameter \( \sigma \), and applying the linear transformation

\[
X = \mu + \sigma T.
\]

(1.17)

If \( T \sim t_r \), then \( X \) is said to follow the non-standard Student’s \( t \) distribution with the parameters \( \mu, \sigma, r \). Some authors refer to the \( t \) distribution as the standard \( t \) distribution and to the location-scale \( t \) distribution as the three-parameter \( t \) distribution.
1.5.3 MULTIVARIATE AND NON-CENTRAL T DISTRIBUTIONS

The bivariate $t$ distribution has attracted somewhat limited attention of researchers for the last 70 years in spite of their increasing importance in classical as well as in Bayesian statistics. Nadarajah and Kotz (2003) provide a comprehensive review of the known bivariate $t$ distributions. We present some key ideas from this paper as follows.

Definition 1.5.7. (Bivariate $t$ Distribution): A two dimensional random vector $X = (X_1, X_2)^T$ is said to have the bivariate $t$ distribution with degrees of freedom $r$, mean vector $\mu$, and with $\Sigma$ denoting the covariance matrix if its joint probability density function is given by:

$$ f(x) = \frac{\Gamma \left( \frac{r+1}{2} \right)}{\sqrt{\pi r} \Gamma(r/2)|\Sigma|^{1/2}} \left[ 1 + \frac{1}{r} (x - \mu)^T \Sigma^{-1} (x - \mu) \right]^{-(r+2)/2} \quad (1.18) $$

The degrees of freedom parameter, $r$, is also referred to as the shape parameter, because the peakedness of the density may be diminished, preserved, or increased by varying $r$. The distribution is said to be central if $\mu = 0$; otherwise, it is said to be non-central.

If $r = 1$, then equation (1.18) is the bivariate Cauchy distribution. The limiting form of equation (1.18) as $r \to \infty$ is the joint p.d.f. of the bivariate normal distribution with the mean vector $\mu$ and the covariance matrix $\Sigma$.

It is worth mentioning that there are many other ways to generalize the Student’s $t$ distribution to the multivariate version. One of the earliest forms of bivariate $t$ distributions was proposed by Owen (1965) in a paper that deals mainly with the univariate non-central $t$ distribution. Nadarajah and Kotz (2003) provide a review of the variants of the bivariate $t$ distribution which are discussed in the literature periodically but are not widely known. Kotz and Nadarajah (2004) provided detailed discussion of both the bivariate and the multivariate variations of the $t$ distribution in their book entitled “Multivariate $t$ distributions and their applications”. In Chapter 4 of their book, they discussed 11 types of bivariate
generalizations of the $t$ distribution. In Chapter 5, they discussed 14 types of multivariate generalizations of the $t$ distribution. Nadarajah and Dey (2005) presented further study of the multivariate generalizations of the $t$ distribution.

Most of the definitions of the non-central $t$ distribution rely on constructing the $t$ distribution by using a transformation that starts with a non-standard normal distribution and an independent chi-squared one. Some generalizations in the literature replace the chi-squared variable in the denominator of the transformation in equation (1.8) by a more general gamma variable. The readers are referred to Kotz and Nadarajah’s (2004) book for more detailed discussion of the literature on the non-central and the multivariate $t$ distributions and their applications.

1.6 LITERATURE REVIEW OF SKEW $t$ DISTRIBUTIONS

A large number of papers have been published on the applications of skewed distributions. An extensive list of applications is provided in Genton (2004) and Azzalini (2005). The skew $t$ distribution received special attention after the introduction of the skew multivariate normal distribution in Azzalini and Dalla Valle (1996). The results of the multivariate normal distribution motivated researchers to study possible applications that adapt the same ideas to different distribution families. However, most of them are simply a variation of the form of Azzalini’s $p.d.f.$ of the skew normal distribution.

Gupta (2000) defined a skew multivariate $t$ distribution using a pair of independent standard skew normal and chi-squared random variables. He also studied a skew Cauchy distribution and some other skew distributions. His approach is different from the approaches
proposed by Azzalini and his collaborators, but he was able to obtain a closed form for his
density that is similar to Azzalini’s form. Azzalini and Capitanio (2002) defined a skew t
variate as a scale mixture of skew normal and chi-squared variables. Huang, et al. (2007)
studied generalized skew t distributions and used it in data analysis. Arellano-Valle, et al.
(2005) discussed generalized skew distributions in the multivariate setting.

The following Definition was used in the literature to introduce generalized skew t dis-
tributions:

Definition 1.6.8. (Generalized Skew t Distribution): A random variable $X$ is said to
have a generalized skew t distribution with skewness coefficient $\lambda$ and degrees of freedom $r$ if
its p.d.f. is given by:

$$f_X(x) = 2f_{t_r}(x) \cdot F_{t_r}(\lambda x),$$  \hspace{1cm} (1.19)

where $f$ and $F$ represent the p.d.f. and c.d.f. of the Student’s t distribution respectively, and
$\lambda \in \mathbb{R}$ is the shape parameter.

Several variations and extensions of Definition (1.6.8), have been studied in the litera-
ture. More recent work focused on applications to this approach in data analysis. Some
authors proposed replacing the term $\lambda x$ in the argument of the c.d.f. in equation (1.19)
by a general function of $\lambda$ and $x$. Nadarajah and Kotz (2003) discussed the skewed distri-
butions generated by the normal kernel. Two unpublished manuscripts by Huang, et al.
(2009) discussed variations of Definition (1.6.8). Kim and Mallick (2003) studied the mo-
ments of random vectors with skew t distribution and their quadratic forms. Jamaliadeneh, et
al. (2008) discussed order statistics from bivariate skew normal and skew t distributions.
Lin (2009, 2010) discussed applications of mixtures of skew t distributions. For more de-
tails and additional references, the readers are referred to the bibliography of my dissertation.
CHAPTER 2

A NON-CENTRAL SKEW $t$ DISTRIBUTION

2.1 INTRODUCTION

Student’s $t$ distribution has been used in the literature as a model to fit symmetric data with tails longer than those of the normal distribution. Several authors have studied possible extensions of the $t$ distribution to a non-central setting or to a multivariate one. Chapter 1 highlights some of the progress that has been made in this field over the past few decades. A more important problem is studying a skew version of the $t$ distribution. Section (1.6) has a brief review of the literature on this problem.

In this chapter, we present a new approach to define a non-central skew $t$ distribution. Theoretical properties of this distribution will be discussed and some graphical and numerical properties will be presented. The chapter will also discuss a Definition that generalizes our non-central skew $t$ distribution to the multivariate case. We will use the terms “Extended Skew $t$” or “Non-central Skew $t$” to refer to our distribution and to distinguish it from the generalized skew $t$ distributions discussed in the literature.
2.2 A UNIVARIATE NON-CENTRAL SKEW $t$ DISTRIBUTION

2.2.1 INTRODUCTION

In this section a new approach to define a univariate non-central skew $t$ random variable is presented. Definition 1.6.8 introduced a skew $t$ random variable by specifying the form of its density function. Instead of starting from the density function, we will start from a transformation of variables that resembles the one discussed in section 1.5 of chapter 1.

Definition 2.2.1. Let $X$ be a skew normal variable with parameters $(\mu, \sigma, \alpha)$. Let $Y$ be a $\chi^2$ variable with $r$ degrees of freedom. Assume further that $X, Y$ are independent, then the non-central skew $t$ random variable, $T$, is defined by the transformation:

$$T = \frac{X}{\sqrt{Y/r}}.$$

(2.1)

We will use $St'(\mu, \sigma, \alpha)$ to denote the non-central skew $t$ distribution throughout this dissertation. Definition 2.2.1 follows the original construction of the Student’s $t$ distribution that was reviewed in Chapter 1. It replaces the standard normal variable in the numerator by a location-scale skew normal variable. Hence, it has the added feature of skewness and its shape can be controlled by two additional parameters. Unlike the skew normal distribution, the non-central skew $t$ distribution has the additional feature of a tuning parameter, $r$, which can be used to control the rate at which the tail of the distribution decays.
2.2.2 PROBABILITY DENSITY FUNCTION OF THE $St'_r$ DISTRIBUTION

Let $\phi(.)$, $\Phi(.)$ denote the standard normal density and the cumulative distribution functions respectively. The inverse transformation method will be used to obtain the expression for the p.d.f. of the random variable $T$ defined above. The joint density of $X,Y$ is given by:

$$h(x, y) = \frac{2y^{\frac{r}{2}-1}e^{-\frac{x^2}{2}}}{\sigma \Gamma\left(\frac{r}{2}\right)} \phi\left(\frac{x-\mu}{\sigma}\right) \Phi\left(\frac{\alpha(x-\mu)}{\sigma}\right), \quad -\infty < x < \infty, 0 < y < \infty. \tag{2.2}$$

Or equivalently

$$h(x, y) = \frac{2y^{\frac{r}{2}-1}e^{-\frac{y^2}{2}}}{\sqrt{2\pi\sigma} \Gamma\left(\frac{r}{2}\right)} e^{-\frac{1}{2}(\frac{x-\mu}{\sqrt{\sigma}})^2} \phi\left(\frac{x-\mu}{\sigma}\right) \Phi\left(\frac{\alpha(x-\mu)}{\sigma}\right), \quad -\infty < x < \infty, 0 < y < \infty. \tag{2.3}$$

The Jacobian of the transformation is given by:

$$|J| = \sqrt{\frac{y}{r}}. \tag{2.4}$$

The joint density of $T,Y$ is given by:

$$f(t, y) = \frac{2y^{\frac{r}{2}-1}e^{-\frac{y^2}{2}}}{\sqrt{2\pi\sigma} \Gamma\left(\frac{r}{2}\right)} e^{-\frac{1}{2}(\frac{t\sqrt{\frac{r}{\sigma}}-\mu}{\sqrt{\sigma}})^2} \phi\left(\frac{t\sqrt{\frac{r}{\sigma}}-\mu}{\sigma}\right) \Phi\left(\frac{\alpha(t\sqrt{\frac{r}{\sigma}}-\mu)}{\sigma}\right) \sqrt{\frac{y}{r}}, \tag{2.5}$$

To get the marginal p.d.f. of $T$, we integrate the expression in (2.6) with respect to $y$. Thus,

$$f_T(t) = \frac{2e^{-\frac{t^2}{2\sigma^2}}}{\sqrt{2\pi\sigma} \cdot \Gamma\left(\frac{r}{2}\right)^{2r/2}} \int_0^\infty y^{\left(\frac{r}{2}-1\right)} e^{-\frac{y^2}{2}(t^2r^{-1}\sigma^{-2}+1)+\frac{\mu\nu\pi}{\sigma^2\sqrt{\sigma}}} \phi\left(\frac{t\sqrt{\frac{r}{\sigma}}-\mu}{\sigma}\right) dy. \tag{2.6}$$
Using the following substitutions:

\[ \beta = \frac{2r\sigma^2}{r\sigma^2 + t^2}, \quad (2.7) \]
\[ \gamma = t\mu - r^{-\frac{1}{2}} = \frac{t\mu}{\sigma^2\sqrt{r}}, \quad (2.8) \]
\[ u = \sqrt{y}. \quad (2.9) \]

Equation (2.6) is reduced to:

\[ f_T(t) = \frac{4e^{-\frac{t^2}{2\sigma^2 + t^2}}}{\sqrt{2\pi r\sigma\Gamma(\frac{1}{2})2r^{1/2}}} \int_0^\infty u e^{-\frac{1}{\beta}(u-\frac{\beta\gamma}{2})^2} \Phi\left(\frac{tu/\sqrt{r} - \mu}{\sigma}\right) du. \quad (2.10) \]

Let \( I \) be the integral part of equation (2.10). Using the substitution \( v = u - \frac{\beta\gamma}{2} \), we obtain:

\[ I = \int_{-\frac{\beta\gamma}{2}}^\infty e^{-v^2/\beta} \left( v + \frac{\beta\gamma}{2} \right)^r \Phi\left(\frac{\alpha(t\sqrt{r} - \mu)}{\sigma}\right) dv. \quad (2.11) \]

Let

\[ a = \frac{\alpha t}{\sigma\sqrt{r}}, \quad b = \frac{\alpha}{\sigma} \left( \frac{t\gamma\beta}{2\sqrt{r}} - \mu \right). \quad (2.12) \]

Using the binomial expansion and the above substitutions \( I \) can be simplified as follows:

\[ I = \int_{-\frac{\beta\gamma}{2}}^\infty e^{-v^2/\beta} \sum_{k=0}^r \binom{r}{k} \left( \frac{\beta\gamma}{2} \right)^{r-k} v^k \Phi(a v + b) \ dv \]

\[ = \sum_{k=0}^r \left\{ \binom{r}{k} \left( \frac{\beta\gamma}{2} \right)^{r-k} \int_{-\frac{\beta\gamma}{2}}^\infty e^{-v^2/\beta} v^k \Phi(a v + b) \ dv \right\} \]

\[ = \sqrt{\pi\beta} \sum_{k=0}^r \left\{ \binom{r}{k} \left( \frac{\beta\gamma}{2} \right)^{r-k} \int_{-\frac{\beta\gamma}{2}}^\infty \left[ \frac{1}{\sqrt{\pi\beta}} e^{-v^2/\beta} v^k \Phi(a v + b) \right] \ dv \right\} \]

\[ = \sqrt{\pi\beta} \sum_{k=0}^r \left\{ \binom{r}{k} \left( \frac{\beta\gamma}{2} \right)^{r-k} \int_{-\infty}^{\infty} \left[ 1_{[-\infty,\infty)}(v) v^k \Phi(a v + b) \right] \ dv \right\} \]

\[ = \sqrt{\pi\beta} \sum_{k=0}^r \left\{ \binom{r}{k} \left( \frac{\beta\gamma}{2} \right)^{r-k} \mathbb{E}_V \left[ 1_{[-\infty,\infty)}(V) V^k \Phi(a V + b) \right] \right\} \]

where \( V \sim N(0, \beta/2) \).
Thus, the p.d.f. of T can be given by:

\[
f_T(t) = \frac{4\sqrt{\beta e^{-\frac{\mu^2}{2\sigma^2}} + \frac{\beta^2}{4\sigma^2}}}{\sqrt{2\pi} r \sigma \Gamma(\frac{r}{2}) 2^{\frac{r}{2}}} \sum_{k=0}^{r} \binom{r}{k} \left(\frac{\beta \gamma}{2}\right)^{r-k} \mathbb{E}_V \left[1_{[-\frac{\mu^2}{2\sigma^2}, \infty)}(V) \ V^k \Phi(aV + b)\right],
\]

where \( V \sim N(0, \beta/2) \).

Let \( W = \sqrt{\frac{\beta}{2}} V \), then

\[
\mathbb{E}_V \left[1_{[-\frac{\mu^2}{2\sigma^2}, \infty)}(V) \ V^k \Phi(aV + b)\right] = \int_{-\infty}^{\infty} \left[1_{[-\frac{\mu^2}{2\sigma^2}, \infty)}(v) \ v^k \frac{1}{\sqrt{\pi} \beta} e^{-v^2/\beta} \Phi(av + b)\right] dv
\]
\[
= \left(\frac{\beta}{2}\right)^{(k+1)/2} \int_{-\infty}^{\infty} \left[1_{[-\gamma\sqrt{\frac{\beta}{2}}, \infty)}(w) \ w^k e^{-w^2/2} \Phi(\hat{a}w + b)\right] dw
\]
\[
= \sqrt{2\pi} \left(\frac{\beta}{2}\right)^{(k+1)/2} \mathbb{E}_W \left[1_{[-\gamma\sqrt{\frac{\beta}{2}}, \infty)}(W) \ W^k \Phi(\hat{a}W + b)\right],
\]

where \( W \sim N(0, 1) \) and \( \hat{a} = a\sqrt{\frac{\beta}{2}} \).

(2.15)

Now the p.d.f. of T is given by:

\[
f_T(t) = \frac{\sqrt{\pi} \beta^{(r+1)} e^{\left(\frac{\beta^2}{4} - \frac{\mu^2}{2\sigma^2}\right)}}{\sqrt{r\sigma^2 \Gamma(r/2) 2^{1.5(r-1)}}} \sum_{k=0}^{r} \binom{r}{k} \left(\frac{2}{\beta}\right)^{(k/2)} \gamma^{r-k} \mathbb{E}_W \left[1_{[-\gamma\sqrt{\frac{\beta}{2}}, \infty)}(W) \ W^k \Phi(\hat{a}W + b)\right],
\]

where \( W \sim N(0, 1) \).

(2.16)

Equation (2.16), provides a closed form for the marginal density of \( St_r^r \). However, in order for this expression to be useful in practice, one needs to calculate the expected value involved. To study the expectation in equation (2.16), some results from the literature need to be introduced.

**Definition 2.2.2.** (Azzalini, 1985) A random variable \( Z_{\lambda, \gamma} \) is said to have an extended standard skew normal distribution with parameters \( \lambda, \gamma \in \mathbb{R} \), denoted by \( Z_{\lambda, \gamma} \sim ESN(\lambda, \gamma) \),
if its p.d.f. is given by:

$$\phi_{SN}(z; \lambda, \gamma) = \frac{\phi(z) \Phi(\lambda z + \gamma)}{\Phi\left(\frac{\gamma}{\sqrt{1+\lambda^2}}\right)}.$$  \hspace{1cm} (2.17)

This Definition was proposed by Azzalini (1985) as a two variable extension of his original Definition. We will denote this distribution by $ESN$ in the following discussion. Based on this Definition Jamalizadeh, et al. (2009) provided the following Definition of the truncated skew normal distribution.

**Definition 2.2.3. (Jamalizadeh, et al. 2009):** A random variable $Z$ is said to have a truncated skew-normal random variable $Z_{T(a,b); \lambda, \gamma} \sim TSN((a,b), \lambda, \gamma)$, with parameters $\lambda, \gamma \in \mathbb{R}$ and $a, b : (a < b)$ and possibly one of them is infinite but not both), if and only if

$$Z_{T(a,b); \lambda, \gamma} \overset{D}{=} Z_{\lambda, \gamma} | (a < Z_{\lambda, \gamma} < b),$$  \hspace{1cm} (2.18)

Where $\overset{D}{=} \text{means equal in distribution}.$

Let the c.d.f. of $Z_{\lambda, \gamma}$ be denoted by $\Phi_{SN}(z, \lambda, \gamma)$, then the c.d.f. and p.d.f. of $Z_{T(a,b); \lambda, \gamma}$ will be given by:

$$\Phi_{TSN}(z; (a, b), \lambda, \gamma) = \begin{cases} 
0 & \text{if } z < a \\
\frac{\Phi_{SN}(z; \lambda, \gamma) - \Phi_{SN}(a; \lambda, \gamma)}{\Phi_{SN}(b; \lambda, \gamma) - \Phi_{SN}(a; \lambda, \gamma)} & \text{if } a \leq z < b \\
1 & \text{if } z > b
\end{cases},$$  \hspace{1cm} (2.19)

and

$$\phi_{TSN}(z; (a, b), \lambda, \gamma) = c(\lambda, \gamma, a, b) \phi(z) \Phi(\lambda z + \gamma), \hspace{1cm} a < z < b,$$  \hspace{1cm} (2.20)

respectively, where

$$c(\lambda, \gamma, a, b) = \frac{1}{\Phi\left(\frac{\gamma}{\sqrt{1+\lambda^2}}\right)\{\Phi_{SN}(b; \lambda, \gamma) - \Phi_{SN}(a; \lambda, \gamma)}\}. \hspace{1cm} (2.21)$$

The following Theorem provides the moment generating function of $Z_{T(a,b); \lambda, \gamma}.$
Theorem 2.2.4. (Jamalizadeh, et al. 2009): The moment generating function of $Z_{T(a,b),\lambda,\gamma}$ is given by:

$$
M(s; a, b, \lambda, \gamma) = c(\lambda, \gamma, a, b) e^{s^2/2} \{ \Phi_2 \left( \frac{\lambda s + \gamma}{\sqrt{1 + \lambda^2}}, b - s; \frac{-\lambda}{\sqrt{1 + \lambda^2}} \right) \\
- \Phi_2 \left( \frac{\lambda s + \gamma}{\sqrt{1 + \lambda^2}}, a - s; \frac{-\lambda}{\sqrt{1 + \lambda^2}} \right) \}.
$$

(2.22)

where $\Phi_2(., ., \delta)$ denotes the c.d.f. of $N_2(0, 0, 1, 1, \delta)$, the standard bivariate normal with correlation coefficient $\delta$.

Using the m.g.f. one can obtain the following expression for the first two moments of $Z_{T(a,b),\lambda,\gamma}$ as follows:

$$
\mathbb{E}Z_{T(a,b),\lambda,\gamma} = c(\lambda, \gamma, a, b) \{ \phi(a)\Phi(a\lambda + \gamma) - \phi(b)\Phi(b\lambda + \gamma) + \\
\frac{\lambda}{\sqrt{1 + \lambda^2}} \phi \left( \frac{\gamma}{\sqrt{1 + \lambda^2}} \right) \left[ \Phi \left( \frac{b(1 + \lambda^2) + \lambda\gamma}{\sqrt{1 + \lambda^2}} \right) - \Phi \left( \frac{a(1 + \lambda^2) + \lambda\gamma}{\sqrt{1 + \lambda^2}} \right) \right] \}
$$

(2.23)

$$
\mathbb{E}Z^2_{T(a,b),\lambda,\gamma} = 1 + c(\lambda, \gamma, a, b) \{ a\phi(a)\Phi(a\lambda + \gamma) - b\phi(b)\Phi(b\lambda + \gamma) \\
+ \lambda c(\lambda, \gamma, a, b) \times \{ \phi(a)\Phi(a\lambda + \gamma) - \phi(b)\Phi(b\lambda + \gamma) \\
+ \frac{\lambda}{\sqrt{1 + \lambda^2}} \phi \left( \frac{\gamma}{\sqrt{1 + \lambda^2}} \right) \left[ \Phi \left( \frac{b(1 + \lambda^2) + \lambda\gamma}{\sqrt{1 + \lambda^2}} \right) - \Phi \left( \frac{a(1 + \lambda^2) + \lambda\gamma}{\sqrt{1 + \lambda^2}} \right) \right] \}.
$$

(2.24)

Note that the expressions given in equations (2.23) and (2.24) are originally stated in Jamalizadeh, et al. (2009) with a typo that has been corrected here.

Using the moments of the truncated skew normal density, equation (2.16) can be rewritten as a finite sum of terms involving moments of $Z_{T(a,b),\lambda,\gamma}$ as follows:
\[ f_T(t) = \frac{4}{\sqrt{\pi} \sigma \Gamma(\frac{r}{2})} \frac{e^{-\frac{\mu^2 + \beta \gamma}{2\sigma^2}}}{\sqrt{1 + a^2}} \Phi \left( \frac{b}{\sqrt{1 + a^2}} \right) \left[ 1 - \Phi_{SN}( -\gamma \sqrt{\frac{\beta}{2}} ; \hat{a}, \hat{b}) \right] \times \sum_{k=0}^{r} \left\{ \binom{r}{k} \left( \frac{\beta}{2} \right)^{r + \frac{1 - k}{2}} \gamma^{r - k} E Z^k \right\}, \text{where } Z = Z_{T(-\gamma \sqrt{\frac{\beta}{2}}, \infty)}, \hat{a}, \hat{b}. \] (2.25)

Equation (2.25) provides an alternative expression for the p.d.f. of the random variable \( T \sim St_r'(\mu, \sigma, \alpha) \). The expression uses a finite sum that has \( r + 1 \) terms. Note that the summands depend on the moments of Jamalizadeh’s truncated skew normal random variable. The density of \( St_r' \) for any given positive integer \( r \) can be evaluated using this equation since an explicit expression for the moment generating function is available. The complexity of the computations increases as the degrees of freedom increase. It is worth mentioning that this expression is only valid for integer values of \( r \). Despite this limitation, equation (2.25) is a neat result that relates the density of the extended skew t distribution to the moments of the truncated normal distribution.

The computations required to obtain higher moments of the truncated skew normal distribution are time consuming. It is desirable to have a simple algebraic form for the \( k^{th} \) moment but unfortunately there is not. This makes it difficult to use equation (2.25) for higher values of the degrees of freedom. Being able to evaluate the density at any point is necessary to compute the likelihood function of a given data set, and to use the model in data analysis. To overcome this difficulty, numerical integration can be used to evaluate the density of the \( St_r' \) for any positive value of \( r \). The programming language “R” has been used to write a program that can be used to compute the density \( St_r' \) distribution for any value of \( r \). We double checked that the expression given in equation (2.25) agrees with the numerical integral results for \( r = 1 \) and \( r = 2 \). The numerical integral method has been used to produce the graphics and conduct the data analysis throughout this dissertation. Some of the functions that are used to perform the computations are outlined in the Appendix.
2.3 PROPERTIES OF THE \( St'_r \) DISTRIBUTION

In this section, some properties and special cases of the non-central skew \( t \) random variable will be discussed. The first obvious result to discuss is that our non-central skew \( t \) distribution is a generalization of the standard Student’s \( t \) distribution and the non-central \( t \) distribution. However, it is not a generalization of the non-standard \( t \) distribution that is defined in chapter 1. The following lemma summarizes these special cases:

**Lemma 2.3.1.** Let \( St'_r(\mu, \sigma, \alpha) \) denote the non-central skew \( t \) distribution and the vector \((\mu, \sigma, \alpha, r)\) denote its parameter vector.

(i) The random variable \( St'_r(\mu, \sigma, \alpha) \) reduces to the standard Student’s \( t \) variate if the parameter vector is chosen as \((\mu, \sigma, \alpha, r) = (0, 1, 0, r)\).

(ii) The random variable \( St'_r(\mu, \sigma, \alpha) \) reduces to the non-central Student’s \( t \) variate with non-centrality parameters \( \mu, \sigma \) and \( r \) degrees of freedom if the parameter vector is chosen as \((\mu, \sigma, \alpha, r) = (\mu, \sigma, 0, r)\).

**Proof.** Follows immediately from the construction of the non-central skew \( t \) distribution and the fact that the skew normal distribution reduces to the normal distribution if the skewness parameter is set to be zero.

**Lemma 2.3.2.** The random variable \( St'_r \) reduces to the standard Cauchy variate if the parameter vector is chosen as \((\mu, \sigma, \alpha, r) = (0, 1, 0, 1)\).

**Proof.** First, we will need to write an explicit form of the p.d.f. in equation (2.25) for \( r = 1 \). This requires computing the first moment of \( Z_{T(\gamma \sqrt{2}, \infty), \hat{a}, \hat{b}} \) which is given by equation (2.23).

Plugging in the values of \((\mu, \sigma, \alpha) = (0, 1, 0)\) implies that \( \gamma = \hat{a} = b = 0 \) and \( \beta = \frac{2}{1+r^2} \).
Hence, the p.d.f. of $T$ when $(\mu, \sigma, \alpha) = (0, 1, 0)$ and $r = 1$, reduces to:

$$ f_T(t) = \frac{(2\beta)^{3/2}}{\sigma \Gamma(1/2)} \Phi \left( \frac{b}{\sqrt{1 + \tilde{a}^2}} \right) \left[ 1 - \Phi_{SN} \left( -\gamma \sqrt{\frac{\beta}{2}} \tilde{a}, b \right) \right] $$

$$ \times \sum_{k=0}^{1} \left\{ \left( \frac{\beta}{2} \right)^{1-k} \gamma^{1-k} EZ^k \right\}, \text{where } Z \sim Z_{T(0,\infty),0,0}. \quad (2.26) $$

Note that $EZ^0 = 1$. Using equation (2.23), we can easily see that $EZ^1 = \sqrt{\frac{2}{\pi}}$.

Hence, the p.d.f. reduces to:

$$ f_T(t) = \frac{1}{\pi} \frac{1}{1 + t^2}. \quad (2.27) $$

Equation (2.27) is the p.d.f. of the standard Cauchy variable.

So far, the special cases of the our non-central skew $t$ distribution have been discussed. In the following we investigate the moments of the non-central skew $t$ distribution.

**Theorem 2.3.3.** Let $T \sim St_r'(\mu, \sigma, \alpha)$. The $k^{th}$ moment of $T$ is given by

$$ ET^k = r^{k/2} EX^k \cdot EY^{-k/2}, \quad (2.28) $$

where $X \sim SN(\mu, \sigma, \alpha)$ and $Y \sim \chi^2_r$.

**Proof.** One of the results from probability theory states that the expected value of a product of independent random variables is the product of their expected values. Hence, the proof follows directly from Definition (2.2.1) and the fact that $X$ and $Y$ are assumed to be independent. \hfill \Box

Theorem (2.3.3) provides a useful method to compute the $k^{th}$ moment of a non-central skew $t$ random variable in terms of the moments of the skew normal and the chi-squared random variables. To use this result, we need an expression for the $k^{th}$ moment of $Y^{-1/2}$ where $Y \sim \chi^2_r$. The following lemma provides an expression for this moment.
Lemma 2.3.4. Let $Y \sim \chi^2_r$, then the $k^{th}$ moment of $Y^{-1/2}$ is given by:

$$EY^{-k/2} = \frac{\Gamma\left(\frac{r-k}{2}\right)}{\Gamma\left(\frac{r}{2}\right)(2)^{k/2}}, \quad \text{where } r > k. \quad (2.29)$$

Proof.

$$EY^{-k/2} = \int_0^\infty y^{-k/2} y^{(r-1)/2} e^{-y/2} \frac{y^{r/2-1} e^{-y/2}}{\Gamma(r/2)2^{r/2}} dy$$

$$= \frac{1}{\Gamma(r/2)2^{r/2}} \int_0^\infty y^{(r-k)/2} e^{-y/2} dy$$

$$= \frac{1}{\Gamma(r/2)2^{r/2}} \Gamma\left(\frac{r-k}{2}\right) 2^{(r-k)/2}, \quad r > k$$

$$= \frac{\Gamma\left(\frac{r-k}{2}\right)}{\Gamma\left(\frac{r}{2}\right)2^{k/2}}.$$

To compute the moments of $X$, we need some results on the $m.g.f.$ of the non-standard skew normal random variable. The following Theorem provides an expression of the $m.g.f.$ of $SN(\mu, \sigma, \alpha)$.

Theorem 2.3.5. Let $X \sim SN(\mu, \sigma, \alpha)$, then the $m.g.f.$ of $X$ is given by:

$$M(t) = 2e^{t\mu+\frac{1}{2}t^2\sigma^2} \Phi(\delta t), \quad \text{where } \delta = \frac{\alpha}{\sqrt{1+\alpha^2}}. \quad (2.30)$$
Proof.

\[ M(t) = E e^{tz} \]
\[ = \frac{2}{\sigma} \int_{-\infty}^{\infty} e^{tz} \phi \left( \frac{z - \mu}{\sigma} \right) \Phi \left( \alpha \frac{z - \mu}{\sigma} \right) \, dz \]
\[ = \frac{2}{\sqrt{2\pi} \sigma} \int_{-\infty}^{\infty} e^{tz} e^{-\frac{1}{2} \left( \frac{z - \mu}{\sigma} \right)^2} \Phi \left( \alpha \frac{z - \mu}{\sigma} \right) \, dz \]
\[ = \frac{2 e^{t\mu}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tu\sigma} - \frac{1}{2} u^2 \Phi (\alpha u) \, du, \quad \text{where} \quad u = \frac{z - \mu}{\sigma}. \]
\[ = \frac{2 e^{t\mu} + \frac{1}{2} t^2 \sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \phi(u - t\sigma) \Phi(\alpha u) \, du \]
\[ = \frac{2 e^{t\mu} + \frac{1}{2} t^2 \sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \phi(w) \Phi(\alpha w + t\alpha \sigma) \, dw, \quad \text{where} \quad w = u - t\sigma. \quad (2.31) \]

Using lemma (1.4.5) in chapter 1, we obtain:

\[ M(t) = 2 e^{t\mu + \frac{1}{2} t^2 \sigma^2} \Phi(\alpha w + t\alpha \sigma) \]
\[ = 2 e^{t\mu + \frac{1}{2} t^2 \sigma^2} \Phi \left( \frac{\alpha \sigma t}{\sqrt{1 + \alpha^2}} \right) \]
\[ = 2 e^{t\mu + \frac{1}{2} t^2 \sigma^2} \Phi (\delta t), \quad \text{where} \quad \delta = \frac{\alpha}{\sqrt{1 + \alpha^2}}. \]

Using Theorem (2.3.5), one can compute the moments of the non-standard skew normal random variable. To compute the \( k \)th moment, the \( k \)th derivative of the \( m.g.f. \) needs to be evaluated at \( t = 0 \). The following lemma is useful in computing moments of the skew normal random variable. It provides the derivative of the normal p.d.f. and c.d.f. respectively.

**Lemma 2.3.6.** Let \( c \) be a real number. Let \( \phi \) and \( \Phi \) denote the standard normal p.d.f. and
c.d.f. respectively. Then

\[
\frac{d}{dt} [\phi(ct)] = -c^2 t \phi(ct), \quad (2.32)
\]

\[
\frac{d}{dt} [\Phi(ct)] = c \phi(ct). \quad (2.33)
\]

**Proof.** Follows directly from the chain rule of differentiation and the fundamental Theorem of calculus. \(\square\)

Using the above results, one can show that the first two moments of the non-standard skew normal variable are given by the following expressions:

\[
EX = \mu + \sqrt{\frac{2}{\pi}} \delta \sigma, \quad (2.34)
\]

\[
EX^2 = \mu^2 + \sigma^2 + \frac{2\mu}{\sqrt{2\pi}} (\mu + \delta \sigma), \quad \text{where } \delta = \frac{\alpha}{1 + \alpha^2}. \quad (2.35)
\]

To illustrate the use of the above Theorems, we provide the first two moments and obtain the variance of \(X\) and use them to obtain an explicit form of the first two moment of the random variable \(T \sim St_\gamma\) and compute its variance.

\[
ET = r^{1/2} EX \cdot EY^{-1/2}
\]

\[
= \sqrt{r} \left[ \mu + \sqrt{\frac{2}{\pi}} \delta \sigma \right] \frac{\Gamma \left( \frac{r-1}{2} \right)}{\Gamma \left( \frac{r}{2} \right)} \sqrt{2} \frac{\Gamma \left( \frac{r-1}{2} \right)}{\Gamma \left( \frac{r}{2} \right)} \sqrt{2} \pi \delta \sigma 
\]

\[
= \sqrt{r/2} \cdot \frac{\Gamma \left( \frac{r-1}{2} \right)}{\Gamma \left( \frac{r}{2} \right)} \left[ \mu + \sqrt{\frac{2}{\pi}} \delta \sigma \right]. \quad (2.36)
\]

Note that the first moment depends on all four parameters and is defined if and only if \(r > 1\).
The second moment of $T$ can be derived as follows:

$$ET^2 = rEX^2 \cdot EY^{-1}$$

$$= r \left[ \mu^2 + \sigma^2 + \frac{2\mu}{\sqrt{2\pi}}(\mu + \sigma\delta) \right] \frac{\Gamma \left( \frac{r-2}{2} \right)}{2\Gamma \left( \frac{r}{2} \right)}$$

$$= \frac{r\Gamma \left( \frac{r-2}{2} \right)}{2\Gamma \left( \frac{r}{2} \right)} \left[ \mu^2 + \sigma^2 + \frac{2\mu}{\sqrt{2\pi}}(\mu + \sigma\delta) \right]$$

$$= \frac{r}{r-2} \left[ \mu^2 + \sigma^2 + \frac{2\mu}{\sqrt{2\pi}}(\mu + \sigma\delta) \right]$$

where $\delta$ is defined in equation (2.35).

The variance of $St'_r$ can be computed as follows:

$$\text{Var}(T) = ET^2 - (ET)^2$$

$$= \frac{r}{r-2} \left[ \mu^2 + \sigma^2 + \frac{2\mu}{\sqrt{2\pi}}(\mu + \sigma\delta) \right]$$

$$- \frac{r}{2} \left\{ \mu + \sqrt{\frac{2}{\pi}}\sigma\delta \frac{\Gamma \left( \frac{r-1}{2} \right)}{\Gamma \left( \frac{r}{2} \right)} \right\}^2$$

(2.38)

Higher order moments can be computed similarly using equation (2.28) of Theorem (2.3.3).

One of the desirable properties of a model is being closed under scalar multiplication. This is particularly helpful if the there is a need to change the scale or the units of measurements in a data set. The following Theorem states that the family of extended skew $t$ distributions is closed under scalar multiplication and specifies the distribution of a scalar multiple of the non-central skew $t$ variable.

**Theorem 2.3.7.** Let $T \sim St'_r(\mu, \sigma, \alpha)$. Let $c$ be a nonzero real number. Let $U = cT$. Then $U \sim St'_r(c\mu, |c|\sigma, \gamma)$, where

$$\gamma = \frac{\alpha c}{\sqrt{1 + \alpha^2(1-c^2)}}.$$  

(2.39)
Proof. It is sufficient to show that if $X \sim SN(\mu, \sigma, \alpha)$, then $cX \sim SN(c\mu, |c|\sigma, \gamma)$.

$$M_{cX}(t) = Ee^{ctX} = M_X(ct) = 2e^{ct\mu + \frac{1}{2}(ct)^2\sigma^2 - \frac{\alpha c}{\sqrt{1 + \alpha^2}}\Phi\left(\frac{\alpha c}{\sqrt{1 + \alpha^2}}\sigma t\right)}.$$  

Comparing to the $m.g.f.$ of the skew normal distribution given in equation (2.30), we see that $cX \sim SN(c\mu, |c|\sigma, \gamma)$ where $\gamma$ satisfies the equation

$$\frac{\gamma}{\sqrt{1 + \gamma^2}} = \frac{\alpha c}{\sqrt{1 + \alpha^2}}.$$  

(2.40)

Solving for $\gamma$, yields:

$$\gamma = \frac{\alpha c}{\sqrt{1 + \alpha^2(1 - c^2)}}.$$  

(2.41)

From the construction of our $St_r'$ variate we see that if $T \sim St_r' (\mu, \sigma, \alpha)$, then

$$T \overset{D}{=} \frac{X}{\sqrt{Y/r}},$$

where $X \sim SN(\mu, \sigma, \alpha)$, $Y \sim \chi^2_r$ and $X, Y$ are independent. Hence,

$$cT \overset{D}{=} \frac{cX}{\sqrt{Y/r}}.$$  

Consequently, $cT \sim St_r' (c\mu, |c|\sigma, \gamma)$.

In the remaining part of this subsection, some limiting properties of the family of non-central skew $t$ distributions will be explored.

**Theorem 2.3.8.** Let $T_r \sim St_r'(\mu, \sigma, \alpha)$. Then

$$\lim_{r \to \infty} T_r \overset{a.s.}{=} SN(\mu, \sigma)$$.
Proof. Let \( X, Y \) be two independent random variables with \( X \sim SN(\mu, \sigma, \alpha) \), and \( Y \sim \chi^2_r \).

Since \( T \overset{D}{=} \frac{X}{\sqrt{Y/r}} \), it suffices to prove that \( \lim_{r \to \infty} \frac{Y}{r} = 1 \).

Let \( Y_1, Y_2, \ldots, Y_r \) be independent identically distributed random variables with the \( \chi^2_1 \) distribution. A standard result in distribution theory states that:

\[
Y \overset{D}{=} \sum_{i=1}^{r} Y_i.
\]

In other words, \( \frac{Y}{r} \) has the same distribution as the sample mean of \( Y_1, Y_2, \ldots, Y_r \).

The fourth moment of \( Y_1 \) is finite, thus the strong law of large numbers implies that the sample mean converges to the population mean almost surely as the sample size \( r \to \infty \).

With \( EY_1 = 1 \), we have

\[
\frac{Y}{r} \xrightarrow{a.s.} E(Y_1) = 1. \tag{2.42}
\]

Since \( X, Y \) are independent, Slutsky’s Theorem implies that if \( r \to \infty \), then

\[
\frac{X}{\sqrt{Y/r}} \overset{D}{\to} X. \tag{2.43}
\]

Hence, the limiting distribution of \( T \sim St'_r(\mu, \sigma, \alpha) \) as \( r \to \infty \) is \( SN(\mu, \sigma, \alpha) \).

Theorem (2.3.8) states that for large enough values of the degrees of freedom, the non-central skew \( t \) distribution can be approximated by the skew normal distribution that has the same location, scale and shape parameters. The following Theorem explores the limiting distribution of a special class of the non-central skew \( t \) family as the shape parameter tends to infinity.

**Theorem 2.3.9.** Let \( X \sim St'_r(0, 1, \alpha) \). The limiting distribution of \( X \) as \( \alpha \to \infty \) is \( |t_r| \), where \( t_r \) denotes the Student’s \( t \) distribution.

**Proof.** The proof follows from the Definition of the non-central skew \( t \) distribution and property (iii) of Theorem (1.3.2) of chapter 1, which states that as \( \alpha \to \infty \), \( SN(\alpha) \to HN(0, 1) \), where \( HN \) denotes the half normal distribution.
2.4 NUMERICAL RESULTS AND GRAPHICAL ILLUSTRATIONS

2.4.1 INTRODUCTION

This section includes numerical analyses and graphical illustrations of the density curves of the $St'_r$ distribution. The programming language R has been used throughout this section to perform the computations. The graphs are obtained by applying numerical integration to evaluate the integral representation of the density of the $St'_r$ random variable as outlined in equation (2.6).

2.4.2 GRAPHICAL ILLUSTRATION OF THE $St'_r$ DENSITY

Theoretical properties of the non-central skew $t$ density have been discussed in the previous sections. In this section, graphical illustrations of these properties are provided. Figure (2.1) provides an illustration of Theorem (2.3.2), it shows that the graph of $St'_1(0,1,0)$ coincides with the graph of the standard Cauchy density.

To illustrate the effect of each one of our the four parameters in determining the overall shape of the density, we present a series of graphs with three parameters fixed and the fourth one varying. Figure (2.2) provides an illustration of the effect of the degrees of freedom on the shape of the density. The graph of $St'_r(0,1,-1)$ has been given for values of $r = 1, 3, 5, 10, 15, 30$. The graph indicates that the smaller the value of $r$, the thicker the tail of the distribution. The density with $r = 1$ has the thickest tails and the density with $r = 30$ had the thinnest tail in the group. Notice that as $r$ increases the shape of the density gets closer to that of the corresponding skew normal density.
Figure 2.1: The density of $St'_1(0, 1, 0)$ coincides with $Cauchy(0, 1)$ density.

Figure 2.2: Graphs of the $p.d.f.$ of $St'_r(0, 1, -1)$ for several values of $r$.

Figure (2.3) provides an illustration of effect of the shape parameter $\alpha$ on the shape of the density. The graph of $St'_3(0, 1, \alpha)$ has been given for values of $\alpha = 1, -1, 2, -2, 3, -3$. Because we restrict the location and scale parameters to 0, 1 respectively, the graph indicates that $-St'_3(0, 1, \alpha) = St'_3(0, 1, -\alpha)$. This result is true whenever $r$ is fixed and $\mu = 0, \sigma = 1$. It follows directly from part (ii) of Theorem (1.3.2) in chapter 1. The graph illustrates the
special case of Theorem (2.3.7) with $c = -1$.

Figure 2.3: Graphs of the p.d.f. of $St'_3(0,1,\alpha)$ for several values of $\alpha$.

Figure 2.4: Graphs of the p.d.f. of $St'_3(\mu,1,1)$ for several values of $\mu$. 
Figure (2.4) illustrates the effect of the location parameter $\mu$ on determining the overall shape of the density. Notice that the graphs represented agree on all other three parameters yet they are not location shifts of each other. The non-central skew $t$ family of distributions is not a location-scale family. Figure (2.5) provides an illustration of the effect of the scale parameter $\sigma$ on determining the overall shape of the density. It shows that when the degrees of freedom parameter is fixed, the thickness of the tail of the density can still be controlled by the parameter $\sigma$. Thus, the non-central skew $t$ distribution has two parameters $r$ and $\sigma$ that control the thickness of the tail of the density.

Figure 2.5: Graphs of the p.d.f. of $St'_3(0, \sigma, 1)$ for selected values of $\sigma$.

Figure (2.6) provides a graphical illustration of the property that the limiting distribution of the $St'_r(0,1,\alpha)$ as $r$ tends to infinity is the $SN(\alpha)$ density. With $r = 40$ and $\alpha = -1$, the two density curves are coinciding. Figure (2.7) provides another graphical example where the density of $St'_r(2,2,-1)$ approaches the density of $SN(2, 2, -1)$ as $r$ increases. Graphically, at $r = 40$ the two density curves coincide. Figure (2.8) illustrates Theorem (2.3.9). It shows the density of the $|t_3|$ distribution versus the densities of $St'_3(0,1,\alpha)$ for several values of $\alpha$. As $\alpha$ increases the density of $St'_3(0,1,\alpha)$ approaches the density of the $|t_3|$ distribution.
Figure 2.6: Graphs of the p.d.f. of $St'_r(0,1,1)$ as $r$ increases versus the $SN(1)$ density.

Figure 2.7: Graphs of the p.d.f. of $St'_r(2,2,-1)$ as $r$ increases versus the density of $SN(2,2,-1)$. 

Figure 2.8: Graphs of the p.d.f. of $St_3'(0,1,\alpha)$ as $\alpha$ increases they approach the density of $|t_3|$.

2.4.3 GENERATING RANDOM SAMPLES FROM THE $St_r'$ DISTRIBUTION

Definition (2.2.1) of the non-central skew $t$ distribution provides a stochastic representation of the variable $St_r'$. Thus the following algorithm can be used to generate a random sample of size $n$ from the $St_r'(\mu, \sigma, \alpha)$:

(i) Generate a random sample $x_1, x_2, ..., x_n$ of size $n$ from the skew normal distribution $SN(\mu, \sigma, \alpha)$.

(ii) Generate an independent random sample $y_1, y_2, ..., y_n$ of size $n$ from the chi-squared distribution $\chi_r^2$.

(iii) Apply the transformation defined in equation (2.47) to the two independent samples generated above. Namely, set $z_i = \frac{x_i}{\sqrt{y_i/r}}$, $i = 1, 2, ..., n$.

(iv) The resulting sample $z_1, z_2, ..., z_n$ will be a random sample from the $St_r'(\mu, \sigma, \alpha)$ family.
To illustrate the algorithm outlined above, the programming code has been provided in the appendix. Figure 2.9 shows four examples where the algorithm above is used to generate a random sample of size 5000 from the non-central skew $t$ distribution. The summary of the simulated data is presented graphically by the histogram and the theoretical density curve is superimposed on it to verify that they agree.

Figure 2.9: Histograms of simulated $St'_r$ data versus theoretical density curve.
2.5 A MULTIVARIATE NON-CENTRAL SKEW \( t \) DISTRIBUTION

2.5.1 INTRODUCTION

A non-central skew \( t \) distribution in the univariate setting has been introduced in this chapter. In this section we present the multivariate generalization of our \( St_r'(\mu, \sigma, \alpha) \) distribution and present some properties in the multivariate setting.

2.5.2 A MULTIVARIATE NON-CENTRAL SKEW \( t \) DISTRIBUTION

The multivariate standard skew normal distribution is introduced in Chapter 1. The density of this random variable was provided in Definition (1.4.4). In this section, we use this multivariate skew normal distribution to define a multivariate version of our non-central skew \( t \) distribution. Gupta (2003) provided a version of multivariate skew \( t \) distribution as follows:

Definition 2.5.1. (Gupta’s Multivariate Skew \( t \) Distribution):

Let \( \mathbf{X} = (X_1, X_2, \ldots, X_k)^T \) be a \( k \)-dimensional standard skew normal variable with correlation matrix \( (\mathbf{\Omega}) \) and skewness vector \( (\mathbf{\alpha}) \). Let \( W \) be a univariate \( \chi^2 \) variable with \( r \) degrees of freedom. Assume further that \( X_i \) is independent of \( W \) for \( i = 1, 2, \ldots, k \). Then the \( k \)-dimensional skew \( t \) random variable with parameters \( (\mathbf{\Omega}, \mathbf{\alpha}) \) and \( r \) degrees of freedom is defined to be the joint distribution of \( (Y_1, Y_2, \ldots, Y_k)^T \), where

\[
Y_i = \frac{X_i}{\sqrt{W/r}}, \quad i = 1, 2, \ldots, k. \tag{2.44}
\]

Gupta obtained the following expression for the p.d.f. of his multivariate skew \( t \) variate \( \mathbf{Y} \):
\[ f_r(y, \alpha) = 2 f_{r_k}(y) F_{r+k} \left( \frac{\sqrt{r+k} \alpha^T y}{(r+y^T \Sigma^{-1} y)^{1/2}} \right), \quad y \in \mathbb{R}^k, \tag{2.45} \]

where \( f_{r_k} \) and \( F_{r_k} \) denote the p.d.f. and c.d.f. of the central multivariate t distribution with \( r \) degrees of freedom respectively.

Definition (2.5.1) used the multivariate skew normal that has been presented in Definition (1.4.4) of chapter 1. This multivariate normal distribution can be extended to a more general location-scale family of densities that includes it as a special case. The following extended multivariate skew normal distribution was introduced in Azzalini, et al. (1999).

**Definition 2.5.2. (Extended Multivariate Skew Normal Distribution):** A random vector \( Z = (Z_1, Z_2, ..., Z_k)^T \) is said to have the \( k \)-dimensional extended skew normal distribution if its density is given by:

\[ f_k(z; \mu, \lambda, \Omega) = 2 \phi_k(z; \mu, \Omega) \Phi(\lambda^T \omega^{-1}(z - \mu)); \quad z \in \mathbb{R}^k. \tag{2.46} \]

where \( \phi_k(z; \Omega) \) denotes the density of the \( k \)-dimensional multivariate normal distribution with correlation matrix \( \Omega \) and \( \omega \) is the diagonal matrix formed by the standard deviations of \( \Omega \). We denote such a random vector by \( Z \sim SN_k(\mu, \lambda, \Omega) \). The vectors \( \mu, \lambda \) are also called location and shape parameters respectively.

The multivariate skew t distribution that was introduced by Gupta (2003) is restricted to zero location parameter. This restriction simplifies the computations for the p.d.f. and allows for a nice closed form of the density. In this section, we present the following extension of Definition (2.5.1). Our extension provides more flexibility by adding an arbitrary vector for the location parameter.

**Definition 2.5.3. (Multivariate non-central Skew t Distribution):**

Let \( X = (X_1, X_2, ..., X_k)^T \) be a \( k \)-dimensional extended skew normal variable with location
parameter \( \mu \), correlation matrix \( \Omega \) and skewness vector \( \alpha \). Let \( Y \) be a univariate \( \chi^2 \) variable with \( r \) degrees of freedom. Assume further that \( X \) and \( Y \) are independent, that is, \( X_i \) is independent of \( Y \) for \( i = 1, 2, \ldots, k \). Then the \( k \)-dimensional non-central skew \( t \) random variable with parameters \( (\mu, \Omega, \alpha) \) and \( r \) degrees of freedom is defined to be the joint distribution of \( (Z_1, Z_2, \ldots, Z_k)^T \), where

\[
Z_i = \frac{X_i}{\sqrt{Y/r}}, \quad i = 1, 2, \ldots, k.
\]

(2.47)

The notation \( Z \sim MSt_k^r(\mu, \Omega, \alpha, r) \) will be used to denote this density.

There is no closed form for the density of the multivariate non-central skew \( t \) distribution defined above. However, the availability of numerical integration tools makes an integral form for the density sufficient for practical applications. We proceed to study the theoretical properties of the non-central multivariate skew \( t \) distribution.

One of the nice properties that the multivariate skew normal distribution enjoys is being closed under affine transformations. The following Theorem provides the exact distribution of the random variable that results from applying an affine transformation to a multivariate skew normal.

**Theorem 2.5.4. (The Affine Transformation of the Skew Normal Family):**

Let \( Z \sim SN_d(\xi, \Omega, \alpha) \), \( a \in \mathbb{R}^h \) and let \( B \) be an \( h \times d \) matrix with \( h \leq d \). If \( W = a + BZ \), then \( W \sim SN_h(\xi_W, \Omega_W, \alpha_W) \), where

\[
\begin{align*}
\xi_W &= a + B\xi, \\
\Omega_W &= B\Omega B^T, \\
\alpha_W &= \frac{1}{(1 + \alpha^T(\bar{\Omega} - H\Omega_W^{-1}H^T)\alpha)^{1/2}}\omega_W \Omega_W^{-1}H^T\alpha, \quad H = \omega^{-1}\Omega B^T,
\end{align*}
\]

(2.48)

\( \omega_W \) denotes the diagonal matrix of standard deviations of \( \Omega_W \), and \( \bar{\Omega} = \omega^{-1}\Omega \omega^{-1} \) denotes
the correlation matrix associated with $\Omega$.

Proof. See Azzalini (2005) for the proof and more details. \qed

2.5.3 PROPERTIES OF THE $\text{MSt}_k'$ DISTRIBUTION

In this section, the closure under linear transformations of the multivariate non-central skew $t$ distributions is studied. The following Theorem provides the distribution of a linear transformation of a multivariate non-central skew $t$ variate.

Theorem 2.5.5. Let $Z \sim \text{MSt}_k'((\mu, \Omega, \alpha, r)$, $\mu, \alpha \in \mathbb{R}^k$ and let $\mathbf{B}$ be an $h \times k$ matrix with $h \leq k$. If $W = \mathbf{B}Z$, then $W \sim \text{MSt}_h'((\xi_W, \Omega_W, \alpha_W, r)$, where

\begin{align*}
\xi_W &= \mathbf{B}\xi, \\
\Omega_W &= \mathbf{B}\Omega\mathbf{B}^T, \\
\alpha_W &= \frac{1}{(1 + \alpha^T(\Omega - H\Omega_W^{-1}H^T)\alpha)}^{1/2} \omega_W \Omega_W^{-1}H^T\alpha, \\
H &= \omega^{-1}\Omega\mathbf{B}^T, \\
\omega_W &= \text{the diagonal matrix of standard deviations of } \Omega_W.
\end{align*}

(2.49)

Proof. Consider Theorem (2.5.4). By taking $a = 0$, we see that if $X \sim \text{SN}_k(\xi, \Omega, \alpha)$, then $BX \sim \text{SN}_h(\xi_W, \Omega_W, \alpha_W)$. Definition (2.5.3) implies that if $W$ is a univariate $\chi^2$ random variable with $r$ degrees of freedom which is independent of $X$, then

$$
\frac{BX}{\sqrt{W/r}} \sim \text{MSt}_h'((\xi_W, \Omega_W, \alpha_W, r).
$$

Therefore,

$$
W \sim \text{MSt}_h'((\xi_W, \Omega_W, \alpha_W, r).
$$

\qed
2.6 QUADRATIC FORMS OF THE $S_{t'}_r$ DISTRIBUTION

One of the properties of Student’s $t$ distribution is that if $T \sim t_r$, then $T^2 \sim F(1, r)$. Our next goal is to investigate the distribution of $T^2$ if $T$ follows the non-central skew $t$ distribution. To achieve this goal, we need to discuss the distribution of $Z^2$ where $Z \sim SN(\mu, \sigma, \alpha)$.

Distributions of quadratic forms under skew normal settings were studied in many recent research papers. Recall that the non-central $\chi^2$ distribution with $k$ degrees of freedom and non-centrality parameter $\lambda$ is defined as the distribution of $X'X$ where $X \sim N_k(\nu, I_k)$. Wang, et al. (2009) introduced the non-central skew $\chi^2$ distribution using the multivariate skew normal random variable. He defined it as follows:

**Definition 2.6.1. (Wang, et al., 2009):** Let $X \sim SN_k(\nu, I_k, \alpha)$. The distribution of $X'X$ is called the non-central skew $\chi^2$ distribution with $k$ degrees of freedom, the non-centrality parameter $\lambda = \nu'\nu$ and the skewness parameter $\alpha$.

We will denote the non-central skew $\chi^2$ distribution by $S\chi^2_k(\lambda, \alpha)$. Note that when $\alpha = 0$, it is easy to see that the non-central skew $\chi^2$ distribution is reduced to the non-central $\chi^2$ distribution and it is free of the skew parameter $\alpha$. Definition (2.6.1) provides a Definition of the univariate non-central skew $\chi^2$ distribution by taking $k = 1$. The p.d.f. and m.g.f. of this random variable are provided in Wang, et al (2009). A more general result that addresses quadratic forms of the multivariate skew normal distribution as presented in the following Theorem.

**Theorem 2.6.2. (Wang, et al., 2009):** Let $B$ be a $k \times k$ positive definite matrix, and let $R^{k \times k}$ be the class of all $k \times k$ real valued matrices. If $Y \sim SN_k(\mu, B, \alpha)$ and $Q = Y'WY$ with nonnegative definite $W \in R^{k \times k}$, then the necessary and sufficient conditions under which $Q \sim S\chi^2_k(\lambda, \alpha_*)$, for some $\alpha_*$ including $\alpha_* = 0$, are:
(i) $BW B'$ is idempotent with rank $k$,

(ii) $\lambda = \mu' W \mu = \mu' W B' B W \mu$,

(iii) $\alpha' B W \mu = c \alpha', \nu$,

(iv) $\alpha', \alpha = c^{-2} \alpha' P_1 P_1' \alpha$,

where $c = \sqrt{1 + \alpha' P_1 P_1' \alpha}$, $\nu = P_1' B W \mu$, and $P = (P_1, P_2)$ is an orthogonal matrix in $\mathbb{R}^{k \times k}$ such that

$$BW B' = P \begin{pmatrix} I_m & 0 \\ 0 & 0 \end{pmatrix} P' = P_1 P_1'.$$

\(2.50\)

Proof. See the proof of Theorem (4.1) of Wang et al. (2009).

Theorem (2.6.2) states that under certain restrictions, the distribution of a quadratic form of the non-standard multivariate skew normal distribution is a skew chi-squared distribution. Based on this, one would expect the distribution of quadratic forms of the non-central skew $t$ random variable to be related to a generalization of the $F$ distribution. To prove this, we need to study non-central and skewed versions of the $F$ distribution. The non-central $F$ distribution is defined in Johnson et al. (1995) as follows:

**Definition 2.6.3. (Doubly non-central $F$ Distribution):** The doubly non-central $F$ distribution with $\nu_1, \nu_2$ degrees of freedom, and non-centrality parameters $\lambda_1, \lambda_2$ is defined to be the distribution of the following ratio

$$\frac{\chi^2_{\nu_1}(\lambda_1)/\nu_1}{\chi^2_{\nu_2}(\lambda_2)/\nu_2},$$

\(2.51\)

where the numerator and the denominator are assumed to be independent. We use the notation $F_{\nu_1, \nu_2}^\prime(\lambda_1, \lambda_2)$ to denote this distribution. In the case $\lambda_1 = 0$, the distribution resulting from the ratio is referred to as “singly non-central” and denoted by $F_{\nu_1, \nu_2}(\lambda_2)$. 

One can define a skewed version of the non-central $F$ distribution by replacing the numerator of equation (2.51) by a non-central chi-squared random variable. Thus we have the following Definition:

**Definition 2.6.4. (Non-central Skew $F$ Distribution):** Let $X \sim \chi^2_{\nu_1} (\lambda_1, \alpha)$. Let $Y \sim \chi^2_{\nu_2} (\lambda_2)$ be independent of $X$. The random variable $W$ given by

$$W = \frac{X/\nu_1}{Y/\nu_2},$$

(2.52)

is defined to be a doubly non-central skew $F$ distribution with degrees of freedom $\nu_1, \nu_2$, non-centrality parameters $\lambda_1, \lambda_2$ and skewness coefficient $\alpha$. The notation $SF''_{\nu_1,\nu_2} (\lambda_1, \lambda_2, \alpha)$ will be used to denote this random variable. If $\lambda_2 = 0$, then this random variable will reduce to a singly non-central skew $F$ distribution, and will be denoted by $SF'_{\nu_1,\nu_2} (\lambda_1, \alpha)$.

Based on the discussion above, the distribution of quadratic forms involving the non-central skew $t$ distribution can be determined. The following Theorem specifies the distribution of $T^2$, where $T \sim St'_r (\mu, 1, \alpha)$.

**Theorem 2.6.5.** Let $T \sim St'_r (\mu, 1, \alpha)$, then $T^2 \sim SF'_{1,r} (\mu^2, \alpha)$.

*Proof.* The proof is immediate from the construction of the $St'_r$ random variable and Definitions (2.6.1) and (2.6.4). $\square$

Note that Theorem (2.6.5) is restricted to the case where the scale parameter $\sigma = 1$ to be consistent with the results above. However, Theorem (2.3.7) implies a random variable $T \sim St'_r (\mu, \sigma, \alpha)$ can be reduced to a random variable with distribution $St'_r (\mu, 1, \alpha)$ by multiplying by $1/\sigma$. Hence, a quadratic form of $T \sim St'_r (\mu, \sigma, \alpha)$ where $\sigma \neq 1$ will have the distribution of $\sigma^{-2}SF'_{1,r} (\mu^2, \alpha)$. A multivariate version of Theorem (2.6.5) is presented as follows.

**Theorem 2.6.6.** Let $T \sim MSSt'_k (\mu, I_k, \alpha, r)$. The distribution of $T'T$ is the singly non-central skew $F$ distribution with degrees of freedom $(1, r)$, the non-centrality parameter $\lambda = \ldots$
\( \mu' \mu \) and the skewness parameter \( \alpha \).

Proof. Follows immediately from Definition (2.6.1) and Definition (2.5.3).

We conclude this section by a more general result that discusses conditions under which the distribution of quadratic forms involving the multivariate non-central skew \( t \) distribution follows the multivariate non-central skew \( t \) distribution.

**Theorem 2.6.7. (Quadratic Forms of non-central Skew \( t \) Distributions):**

Let \( B \) be a \( k \times k \) positive definite matrix. Let \( \mathbb{R}^{k \times k} \) be the class of all \( k \times k \) real valued matrices. If \( Y \sim \text{MSt}_k(\mu, B, \alpha, r) \) and \( Q = Y' W Y \) with nonnegative definite \( W \in \mathbb{R}^{k \times k} \), then the necessary and sufficient conditions under which \( Q \sim \text{SF}_{r,1}(\lambda, \alpha_*), \) for some \( \alpha_* \) including \( \alpha_* = 0 \), are:

(i) \( BWB' \) is idempotent with rank \( k \),

(ii) \( \lambda = \mu' W \mu = \mu' WB' BW \mu \),

(iii) \( \alpha'BW \mu = c\alpha_*' \nu \),

(iv) \( \alpha_*' \alpha_* = c^{-2} \alpha' P_1 P_1' \alpha \),

where \( c = \sqrt{1 + \alpha' P_1 P_1' \alpha} \), \( \nu = P_1' BW \mu \), and \( P = (P_1, P_2) \) is an orthogonal matrix in \( \mathbb{R}^{k \times k} \) such that

\[
BWB' = P \begin{pmatrix} I_m & 0 \\ 0 & 0 \end{pmatrix} P' = P_1 P_1'.
\]

Proof. The results follows by applying Theorem (2.6.2), and Definitions (2.5.3) and (2.6.4).
2.7 GENERALIZED NON-CENTRAL SKEW \textit{t} DISTRIBUTION

2.7.1 INTRODUCTION

In this section, we focus on an alternative approach to define a family of generalized skew \textit{t} distributions. The work in this chapter is a generalization of the results in the manuscript entitled “A study of generalized skew \textit{t} distribution” by Huang, et al. (2012) (obtained by personal communication with the authors). We will first list the definitions and Theorems from Huang, et al. (2012) that are needed in this section.

2.7.2 LITERATURE RESULTS ON THE GENERALIZED SKEW \textit{t} DISTRIBUTIONS

Definition 2.7.1. (Generalized Skew Distribution) (Huang et al.) Let \( Y \) be a real valued random variable that is symmetric around 0. Let \( g \) be the p.d.f. of \( Y \) such that \( g(z) = g(-z) \), for all \( z \in \mathbb{R} \). A random variable \( Z \) is said to have a generalized skew distribution of \( g \), if

\[
Z^2 \overset{D}{=} Y^2.
\]

(2.54)

where \( \overset{D}{=} \) indicates have the same distribution.

Definition (2.7.1) was introduced in Huang, et al. (2012). An alternative representation of the of the generalized skew distribution that is introduced in Definition (2.7.1) is given by the following Theorem.

Theorem 2.7.2. Let \( g(.) \) be a symmetric p.d.f. for some random variable. A random variable \( Z \) has a generalized skew distribution of \( g \) if and only if its p.d.f. is given by:

\[
f_Z(z) = 2g(z)\pi(z), \text{ for } z \in \mathbb{R},
\]

(2.55)
where \( \pi \) is a Lebesgue measurable function satisfying the conditions

\[
0 \leq \pi(z) \leq 1 \quad \text{and} \quad \pi(z) + \pi(-z) = 1,
\]

almost everywhere on \( \mathbb{R} \). Such \( \pi \) is said to be a skew function.

Proof. The readers are referred to Arellano-Valle and Genton (2005) for the proof and further details. In fact, Arellano-Valle and Genton (2005) proved the equivalence of the representation in this Theorem and the one in Definition (2.7.1).

One can easily see that the generalized skew random variable \( Z \) in Theorem (2.7.2) is symmetric if and only if \( \pi(z) = \frac{1}{2} \), for all \( z \in \mathbb{R} \). A special case of skew function would be taking \( \pi(z) = F(\lambda z) \), where \( F \) denotes a c.d.f. of some random variable and \( \lambda \) is a non-zero constant. Recall that Azzalini’s Definition of the standard skew normal distribution used this choice of the skew function. In fact, most of the literature on the skew \( t \) distribution relies on this particular choice of skew function or some variation of it. In general, the argument, \( \lambda z \), of the skew function can be replaced by any odd function of \( z \). Azzalini and Huang among others used the statement in Theorem (2.7.2) or a special case of it under a specific skew function to provide a Definition of generalized skew \( t \) distribution as follows:

**Definition 2.7.3.** (Generalized Skew \( t \) Distributions) \( X \) is said to have a generalized skew \( t \) distribution with \( r \) degrees of freedom, denoted by \( X \sim \text{GST}_r \), if the p.d.f. of \( X \) is given by:

\[
f_X(x) = 2f_{t_r}(x)\pi(x), \quad x \in \mathbb{R},
\]

where \( f_{t_r} \) is the p.d.f. of the Student’s \( t \) density, and \( \pi \) is a skew function.

The special case of Definition (2.7.3) where \( \pi \) is chosen to be a c.d.f. of an odd function of \( x \) has been used extensively throughout the literature. Nadarajah (2008) studied the generalized skew \( t \) densities in Definition (2.7.3) under specific choices of the skew function
π. He derived some of their theoretical properties and moments.

Recall that the Student’s $t_r$ distribution converges to $N(0,1)$ as $r \to \infty$, thus the following result is immediate.

**Theorem 2.7.4.** Let $X \sim GSt_r$, with the skew function not depending on $r$. Then $X \overset{D}{\to} U$ as $r \to \infty$, where $U$ is a generalized skew $N(0,1)$ distributed.

**Proof.** Let $X \sim GSt_r$. Then there is a skew function $\pi_0$ such that the p.d.f. of $X$ takes the form:

$$f_X(x) = 2f_{t_r}(x)\pi_0(x).$$

Taking the limit of both sides as $r \to \infty$, we get

$$\lim_{r \to \infty} f_X(x) = 2 \lim_{r \to \infty} [f_{t_r}(x)]\pi_0(x),$$

$$= 2\phi(x)\pi_0(x),$$

where $\phi$ denotes the standard normal p.d.f. Definition (2.7.2) implies that the limiting distribution obtained above is the generalized skew normal distribution. 

2.7.3 GENERALIZED NON-CENTRAL SKEW $t$ DISTRIBUTION

Note that Definition (2.7.1) and Theorem (2.7.2) were restricted in Huang, et al. (2012) to the case where the p.d.f. being generalized is symmetric about 0. Of course, this restriction makes the proofs and computations simpler but it is not necessary. In the following, we present a generalization of Definition (2.7.3). In particular; this idea can be used to define generalized non-central skew $t$ distributions as follows:

**Definition 2.7.5.** *(Generalized non-central Skew $t$ Distribution)*: A random variable $X$ is said to have a generalized non-central skew $t$ distribution with $r$ degrees of freedom
and non-centrality parameters $\lambda$, denoted by: $X \sim GSt_r(\lambda)$, if the p.d.f. of $X$ is given by:

$$f_X(x) = 2f_{T_r(\lambda)}(x)\pi(x), \quad x \in \mathbb{R},$$

(2.60)

where $\pi$ is a skew function, and $f_{T_r(\lambda)}$ is the density of the non-central skew $t$ variable presented in Definition (2.2.1) with $r$ degrees of freedom, the location parameter $\lambda$ and scale parameter $\sigma = 1$ and shape parameter $\alpha$.

It is easy to see that this Definition reduces to Definition (2.7.3) when $\lambda = \alpha = 0$. Also note that Definition (2.7.5) provides a family of distributions that vary by choosing a different skew function $\pi$. A similar approach of defining a skew $t$ distribution was presented in several papers in the literature. However, none of them considered using a non-central $t$ density. Another aspect in which the two definitions are different is the fact that the skew function $\pi$ used here need not be a c.d.f.

An alternative method to construct a generalized non-central skew $t$ distribution can be introduced in the following Definition:

**Definition 2.7.6.** Let $U$ follow a generalized skew $N(\mu, 1, \lambda)$ distribution, with $\mu \in \mathbb{R}$. Let $V$ follow a generalized skew distribution of $\chi^2_r$. Suppose further that $U$ and $V$ are independent. Then

$$T = \frac{U}{\sqrt{V/r}},$$

(2.61)

is said to have a generalized skew non central $t$ distribution with $r$ degrees of freedom, shape parameter $\mu$ and skewness coefficient $\lambda$. We will use the notation $GSt_r(\mu, 1, \lambda)$ to denote this density.

Note that this Definition provides an extension to the Definition proposed by Huang, et al. (2012) by allowing $U$ to be a generalized skew normal with an arbitrary mean rather than restricting its mean to zero. It also provides an extension to Definition (2.2.1). The following result illustrates the limit of the random variable $GSt_r$ as $r \to \infty$. 

**Theorem 2.7.7.** Let $T_r \sim GS_{\mu}(\mu, 1, \lambda)$, then

$$\lim_{r \to \infty} T_r \overset{D}{=\rightarrow} GSN(\mu, 1, \lambda).$$ (2.62)

*Proof.* Follows immediately from Definition (2.7.5) and the proof of Theorem (2.3.8). □

Finally, one can introduced generalized non-central skew $t$ distributions in a similar manner by replacing the univariate non-central skew $t$ variable in definitions (2.7.5) and (2.7.6) above by its multivariate extension. Definition (2.7.5) provides a family of generalized non-central skew $t$ distributions that vary in shape and properties depending on the particular choice of the skew function. The readers are referred to Arellano-Valle, et al. (2005) for a discussion of multivariate generalized skew distributions in general.
CHAPTER 3

APPLICATIONS TO DATA ANALYSIS AND MODEL COMPARISONS

3.1 INTRODUCTION

In this chapter we study the applications of the non-central skew $t$ distribution in data analysis. We start by conducting simulation studies to explore the effectiveness of the extended skew $t$ distribution in fitting long-tailed data. Next, we proceed to conduct model comparisons for the skew $t$ distribution using the information criteria. The models that are included in the comparison are:

- Normal Distribution.
- Skew Normal Distribution.
- Our non-central Skew $t$ Distribution.
- Truncated Skew Normal Distribution.
- Truncated Normal Distribution.
The Akaike and Schwartz information criterion will be used to compare the fitted models to the given data set throughout this chapter. The Akaike information criterion was developed by Hirotugu Akaike in 1973, under the name of “an information criterion”.

**Definition 3.1.1.** The Akaike information criterion (AIC) for a given model is given by:

$$AIC = -2 \log L(\hat{\theta}) + 2k,$$

where $L(\hat{\theta})$ is the maximum likelihood of the data under the assumed model and $k$ is the number of parameters in the selected model.

A model that minimizes the AIC is called minimum AIC estimate, abbreviated as MAICE, and is considered to be the most appropriate model. However, the MAICE is not an asymptotically consistent estimator of model order (see Schwarz, 1978). Some authors made efforts to modify the information criterion without violating Akaike’s original principles. One of the modifications is the Schwarz’s information criterion, denoted by SIC, and proposed by Schwarz in 1978. It is defined as:

**Definition 3.1.2.** The Schwarz’s information criterion of a given model is given by

$$SIC = -2 \log L(\hat{\theta}) + k \cdot \log(n),$$

where $L(\hat{\theta})$ is the maximum likelihood of the data under the assumed model, $k$ is the number of parameters in the selected model and $n$ is the sample size.

### 3.2 Simulation Study

In this section we illustrate the superior behavior of the extended skew $t$ distribution in modeling long tailed data. Inverse Gaussian and log normal distributions are typical examples of distributions that can be used to model long tailed data. We simulate a data set form each
one of them and conduct model comparisons to illustrate the superiority of the extended skew $t$ distribution over that of the skew normal and the normal distribution.

We start with a sample of size 5000 generated from the inverse Gaussian distribution with location 0 and scale 1. We use MLE to find the parameters of the fitted normal, skew normal and non-central skew $t$. Due to the domain of the data, we use the truncated normal and truncated skew normal with truncation over the interval $(0, \infty)$. The AIC and SIC are used to compare the goodness of fit for the three models. Table (3.1) illustrates the summary of the three models when used to fit the data. AIC and SIC criteria agree on favoring the extended skew $t$ distribution for fitting the simulated inverse Gaussian data. Figure (3.1) illustrates the histogram of the simulated data and the three fitted density curves. Graphically, it is clear that density curve of the extended skew $t$ distribution provides the best fit to the data compared to skew normal and normal densities.

<table>
<thead>
<tr>
<th>Model</th>
<th>$\mu$</th>
<th>$\sigma$</th>
<th>$\alpha$</th>
<th>$r$</th>
<th>AIC</th>
<th>SIC</th>
<th>LogL</th>
</tr>
</thead>
<tbody>
<tr>
<td>$St_r$</td>
<td>0.12</td>
<td>0.32</td>
<td>5.89</td>
<td>1.18</td>
<td>8961.24</td>
<td>8987.30</td>
<td>-4476.62</td>
</tr>
<tr>
<td>Normal</td>
<td>0.99</td>
<td>1.37</td>
<td>NA</td>
<td>NA</td>
<td>17409.47</td>
<td>17422.50</td>
<td>-8702.73</td>
</tr>
<tr>
<td>Skew Normal</td>
<td>0.04</td>
<td>1.67</td>
<td>346.95</td>
<td>NA</td>
<td>12440.32</td>
<td>12459.87</td>
<td>-6217.16</td>
</tr>
</tbody>
</table>

Our next simulation study involves simulating data from the log normal distribution. We use the R built in routine to produce a simple random sample of size 5000 generated from the log-normal distribution with location parameter 0, and scale parameter 1. We use our extended skew $t$ model, the truncated skew normal distribution and the truncated normal distribution to fit the simulated data. Clearly, the simulated data exhibits long
Figure 3.1: Histogram and Model Comparison for the Simulated Inverse Gaussian Data tail behavior. Using the maximum likelihood estimate, we fit the truncated normal, truncated skew normal and the non-central skew $t$ distributions to the simulated log normal data.

Figure (3.2) illustrates the histogram of the simulated data and the fitted curves. It shows superior performance of the extended skew $t$ model over others. To illustrates how the extended skew $t$ distribution outperforms the other models in fitting the tail of the data, we zoom in to focus on the tail behavior of the density. Figure (3.3) illustrates the improvement that we gain from using the extended skew $t$ distribution compared to the normal or the skew normal distribution.

Table (3.2) displays the summary of the model comparison study for the log normal data set. AIC and SIC are computed to each one of the three models. Based on the AIC and SIC values, the extended skew $t$ distribution outperforms the skew normal and normal distributions. Note here that the fitted value for the degrees of freedom of the extended skew $t$ distribution is $r = 2$. Also note that the fitted value of the shape parameter of the skew normal function is 155.11. The graphical display of the data and the fitted density illustrate the superior performance of the non-central skew $t$ distribution over the skew normal and the normal. The superior performance of the non-central skew $t$ distribution becomes more visible when we focus on the tail of the data. Figure (3.3) illustrates the same graphs of the
fitted density and the data with the vertical scale enlarged to display the tail behavior of the data. It is clear that the skew $t$ density is the best fit for the data compared to the other candidates if an accurate estimate of the tail probability is needed. Overall, our simulation studies indicate that the non-central skew $t$ distribution is a good candidate in fitting long tail data.

<table>
<thead>
<tr>
<th>Model</th>
<th>$\mu$</th>
<th>$\sigma$</th>
<th>$\alpha$</th>
<th>$r$</th>
<th>AIC</th>
<th>SIC</th>
<th>LogL</th>
</tr>
</thead>
<tbody>
<tr>
<td>$St_r$</td>
<td>0.15</td>
<td>1.02</td>
<td>17.02</td>
<td>1.94</td>
<td>14167.2</td>
<td>14193.27</td>
<td>-7079.60</td>
</tr>
<tr>
<td>Normal</td>
<td>1.65</td>
<td>2.12</td>
<td>NA</td>
<td>NA</td>
<td>21709.3</td>
<td>21722.34</td>
<td>-10852.65</td>
</tr>
<tr>
<td>Skew Normal</td>
<td>0.06</td>
<td>2.65</td>
<td>155.11</td>
<td>NA</td>
<td>17047.34</td>
<td>17066.89</td>
<td>-8520.67</td>
</tr>
</tbody>
</table>

Figure 3.2: Histogram and Model Comparison for the Simulated Log Normal Data
3.3 VOLCANO HEIGHTS DATA

This dataset consists of the heights of 1500 of the world’s volcanoes (Source: Global Volcanism Program, http://www.volcano.si.edu/). The basic descriptive statistics for the volcano heights (unit=100 ft) are summarized in table (3.3). The coefficient of skewness is 0.34 and its coefficient of excess kurtosis is 4.93, which indicates that the data is asymmetric and heavy tailed. The histogram of the data is displayed in figure (3.4). Figure (3.5) illustrates the normal quantile-quantile plot of the data. Clearly, the data is not normally distributed. We use the non-central skew $t$ to fit the data and compare to goodness of fit to the skew normal and normal models. The fitted density curves are illustrated in the in figure (3.6). Comparing the AIC and SIC values of the three models, we see that the extended skew $t$ distribution is the best fit for the data. Note again that the degrees of freedom has a MLE of 3.21. For this data set the performance of the normal distribution and the skew normal distribution is comparable. In fact, the fitted normal and skew normal densities in this example seem to coincide in figure (3.6). This can be predicted from table (3.4) since the MLE of the shape parameter of the fitted skew normal density is close to zero.
Table 3.3: Summary of The Volcano Heights Data

<table>
<thead>
<tr>
<th>Min.</th>
<th>Q1</th>
<th>Median</th>
<th>Q3</th>
<th>Max.</th>
<th>Mean</th>
<th>SD</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>-60.00</td>
<td>7.38</td>
<td>14.88</td>
<td>23.77</td>
<td>68.87</td>
<td>16.88</td>
<td>16.12</td>
<td>0.34</td>
<td>4.93</td>
</tr>
</tbody>
</table>

Figure 3.4: Histogram of the Volcano Elevations Data

Figure 3.5: Normal Q-Q Plots of the Volcano Elevations Data
Table 3.4: Models Fitted to the Volcano Data

<table>
<thead>
<tr>
<th>Model</th>
<th>$\mu$</th>
<th>$\sigma$</th>
<th>$\alpha$</th>
<th>$r$</th>
<th>AIC</th>
<th>SIC</th>
<th>LogL</th>
</tr>
</thead>
<tbody>
<tr>
<td>$St_r$</td>
<td>22.096</td>
<td>12.99</td>
<td>-1.76</td>
<td>3.21</td>
<td>12361.59</td>
<td>12382.84</td>
<td>-6176.795</td>
</tr>
<tr>
<td>Normal</td>
<td>16.88</td>
<td>16.11</td>
<td>NA</td>
<td>NA</td>
<td>12599.68</td>
<td>12610.31</td>
<td>-6297.84</td>
</tr>
<tr>
<td>Skew Normal</td>
<td>17.27</td>
<td>16.12</td>
<td>-0.03</td>
<td>NA</td>
<td>12601.68</td>
<td>12617.62</td>
<td>-6297.84</td>
</tr>
</tbody>
</table>

3.4 FIBER GLASS DATA

In this section we study a data set from Smith and Naylor (1987). It consists of the breaking strengths of 63 glass fibers. Breaking strength or tensile strength is the maximum stress applied to a material before it breaks. Although the tensile strength of materials will be different based upon their compositions, the method and formula used to measure tensile strength is the same. A universal testing machine composed of a loading unit and a control unit is used to test any solid, including fiberglass. The loading unit holds the material in place while force is applied with the control unit. Table (3.5) provides the numerical summary of the data. The histogram of the data is displayed in figure (3.4).

Clearly the data is asymmetric so the normal distribution is not expected to be a good
Table 3.5: Summary of The Fiber Glass Data

<table>
<thead>
<tr>
<th>Min.</th>
<th>Q1</th>
<th>Median</th>
<th>Q3</th>
<th>Max.</th>
<th>Mean</th>
<th>SD</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.55</td>
<td>1.38</td>
<td>1.59</td>
<td>1.68</td>
<td>2.24</td>
<td>1.51</td>
<td>0.32</td>
<td>-0.9</td>
<td>3.92</td>
</tr>
</tbody>
</table>

Figure 3.7: Histogram of the fiber glass data

Figure 3.8: Normal Q-Q Plots of the Fiber Glass Data

candidate to model the data. However, the data does not have long tails. This suggests that the skew normal density might be a good fit for the data. In fact table (3.6) confirms this descriptive analysis. The best fit for the data based on the information criterion is the skew normal distribution followed by the non-central skew $t$ distribution. Note that the MLE of the degrees of freedom for the fitted skew $t$ density is 65.31. Recall that Theorem (2.3.8) of chapter 2 states that for large enough values of the degrees of freedom the non-central skew
Figure 3.9: Fitted density curves to the fiber glass data

The density can be approximated by the skew normal one. Hence based on the fitted value of $r$ the fact that the skew normal is adequate in this case is predictable. Figure (3.9) provides a graphical illustration of the fitted density curves and the histogram of the data. It shows that the density of the fitted skew normal and skew $t$ densities are almost coinciding. Of course in this case the simpler model will win in terms of information criterion. This is our first example of a situation where the skew normal is outperforming the non-central skew $t$ density.

<table>
<thead>
<tr>
<th>Model</th>
<th>$\mu$</th>
<th>$\sigma$</th>
<th>$\alpha$</th>
<th>$r$</th>
<th>AIC</th>
<th>SIC</th>
<th>LogL</th>
</tr>
</thead>
<tbody>
<tr>
<td>$St_t$</td>
<td>1.83</td>
<td>0.44</td>
<td>-4.29</td>
<td>65.31</td>
<td>37.04</td>
<td>43.64</td>
<td>-14.17</td>
</tr>
<tr>
<td>Normal</td>
<td>1.51</td>
<td>0.32</td>
<td>NA</td>
<td>NA</td>
<td>39.82</td>
<td>44.11</td>
<td>-17.91</td>
</tr>
<tr>
<td>Skew Normal</td>
<td>1.85</td>
<td>0.47</td>
<td>-2.68</td>
<td>NA</td>
<td>33.91</td>
<td>40.34</td>
<td>-13.96</td>
</tr>
</tbody>
</table>

### 3.5 ENVIRONMENTAL DATA

In this section we use our model to fit data sets obtained by measuring the concentration of heavy metals in soil. These measurements are obtained during a sampling process that is
intended to measure the pollution of soil in certain locations in the USA. Measurements are
given in milligrams of heavy metal per 1 kilogram of soil. We will study three data sets each
with sample size of 1717. Many thanks to Dr. John Carson, a senior statistician in Shaw
Environmental, Inc." for providing the data sets that will be studied in this section.

![Probability Histogram of the As Data](image)

Figure 3.10: Histogram of the Concentration of As in mg/kg of soil

### 3.5.1 THE CONCENTRATION OF ARSENIC IN SOIL SAMPLES

The first data set is obtained by measuring the concentration of the element Arsenic in soil
samples from selected locations. Figure (3.10) provides a histogram of the concentration of
Arsenic in the soil samples data. Table (3.7) provides a numerical summary of the data.
The excessive skewness and long tailed behavior of this histogram makes using skew normal
distribution not optimal. Figure (3.11) illustrates the normal quantile-quantile plot of the
data. It shows that the normal model is not optimal for the data. We use truncated normal,

---

1Shaw Environmental currently joined forces with CB & I a complete energy infrastructure focused
company and a major provider of government services.
truncated skew normal and the proposed skew $t$ distribution to fit the data set from this distribution. The numerical summary of the fitted models and the information criterion for each of them is given in table 3.8. Figure 3.12 illustrates the fitted density curves superimposed on the histogram of the data. A better illustration of the superior behavior of the non-central skew $t$ density is provided in figure 3.13. Both the skew normal and normal fitted densities fail to predict that the random variable being modeled can exceed 5 units, while the non-central skew $t$ does a better job in predicting observations higher than 5 units.

<table>
<thead>
<tr>
<th>Min.</th>
<th>Q1</th>
<th>Median</th>
<th>Q3</th>
<th>Max.</th>
<th>Mean</th>
<th>SD</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>0.49</td>
<td>0.60</td>
<td>0.82</td>
<td>32.70</td>
<td>0.78</td>
<td>1.17</td>
<td>17.09</td>
<td>385.16</td>
</tr>
</tbody>
</table>

Figure 3.11: Normal Q-Q Plots of the Concentration of As in mg/kg of soil
Table 3.8: Models Fitted to the As Data

<table>
<thead>
<tr>
<th>Model</th>
<th>$\mu$</th>
<th>$\sigma$</th>
<th>$\alpha$</th>
<th>$r$</th>
<th>AIC</th>
<th>SIC</th>
<th>LogL</th>
</tr>
</thead>
<tbody>
<tr>
<td>$St_r$</td>
<td>0.49</td>
<td>0.13</td>
<td>1.78</td>
<td>4.31</td>
<td>306.52</td>
<td>327.77</td>
<td>-149.26</td>
</tr>
<tr>
<td>Normal</td>
<td>0.78</td>
<td>1.17</td>
<td>NA</td>
<td>NA</td>
<td>5421.73</td>
<td>5432.63</td>
<td>-2708.87</td>
</tr>
<tr>
<td>Skew Normal</td>
<td>0.27</td>
<td>1.28</td>
<td>23.38</td>
<td>NA</td>
<td>3403.48</td>
<td>3419.83</td>
<td>-1698.74</td>
</tr>
</tbody>
</table>

Figure 3.12: Fitted density curves to the Arsenic data

Figure 3.13: illustration of the tails of the fitted density curves to the Arsenic data
3.5.2 THE CONCENTRATION OF CADMIUM IN SOIL SAMPLES

Our next data set comes from the analysis of the heavy metal Cadmium (denoted by Cd). Table (3.9) shows the numerical summary of the data. Figure (3.15) illustrates the normal quantile-quantile plot of the data. Again the normal model and the skew normal are not optimal to fit the data. Figure (3.16) illustrates the density curves of the fitted models and the histogram of the original data. To illustrate the tail behavior of the data we magnify the scale and produce the zoomed in graph in figure (3.17). Figure (3.17) shows the superior fit that our distribution produces compared to both the normal and the skew normal. The normal distribution was the worst fit for the data with its tail probability approaching zero before the random variable approaches 5 units. The fitted skew normal probability approaches zero slower, but it overestimate the value of the p.d.f. compared to the original data. This data set provides another typical example where our density is the best choice in fitting the data. Table (3.10) lists the fitted MLE values for each model and the computed values of AIC, SIC and log-likelihood. It shows that the non-central skew t distribution is a much better fit for the data than the other two models.

<table>
<thead>
<tr>
<th>Min.</th>
<th>Q1</th>
<th>Median</th>
<th>Q3</th>
<th>Max.</th>
<th>Mean</th>
<th>SD</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.04</td>
<td>0.17</td>
<td>0.34</td>
<td>0.73</td>
<td>72.80</td>
<td>0.74</td>
<td>2.41</td>
<td>20.85</td>
<td>548.23</td>
</tr>
</tbody>
</table>
Figure 3.14: Histogram of the Concentration of Cd in mg/kg of soil

Figure 3.15: Normal Q-Q Plots of the Concentration of Cd in mg/kg of soil

<table>
<thead>
<tr>
<th>Model</th>
<th>$\mu$</th>
<th>$\sigma$</th>
<th>$\alpha$</th>
<th>$r$</th>
<th>AIC</th>
<th>SIC</th>
<th>LogL</th>
</tr>
</thead>
<tbody>
<tr>
<td>$St_r$</td>
<td>0.038</td>
<td>0.37</td>
<td>21.43</td>
<td>1.71</td>
<td>1704.37</td>
<td>1726.16</td>
<td>-848.18</td>
</tr>
<tr>
<td>Normal</td>
<td>0.74</td>
<td>2.41</td>
<td>NA</td>
<td>NA</td>
<td>7894.29</td>
<td>7905.18</td>
<td>-3945.14</td>
</tr>
<tr>
<td>Skew Normal</td>
<td>0.036</td>
<td>2.51</td>
<td>20</td>
<td>NA</td>
<td>5655.12</td>
<td>5671.46</td>
<td>-2824.12</td>
</tr>
</tbody>
</table>
Fitted Density Curves to the Cd Data

Figure 3.16: Fitted density curves to the Cadmium data

Figure 3.17: Illustration of the tail of the fitted density curves to the Cadmium data
3.5.3 THE CONCENTRATION OF PLUMBUM (LEAD) IN SOIL SAMPLES

The last data set to be discussed in this section is obtained by measuring the concentration of the element Plumbum, commonly known as Lead and denoted by Pb in chemistry. The numerical summary is given in table (3.11). The values of the sample skewness and kurtosis indicate that the normal model will not be an appropriate choice. The data set has tails longer than the normal distribution. Figure (3.19) illustrates the normal quantile-quantile plot of the data, it shows that the data is not normally distributed. The histogram of the data is displayed in figure (3.18). The fitted density curves to the data are displayed in figure (3.20) with the tails of the fitted density illustrated in figure (3.20). The MLE estimates, AIC, and SIC values of each distribution are summarized in table (3.12). As expected, the non-central skew t distribution is the best fit for the data compared to the normal and skew normal models.

Figure 3.18: Histogram of the concentration of Lead in soil samples measured in 10 mg/ kg of soil.
Table 3.11: Summary of The Plumbum Concentration Data

<table>
<thead>
<tr>
<th>Min.</th>
<th>Q1</th>
<th>Median</th>
<th>Q3</th>
<th>Max.</th>
<th>Mean</th>
<th>SD</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>1.67</td>
<td>2.20</td>
<td>3.35</td>
<td>428.00</td>
<td>4.66</td>
<td>16.78</td>
<td>18.87</td>
<td>416.94</td>
</tr>
</tbody>
</table>

Figure 3.19: Normal Q-Q Plots of the Concentration of Pb in mg/kg of soil

Table 3.12: Summary of the Models Fitted to the Pb Data

<table>
<thead>
<tr>
<th>Model</th>
<th>$\mu$</th>
<th>$\sigma$</th>
<th>$\alpha$</th>
<th>$r$</th>
<th>AIC</th>
<th>SIC</th>
<th>LogL</th>
</tr>
</thead>
<tbody>
<tr>
<td>$St_r$</td>
<td>1.20</td>
<td>1.42</td>
<td>6.12</td>
<td>2.34</td>
<td>7047.43</td>
<td>7069.22</td>
<td>-3519.72</td>
</tr>
<tr>
<td>Normal</td>
<td>4.61</td>
<td>16.74</td>
<td>NA</td>
<td>NA</td>
<td>14561.6</td>
<td>14572.52</td>
<td>-7278.32</td>
</tr>
<tr>
<td>Skew Normal</td>
<td>0.46</td>
<td>17.30</td>
<td>82.54</td>
<td>NA</td>
<td>12297.45</td>
<td>12313.8</td>
<td>-6145.228</td>
</tr>
</tbody>
</table>
Figure 3.20: Fitted density curves to the concentration of Lead in soil samples measured in 10 mg/kg of soil.

Figure 3.21: Illustration of the tails of the fitted density curves to the concentration of Lead in soil samples measured in 10 mg/kg of soil.
3.6 CONCLUSIONS AND RECOMMENDATIONS

Based on our simulation studies, the real life data analysis and the model comparisons discussed throughout this chapter we can make the following recommendations about using skew normal or the non-central skew $t$ distributions in modeling data sets.

- Using descriptive analysis is always recommended at the beginning of studying a given data set. The histogram of the data can give insights about whether the data is symmetric or not and in some examples one can infer that the data is long tailed by looking at the histogram. We also recommend computing the sample skewness and sample kurtosis as a more precise indicator that can show whether a normal or skew normal model would be adequate for the data.

- We developed an algorithm for computing the MLE estimates of the parameters for the non-central skew $t$ density that fits the data. This algorithm is easy to use regardless to the size of the data. Hence we recommend starting by fitting a given data set with the $St_r'(\mu, \sigma, \alpha)$ density. The next step is to check if the skew normal would be good enough to model that data. In general, if the fitted value of the degrees of freedom exceeds 30, we recommend computing the MLE estimates for the skew normal density and running model comparison analysis. In all the examples that I worked on, the skew normal density had lower AIC, SIC values compared to the $St_r'(.)$ one when the fitted value of $r$ exceeded 30.

- The special cases of our density as outlined in lemma (2.3.1) and lemma (2.3.2) need to be kept in mind during the model comparison. If the fitted values of the $St_r'(.)$ parameters coincide with one of the special cases then the special case need to be used.
The $St_r(.)$ density significantly outperforms the skew normal density for data with large values of the kurtosis coefficient. Examples are provided in the environmental data section of this chapter. Rare events that are supernormally far from the mean of the data can not be predicted or modeled using the skew normal density. In some applications such rare events might be of extreme importance when it comes to the financial risk associated with them. Hence the $St_r(.)$ distribution is always recommended in such examples as it does a good job in detecting such rare events or outliers.
CHAPTER 4

AN INFORMATION APPROACH FOR THE CHANGE POINT PROBLEM UNDER THE NON-CENTRAL SKEW $t$ DISTRIBUTION

4.1 INTRODUCTION

One of the goals of statistics is to find models that can best describe a given data set. Such models can be useful in understanding the data and predicting future outcomes. In practice, data may follow a certain distribution up to a certain point until a dramatic change occurs due to some extraneous factors. After a certain point the data starts to follow a different distribution. A point after which the change in the distribution of the data occurs is commonly referred to as a change point. The change point problem is the problem of detecting the number of change points and estimating their locations. The following setting for the
Let \( x_1, x_2, \ldots, x_n \) be a sequence of independent random variables with probability distribution functions \( F_1, F_2, \ldots, F_n \), respectively. Then in general, the change point problem is to test the following null hypothesis,

\[
H_0 : F_1 = F_2 = \cdots = F_n
\]  

(4.1)

versus the alternative:

\[
H_1 : F_1 = \cdots = F_{k_1} \neq F_{k_1+1} = \cdots = F_{k_2} \neq F_{k_2+1} = \cdots = F_{k_q} \neq F_{k_q+1} = \cdots = F_n,
\]  

(4.2)

where \( 1 < k_1 < k_2 < \cdots < k_q < n \), \( q \) is the unknown number of change points and \( k_1, k_2, \cdots, k_q \) are the respective unknown positions that have to be estimated.

If the distributions \( F_1, F_2, \cdots, F_n \) belong to a common parametric family \( F(\theta) \), where \( \theta \in \mathbb{R}^p \), then the change point problem is to test the null hypothesis about the population parameters \( \theta_i, i = 1, \cdots, n \):

\[
H_0 : \theta_1 = \theta_2 = \cdots = \theta_n = \theta \text{ (unknown)},
\]  

(4.3)

versus the alternative hypothesis:

\[
H_1 : \theta_1 = \cdots = \theta_{k_1} \neq \theta_{k_1+1} = \cdots = \theta_{k_2} \neq \theta_{k_2+1} = \cdots = \theta_{k_q-1} \neq \theta_{k_q} = \cdots = \theta_n,
\]  

(4.4)

where \( q \) and \( k_1, k_2, \cdots, k_q \) have to be estimated. These hypotheses together reveal the aspects of change point inference: determining if any change point exists in the process and estimating the number and locations of change points. In fact, the multiple change point
problem can be treated as a single change point problem by using the binary segmentation method that was proposed by Vostrikova (1981).

A special multiple change points problem is the epidemic change point problem, which is defined by testing the following null hypothesis,

\[ H_0 : \theta_1 = \theta_2 = \cdots = \theta_n = \theta \text{ (unknown)} \quad (4.5) \]

versus the alternative:

\[ H_1 : \theta_1 = \cdots = \theta_k = \alpha \neq \theta_{k+1} = \cdots = \theta_t = \beta \neq \theta_{t+1} = \cdots = \theta_n = \alpha, \quad (4.6) \]

where \( 1 \leq k < t \leq n \), and \( \alpha \) and \( \beta \) are unknown.

\section*{4.2 LITERATURE REVIEW OF THE CHANGE POINT PROBLEM}

Change point problems have interested statisticians since 1975. The frequently used methods for change point inference in the literature are the maximum likelihood ratio test, Bayesian test, nonparametric test, stochastic process, information theoretic approach and semi-parametric test. The change point problem for a sequence of normal random variables was studied thoroughly in the literature. Most of the literature on the change point problem assumes a single change point in the data. The binary segmentation method can be used to detect multiple change points in a data set.

A comprehensive literature review of the change point problem is provided in the book by Chen and Gupta (2012). Sen and Srivastava (1975a) derived the exact and asymptotic

4.3 METHODOLOGY

In this section, the Schwartz information criterion will be used to detect the location of the change point for a given data set under the assumption that the data follows the non-central skew $t$ distribution. Akaike (1973) proposed the following information criterion,

$$AIC_t = -2\log L(\hat{\theta}_t) + 2t, \quad t = 1, 2, ..., K. \quad (4.7)$$
where \( L(\hat{\theta}_t) \) is the maximum likelihood for model with \( t \) parameters, as a measure of model evaluation. A model that minimizes the AIC is called Minimum AIC estimate, (abbreviated as MAICE) is considered to be the most appropriate model. However, the MAICE is not an asymptotically consistent estimator of model order (see Schwarz, 1978). Some authors made efforts to modify the information criterion without violating Akaike’s original principles. One of the modifications is the Schwarz’s Information Criterion, denoted as SIC, and proposed by Schwarz in 1978. It is expressed as

\[
SIC_t = -2\log L(\hat{\theta}_t) + t\log(n),
\]

where \( t \) is the number of parameters to be estimated and \( n \) is the sample size. Assuming that the change point of the data occurs in location \( k \), we can compute the following information criteria for the model

\[
SIC_t(k) = -2\log L(\hat{\theta}_t) + t\log(n), \quad k = 2, ..., n-2, \quad t = 1, 2, \cdots, K.
\]

where \( k \) indicates the location of the change point and \( t \) is the total number of parameters in the two models.

For computational purposes, we need enough observations to obtain a sufficient estimator for the parameters of the model. Therefore, we compute \( SIC(k) \) for \( k_0 \leq k \leq n-k_0 \), where in practice we choose \( k_0 \) large enough so that the MLE can be computed accurately. Clearly, the \( k_0 \) will depend on the sample size \( n \) and the number of parameters to be estimated. According to the information criteria principle, the position of the change point \( k \) can be estimated by \( \hat{k} \) such that \( SIC(\hat{k}) \) is the minimal. In other words, under the null hypothesis \( H_0 \), \( SIC(n) \) is computed under the assumption of no change in the distribution that the data comes from. Corresponding to the \( H_1 \) there are \( k \) possible change points and each of them has a corresponding SIC, denoted by \( SIC(k) \) for \( k_0 \leq k \leq n-k_0 \). According to the
information criterion principle, we accept $H_0$ if

$$SIC(n) < \min_{k_0 \leq k \leq n-k_0} SIC(k),$$  \hspace{1cm} (4.10)

and accept $H_1$ if

$$SIC(n) > SIC(k), \quad \text{for some } k_0 \leq k \leq n - k_0.$$  \hspace{1cm} (4.11)

The position of the change point can be estimated by $\hat{k}$ such that

$$SIC(\hat{k}) = \min_{k_0 \leq k \leq n-k_0} SIC(k).$$  \hspace{1cm} (4.12)

4.4 SIMULATION STUDY

In this section, we investigate change point problems of the non-central skew $t$ distribution under different scenarios. In each case, the problem is defined as a hypothesis testing problem. A testing procedure based on $SIC$ is proposed to detect and estimate the location of the change points if there is any. To study the effectiveness of the method, simulations of the power are conducted to illustrate the performance of the proposed method for various values of the parameters and the sample size.

4.4.1 THE CHANGE OCCURS IN THE SHAPE PARAMETER

In this section, we will study the change point problem for the shape parameter while the location, scale and degrees of freedom are unknown but fixed. During this discussion, we assume that there is at most one change point since the multiple change point problem can be transformed into a single change point problem with the binary segmentation method. That is, we are testing the null hypothesis:

$$H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_n = \alpha \text{ (unknown)},$$  \hspace{1cm} (4.13)
versus the alternative:

\[ H_1 : \alpha_1 = \cdots = \alpha_k = \beta \neq \alpha_{k+1} = \cdots = \alpha_n = \gamma, \tag{4.14} \]

where \( k_0 \leq k < n - k_0 \), and \( \beta \) and \( \gamma \) are unknown.

Under the null hypothesis, \( H_0 \), the \( SIC \) is given by:

\[ SIC_4(n) = -2 \sum_{i=1}^{n} \log(f_{St_{r_0}'}(x_i; \mu_0, \sigma_0, \alpha)) + 4 \log(n), \tag{4.15} \]

where \((\mu_0, \sigma_0, \alpha, r_0)\) is the vector of MLE fitted to the whole data set, \( f_{St_{r_0}'} \) denotes the p.d.f. of the non-central skew \( t \) distribution with \( r_0 \) degrees of freedom and \( x_i \) is the \( i^{th} \) observed value.

Under the alternative \( H_1 \), the \( SIC \) is given by:

\[ SIC_5(k) = -2 \sum_{i=1}^{k} \log(f_{St_{r_0}'}(x_i; \mu_0, \sigma_0, \beta)) - 2 \sum_{i=k+1}^{n} \log(f_{St_{r_0}'}(x_i; \mu_0, \sigma_0, \gamma)) + 5 \log(n), \tag{4.16} \]

where \( k \) is the location of the change point, \((\mu_0, \sigma_0, \beta, r_0)\) is the vector of MLE of the parameters fitted to the first segment of the data and \((\mu_0, \sigma_0, \gamma, r_0)\) is the MLE vector fitted to the second segment.

We conduct the simulations to study the performance of the change point detecting procedure based on \( MSICE \). In the simulation study, we generate the data under the non-central skew \( t \) distribution with various change point locations of the shape parameter and various sample sizes while the other parameters are fixed. The procedure successfully detects the location of the change point if \( H_0 \) is rejected and the change point \( \hat{k} \) is within \([k - 1, k + 1]\), where \( k \) is the true value of the change point. The power of the proposed test will be calculated as the percentage of successful detection among the total number of simulations.
To illustrate the use of the MSICE based change point and study its power, we run a series of simulations studies. The algorithm proposed in Chapter 2 to generate a random sample from the $St_r$ density is used. A sample of size $n$ is chosen. The sample is designed to have one change point at a predetermined location $k$. We generate a sample of size $k$ from the $St_r(\mu, \sigma, \alpha_1)$ family and then we generate a random sample of size $n - k$ from the $St_r(\mu, \sigma, \alpha_2)$ family, where $\alpha_1 \neq \alpha_2$. Then we compute the SIC of the resulting sample assuming that the location, scale and degrees of freedom are fixed and the shape parameter $\alpha$ is unknown and possibly changing. Values of SIC are computed assuming the location of the change point varies from 2 to $n - 2$. The values of the resulting SIC are stored in a vector and the location of the minimum value corresponds to the estimated change point $\hat{k}$. Finally, the probability $P(k-1 \leq \hat{k} \leq k+1)$ is computed to quantify the power of the method.

Table (4.1) summarizes the power of the MSICE method when the sample size $n = 40$, the parameters $\mu = 0, \sigma = 1, r = 3$ are fixed and the shape parameter $\alpha$ changes from 1 to -1. The computations of the power are based on 1000 replicates of the experiment. Table (4.2) summarizes the power under the same assumptions but now with the larger sample size of $n = 60$. Comparing the powers at the common change points between the two scenarios, we notice that the power of the method improves as the sample size increases.

Table (4.3) summarizes the power of the MSICE method when the sample size $n = 40$ and $\mu = 0, \sigma = 1, r = 3$ are fixed while $\alpha$ changed from 5 to -1. These are the same assumptions used to generate table (4.1) except here that the difference in the shape parameters $\alpha_1, \alpha_2$ has increased. Notice that the power at each change point is higher than the power at the corresponding point in table (4.1). This can be explained by the higher discrepancy in the densities under the two models used to generate the samples in the latter case. Figure (4.1) illustrates the density curves of the densities included in this section. Notice that the values of $t$ that are in the intersection of the supports of the two densities summarized in
table (4.1) have higher probabilities in general compared to the overlapping values for the
densities in table (4.3). This explains the higher power in the latter case. In fact, simulations
under the same assumptions with $\alpha$ changing from 1 to 5 produce the lowest power.

Table 4.1: Power of the MSICE change point detection as the shape parameter $\alpha$ changed
from 1 to -1, $\mu = 0, \sigma = 1, r = 3$ and sample size $n = 40$

<table>
<thead>
<tr>
<th>Location of the change point $k$</th>
<th>Power $= P(k - 1 \leq \hat{k} \leq k + 1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.550</td>
</tr>
<tr>
<td>10</td>
<td>0.584</td>
</tr>
<tr>
<td>20</td>
<td>0.584</td>
</tr>
<tr>
<td>30</td>
<td>0.606</td>
</tr>
<tr>
<td>35</td>
<td>0.594</td>
</tr>
</tbody>
</table>

Table 4.2: Power of the MSICE change point detection as the shape parameter $\alpha$ changed
from 1 to -1, $\mu = 0, \sigma = 1, r = 3$ and sample size $n = 60$

<table>
<thead>
<tr>
<th>Location of the change point $k$</th>
<th>Power $= P(k - 1 \leq \hat{k} \leq k + 1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.570</td>
</tr>
<tr>
<td>10</td>
<td>0.594</td>
</tr>
<tr>
<td>20</td>
<td>0.602</td>
</tr>
<tr>
<td>30</td>
<td>0.618</td>
</tr>
<tr>
<td>40</td>
<td>0.620</td>
</tr>
<tr>
<td>50</td>
<td>0.630</td>
</tr>
<tr>
<td>55</td>
<td>0.554</td>
</tr>
</tbody>
</table>
Table 4.3: Power of the MSICE change point detection as the shape parameter $\alpha$ changed from 5 to -1, $\mu = 0, \sigma = 1, r = 3$ and sample size $n = 40$

<table>
<thead>
<tr>
<th>Location of the change point $k$</th>
<th>Power $= P(k - 1 \leq \hat{k} \leq k + 1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.816</td>
</tr>
<tr>
<td>10</td>
<td>0.878</td>
</tr>
<tr>
<td>20</td>
<td>0.862</td>
</tr>
<tr>
<td>30</td>
<td>0.852</td>
</tr>
</tbody>
</table>

Figure 4.1: Graphs of the densities included in the power comparisons of the MSICE of change point as the shape parameter changes: The p.d.f. of $St'_3(0,1,\alpha)$ for $\alpha = 1, -1, -5$ are displayed.

4.4.2 THE CHANGE OCCURS IN THE LOCATION PARAMETER ONLY

In this section, we study the change point problem of the non-central skew $t$ distribution under the assumption that the location parameter $\mu$ is changing, while the scale, shape and degrees of freedom are known and fixed. We also assume that the data has exactly one
change point. In other words, we are testing the hypothesis:

\[ H_0 : \mu_1 = \mu_2 = \cdots = \mu_n = \mu \text{ (unknown)} \quad (4.17) \]

versus the alternative:

\[ H_1 : \mu_1 = \cdots = \mu_k = \beta \neq \mu_{k+1} = \cdots \mu_n = \gamma, \quad (4.18) \]

where \( 2 \leq k \leq n - 2 \), and \( \beta \) and \( \gamma \) are unknown.

Under the null hypothesis, \( H_0 \), the SIC is given by:

\[ SIC_4(n) = -2 \sum_{i=1}^{n} \log(f_{St_{r_0}}(x_i; \mu, \sigma_0, \alpha_0)) + 4 \log(n), \quad (4.19) \]

where \((\mu, \sigma_0, \alpha_0, r_0)\) is the vector of MLE fitted to the whole data set, \(f_{St_{r_0}}\) denotes the p.d.f. of the non-central skew t distribution with \(r_0\) degrees of freedom and \(x_i\) is the \(i^{th}\) observed value.

Under the alternative \( H_1 \), the SIC is given by:

\[ SIC_5(k) = -2 \sum_{i=1}^{k} \log(f_{St_{r_0}}(x_i; \beta, \sigma_0, \alpha_0)) - 2 \sum_{i=k+1}^{n} \log(f_{St_{r_0}}(x_i; \gamma, \sigma_0, \alpha_0)) + 5 \log(n), \quad (4.20) \]

where \(k\) is the location of the change point, \((\beta, \sigma_0, \alpha_0, r_0)\) is the vector of MLE of the parameters fitted to the first segment of the data and \((\gamma, \sigma_0, \alpha_0, r_0)\) is the MLE vector fitted to the second segment.

The method outlined in the previous section is used here to estimate the power of the MSICE change point detection as the location parameter \(\mu\) changed while the remaining parameters are assumed fixed. Table (4.4) summarizes the power as \(\mu\) changes from 0 to 2 ,
while $\sigma = 1, \alpha = 1, r = 3$ and the sample size $n = 40$. The first observation here is that the power is higher than the similar case that was involving a change in the shape parameter. This indicates that the MSICE change point detection method is more sensitive to the change in location than the change in shape of the $S_{\mu}^r$ random samples.

Table (4.5) summarizes the power as $\mu$ changes from 0 to -2, while $\sigma = 1, \alpha = 1, r = 3$ and sample size $n = 40$. Comparing the powers computed at each change point here to the corresponding powers in table (4.4), we see an improvement in table (4.5). Figure (4.2) illustrated the three density curves under study. It is clear from the probabilities of the points in the overlapping supports or every pair of density curves, that it is easier to detect the change point as $\mu$ changes from 0 to -2 compared to the case where $\mu$ changes from 0 to 2. Table (4.6) summarizes the power under the same parameter assumptions in table (4.5) but with a sample size of 60 instead of 40. We see that the values of the power in the last two tables are comparable.

Table 4.4: Power of the MSICE change point detection as the location parameter $\mu$ changed from 0 to 2, $\sigma = 1, \alpha = 1, r = 3$ and sample size $n = 40$

<table>
<thead>
<tr>
<th>Location of the change point $k$</th>
<th>Power $= P(k - 1 \leq \hat{k} \leq k + 1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.89</td>
</tr>
<tr>
<td>10</td>
<td>0.89</td>
</tr>
<tr>
<td>20</td>
<td>0.80</td>
</tr>
<tr>
<td>30</td>
<td>0.79</td>
</tr>
<tr>
<td>35</td>
<td>0.83</td>
</tr>
</tbody>
</table>
Table 4.5: Power of the MSICE change point detection as the location parameter $\mu$ changed from 0 to -2, $\sigma = 1$, $\alpha = 1$, $r = 3$ and sample size $n = 40$

<table>
<thead>
<tr>
<th>Location of the change point $k$</th>
<th>Power $= P(k - 1 \leq \hat{k} \leq k + 1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.908</td>
</tr>
<tr>
<td>20</td>
<td>0.888</td>
</tr>
<tr>
<td>30</td>
<td>0.912</td>
</tr>
<tr>
<td>35</td>
<td>0.862</td>
</tr>
</tbody>
</table>

Figure 4.2: Graphs of the densities included in the power comparisons of the MSICE of change point as the location parameter changes: The p.d.f. of $St'_{3}(\mu, 1, 1)$ for $\mu = 0, 2, -2$ are displayed.

Table 4.6: Power of the MSICE change point detection as the location parameter $\mu$ changed from 0 to -2, $\sigma = 1$, $\alpha = 1$, $r = 3$ and sample size $n = 60$

<table>
<thead>
<tr>
<th>Location of the change point $k$</th>
<th>Power $= P(k - 1 \leq \hat{k} \leq k + 1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.904</td>
</tr>
<tr>
<td>20</td>
<td>0.900</td>
</tr>
<tr>
<td>30</td>
<td>0.900</td>
</tr>
<tr>
<td>40</td>
<td>0.912</td>
</tr>
<tr>
<td>50</td>
<td>0.856</td>
</tr>
</tbody>
</table>
4.4.3 *THE LOCATION AND SCALE PARAMETERS CHANGE SIMULTANEOUSLY*

In this subsection, we study the change point problem of the non-central skew $t$ distribution under the assumption that both the location parameter $\mu$ and the scale parameter $\sigma$ are changing at the same point. The shape parameter $\alpha$ and the degrees of freedom $r$ are known and fixed. We also assume that the data has exactly one change point. In other words, we are testing the hypothesis:

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_n = \mu$$
and $$\sigma_1 = \sigma_2 = \cdots = \sigma_n = \sigma$$ \hspace{1cm} (4.21)

where $\mu, \sigma$ are known.

versus the alternative:

$$H_1: \mu_1 = \cdots = \mu_k = \beta_1 \neq \mu_{k+1} = \cdots = \mu_n = \beta_2,$$
and $$\sigma_1 = \cdots = \sigma_k = \gamma_1 \neq \sigma_{k+1} = \cdots = \sigma_n = \gamma_2,$$ \hspace{1cm} (4.22)

where $2 \leq k \leq n - 2$, and $\beta_1, \beta_2$ and $\gamma_1, \gamma_2$ are unknown.

Under the null hypothesis, $H_0$, the $SIC$ is given by:

$$SIC_4(n) = -2 \sum_{i=1}^{n} \log(f_{St'_r}(x_i; \mu, \sigma, \alpha)) + 4 \log(n),$$ \hspace{1cm} (4.23)

where $(\mu, \sigma, \alpha, r)$ is the vector of $MLE$ fitted to the whole data set, $f_{St'_r}$ denotes the $p.d.f.$ of the non-central skew $t$ distribution with $r$ degrees of freedom and $x_i$ is the $i^{th}$ observed value.

Under the alternative $H_1$, the $SIC$ is given by:

$$SIC_6(k) = -2 \sum_{i=1}^{k} \log(f_{St'_r}(x_i; \beta_1, \gamma_1, \alpha)) - 2 \sum_{i=k+1}^{n} \log(f_{St'_r}(x_i; \beta_2, \gamma_2, \alpha)) + 6 \log(n),$$ \hspace{1cm} (4.24)
Table (4.7) summarizes the power as \( \mu \) changes from 0 to -2, and \( \sigma \) changes from 1 to 2 at the same change point \( k \). The other parameters are kept fixed at \( \alpha = 1, r = 3 \) and the sample size \( n = 40 \). The first observation here is that the power is lower than the similar case that was involving a change in the shape parameter. This is due to the fact that there are two parameters to estimate at each possible change point location, which requires more computing time and higher sample size of the \( St'_r \) random samples. This can be seen from the improvement in the power as the sample size increased to \( n = 60 \) in table (4.8).

Table 4.7: Power of the MSICE change point detection as the location parameter \( \mu \) changed from 0 to -2, \( \sigma \) changes from 1 to 2, and \( \alpha = 1, r = 3 \) and sample size \( n = 40 \)

<table>
<thead>
<tr>
<th>Location of the change point ( k )</th>
<th>Power = ( P(k - 1 \leq \hat{k} \leq k + 1) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.536</td>
</tr>
<tr>
<td>20</td>
<td>0.594</td>
</tr>
<tr>
<td>30</td>
<td>0.606</td>
</tr>
</tbody>
</table>

Table 4.8: Power of the MSICE change point detection as the location parameter \( \mu \) changed from 0 to -2, \( \sigma \) changes from 1 to 2, and \( \alpha = 1, r = 3 \) and sample size \( n = 60 \)

<table>
<thead>
<tr>
<th>Location of the change point ( k )</th>
<th>Power = ( P(k - 1 \leq \hat{k} \leq k + 1) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.574</td>
</tr>
<tr>
<td>20</td>
<td>0.616</td>
</tr>
<tr>
<td>30</td>
<td>0.632</td>
</tr>
<tr>
<td>40</td>
<td>0.616</td>
</tr>
<tr>
<td>50</td>
<td>0.558</td>
</tr>
</tbody>
</table>
Distributions of emerging market stock returns often exhibit skewness and tails that are lighter or longer than the normal distribution. Emerging markets are known to be more susceptible to the political scenario than developed markets, thus their indices tend to have more outliers and structural changes. In this section we apply the MSICE change point detection method to three data sets. The data is obtained from “The MSCI Global Equity indices”. These indices are widely tracked global equity benchmarks and serve as the basis for over 500 exchanged traded funds throughout the world. The data files contain weekly prices over a few years for three countries in Latin America. The goal is to use the data set to detect change point that can possibly be used to detect a change in the overall economy of the country. I am grateful to Loschi, R. H., Arellano-Valle, R. B. and Castro, L. M. for providing the data that has been used in this section.

4.5.1 CHILE’S WEEKLY RETURNS DATA

The data set lists the weekly exchange prices are used to compute weekly returns from Chile starting on the November, 3, 1995 to November, 3, 2000. The weekly prices denoted by $P_t$ are typically dependent. To obtain an independent set of random variables we compute the rate of weekly returns using the following transformation:

$$R_t = \frac{P_{t+1} - P_t}{P_t}, \quad t = 1, 2, \cdots, 261.$$  \hspace{1cm} (4.25)  

The reader is referred to Hsu (1979) for more details on this transformation. The histogram of the rates of weekly returns for Chile is given in figure 4.3. The numerical summary of the data is given in table 4.9. Based on the computed values of the sample skewness and sample kurtosis of the data, we see that the normal model is not the best op-
tions to study this data. To detect the location of the change point we start by standardizing the data so that the mean = 0 and scale = 1. Then we use the standard version of the $St_r'$ model to fit the data. We assume that the location and scale are fixed at 0, 1 respectively and that the shape parameter is changing. Figure (4.5) shows the probability histogram of the standardized and transformed data set. The problem reduces to the testing the hypothesis outlined in equation (4.13) versus the alternative outlined in equation (4.14). We use a sequential approach that assumes one change point at most at each step and keeps testing until no further change points are detected. Figure (4.4) shows that the normal distribution is not an appropriate model for the data.

![Histogram for Chile's Weekly Return Rates](image)

**Figure 4.3:** Histogram for Chile’s Weekly Return Rates.

<table>
<thead>
<tr>
<th>Min.</th>
<th>Q1</th>
<th>Median</th>
<th>Q3</th>
<th>Max.</th>
<th>Mean</th>
<th>SD</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.14010</td>
<td>-0.01984</td>
<td>-0.00135</td>
<td>0.01823</td>
<td>0.11650</td>
<td>-0.00074</td>
<td>0.03356</td>
<td>-0.16606</td>
<td>5.29136</td>
</tr>
</tbody>
</table>
Figure 4.4: Normal Q-Q Plots for Chile’s Weekly Return Rates.

Figure 4.5: Probability Histogram of the Standardized Rates of Chile’s Weekly Returns from 11/3/1995 to 11/3/2000.
Figure 4.6: SIC(k) Values for Chile’s Standardized Return Rates at each Location with a Vertical Line at the First Detected Change Point $\hat{k} = 149$.

Figure 4.7: SIC(k) Values for Chile’s Standardized Return Rates for the First Segment of the Data.
SIC(k) Values for the Second Segment of Chile’s Weekly Returns Data

Figure 4.8: SIC(k) Values for Chile’s Standardized Return Rates for the Second Segment of the Data.

Figure (4.6) shows the computed value of SIC assuming the change point ranges from 2 to \( n - 2 \). Based on the SIC values computed assuming a change in the skewness parameter under the \( St' \) model, we detect the change point to be \( \hat{k} = 149 \). This corresponds to the date September, 4, 1998. Next, we consider the time segments before and after the computed value of the change point \( \hat{k} = 149 \). Figure (4.7) shows the computed values of the SIC for the first segment with a vertical line that marks the detected change point in the segment. Figure (4.8) shows the \( SIC(k) \) values over the second segment. Using the binary segmentation method, the following additional change points are detected: \( \hat{k}_1 = 141 \), and \( \hat{k}_2 = 238 \). These points correspond to the dates July, 10, 1998 and May, 19, 2000. Figure (4.9) displays the weekly return data with vertical lines that mark the three detected locations of the change points.

One needs to take a look at the history of the Chilean economy to provide an explanation to the computed change points. During the early 1990s, Chile’s had a significant
improvement in its economy with the growth in real GDP averaged 8 percent from 1991-1997, but fell to half that level in 1998 because of tight monetary policies and because of lower export earnings that was a product of the Asian financial crisis. Chile’s economy has since recovered over the past several years. After a decade of impressive growth rates, Chile began to experience a moderate economic downturn in 1999, due to global economic conditions related to the Asian financial crisis in 1997 and the 1998 Russian financial crisis. So the detected change points can be justified by the tight monetary policies which were implemented in 1998 and the global economical downturn that occurred in 1999 in connection with the Asian and Russian financial crises. In March, 10, 2000 there was a collapse of a technology bubble that might have contributed to the change point we detected in May, 19, 2000.

Figure 4.9: Data Points of the Weekly Returns of Chile from 11/3/1995 to 11/3/2000 with three vertical lines that mark the detected change points.
4.5.2 BRAZIL'S WEEKLY RETURNS DATA

In this section we analyze the change points of the Brazilian weekly returns. The data set lists the weekly exchange prices are used to compute weekly returns from Brazil starting on the November, 3, 1995 to November, 3, 2000. To analyze the data, we apply the transformation in equation (4.25) to compute the weekly return rates. The transformation gives a data set of size 262. The histogram of the rates of weekly returns for the Brazilian market is given in figure (4.10). The numerical summary of the data is given in table (4.10). The normal model is inappropriate to fit this data as shown by the quantile plot in figure (4.11).

![Histogram of the Standardized Weekly Returns of Brazil](image)

Figure 4.10: Histogram of the standardized weekly returns of the Brazilian market from 11/3/1995 to 11/3/2000.

<table>
<thead>
<tr>
<th>Min.</th>
<th>Q1</th>
<th>Median</th>
<th>Q3</th>
<th>Max.</th>
<th>Mean</th>
<th>SD</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.20450</td>
<td>-0.02346</td>
<td>0.00275</td>
<td>0.03143</td>
<td>0.21240</td>
<td>0.00423</td>
<td>0.05282</td>
<td>-0.20695</td>
<td>5.29846</td>
</tr>
</tbody>
</table>
Figure 4.11: Normal Q-Q Plots for Brazil’s Weekly Return Rates.

The MSICE algorithm with binary segmentation detects two change points in the data set. The values of SIC of the complete data set are displayed in figure 4.12. It is clear from the graph that there are two change points in the data set. The detected points are at weeks number 87 and 253. These correspond to the dates June, 27, 1997 and September, 1, 2000. Figure 4.13 displays the weekly return data with vertical lines that mark the detected locations of the change points. The first detected change point coincides with the 1997 Asian financial crisis and the second one coincides with the collapse of technology bubble, which started in March 2000 and affected the world’s economy for a three year period.

4.5.3 ARGENTINA’S WEEKLY RETURNS DATA

In this section, we study the weekly returns of Argentina’s market. The data set lists the weekly exchange prices are used to compute weekly returns from Argentina starting on the November, 3, 1995 to January, 21, 2000. To analyze the data, we apply the transformation
Figure 4.12: SIC Values for the the Brazilian Returns at Each Week with Vertical Lines to Mark the Detected Change Point.

Figure 4.13: Data Points of the Weekly Returns of Brazil from 11/3/1995 to 11/3/2000 with a Vertical Line that Marks the Detected Change Point.
in equation (4.25) to compute the weekly returns. The histogram of the weekly return data for the Argentina’s market is given in figure (4.14). The sample size of the transformed data is 221. The numerical summary of the data is given in table (4.11). The normal quantile plot is illustrated in figure (4.15). Clearly, the normal model is inappropriate for this data set.


<table>
<thead>
<tr>
<th>Weekly Returns of the Argentina's Market</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weekly Return Rates</td>
</tr>
<tr>
<td>Frequency</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>20</td>
</tr>
<tr>
<td>30</td>
</tr>
<tr>
<td>40</td>
</tr>
<tr>
<td>50</td>
</tr>
<tr>
<td>60</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Weekly Return Rates</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.15 -0.10 -0.05 0.00 0.05 0.10 0.15 0.20</td>
</tr>
</tbody>
</table>

Table 4.11: Summary of Argentina’s Weekly Returns Data

<table>
<thead>
<tr>
<th>Min.</th>
<th>Q1</th>
<th>Median</th>
<th>Q3</th>
<th>Max.</th>
<th>Mean</th>
<th>SD</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.15330</td>
<td>-0.02148</td>
<td>0.00554</td>
<td>0.02561</td>
<td>0.18110</td>
<td>0.00369</td>
<td>0.04575</td>
<td>-0.02599</td>
<td>5.13771</td>
</tr>
</tbody>
</table>

The MSICE algorithm detects two change points in the data set. The values of SIC of the complete data set are displayed in figure (4.16). It is clear from the graph that the first
change point occurs on the ninth week. Running the algorithm again for the segment from week 1 to week 9, we find that there are no further change points. Analysis of the segment from week 9 to week 263, we detect another change point at week 17. These correspond to the dates December, 29, 1995 and February, 23, 1996. Figure 4.17 displays the weekly return rates with vertical lines that mark the detected locations of the change points. Looking at the history of the Argentinean economy we see that the during the early nineties, the economy was blossoming due to trade liberalization and privatization. But the Argentinean economy had several external shocks and witnessed failures of the system that caused it to crumble slowly from 1995 until the collapse in 2001.
Figure 4.16: SIC Values for the Argentina’s Returns at Each Week with Vertical Lines to Mark the Detected Change Point.

Figure 4.17: Data Points of the Weekly Returns of Argentina from 11/3/1995 to 11/3/2000 with a Vertical Line that Marks the Detected Change Point.
4.6 CONCLUDING REMARKS

The change point problem for the non-central skew $t$ distribution can be discussed under several settings. The information based approach for detecting the change point is a good option that has high power. One additional desirable aspect of the $MSICE$ approach to detecting the location of the change point is that it is easy to implement and the algorithm converges relatively fast. The non-central skew $t$ distribution is ideal for modeling data with excess kurtosis or longer tails than the normal model. Hence, the change point of this distribution is very useful in applications that produce data that has such properties.

Detecting the location of the change point is easier if only one parameter is changing over time. As the number of parameters that are changing simultaneously increases, the computations get more complicated, the convergence of the algorithm gets slower and the power decreases. Finally, one can use the algorithm under the non-central skew $t$ model to detect the change point in skew normal data. This is due to the fact that the skew normal model is a special case of our model and they coincide for larger values of the degrees of freedom. Of course the computations under the skew normal model are easier because it requires 3 parameters to be estimated in each iteration rather than 4. In practice one can start by computing the MLE for the whole data. If the value of the fitted degrees of freedom $r$ exceeds 40, then the skew normal might be sufficient to detect the change point of the data.
APPENDIX: SAMPLE OF PROGRAMMING CODE USING R

In this section some of the programming codes that have been used in the data analysis parts of the dissertation are outlined. The functions are written using the software R. The main functions and procedures are outlined in this section. Specific details are left out.

- The following function is used to compute the *p.d.f.* of our skew *t* distribution:

  pdf.t = function(t, mu,sigma, alpha,r)
  {
    g = function (u){
      w = (t*sqrt(u/r)-mu)/sigma
      return(dchisq(u,df = r)* dnorm(w)*pnorm(alpha*w)*sqrt(u/r))
    }
    I = integrate(g, lower = 0, upper = Inf, stop.on.error = TRUE)$value
    c = 2/sigma
    return(c*I)
  }

- This function is used to compute the log likelihood of the density of the skew *t* distribution:

  LL1 = function (x )
\{
\text{mu = x}[1] ; \text{sigma = x}[2] ; \text{alpha=x}[3] ; \text{r=x}[4] \\
\text{sum = 0} \\
\text{for (i in 1:m)}
\{ \\
\text{sum = sum + log (pdf.t (data2[i], mu, sigma, alpha, r ))} \\
\text{return (- sum)}
\}\}

- The following part calls a built-in routine to maximize the likelihood function by minimizing its negative value:
  \text{fit1= optim (c(0, 0.3, 6, 2), LL1)}
  \text{lambda.est = fit1$\text{value}}

- The following part is used to plot the fitted density curve to the given data:
  \text{par = fit1$\text{par}}
  \text{mu =par[1]; sigma= par[2] ; alpha=par[3]; r=par[4]} \\
  \text{x = seq(a, b, by= 0.05)}
  \text{n = length (x)}
  \text{y = numeric (n)}
  \text{for (i in 1:n)}
  \text{y[i] = pdf.t (x[i], mu, sigma, alpha,r)}
  \text{lines (x, y, main ="", type="l", lwd= 2, col="red")}

- The following function computes the log likelihood of the normal density:
  \text{LL.norm = function (x )}
  \{ \text{mu = x}[1] ; \text{sigma = x}[2];
sum = 0
for (i in 1:m){
    sum = sum + dnorm(data2[i], mean = mu, sd = sigma, log = TRUE)}
return(- sum)
}

• The following part computes the MLE using the truncated normal density:
  
  fit.norm= optim(c(1.5, 0.3), LL.norm)
  par = fit.norm$par
  mu = par[1]; sigma = par[2] ; 
  c = 1/(1-pnorm(0, mean=mu , sd= sigma, log = FALSE))
  lambda.norm = log( c) + fit.norm$value

• The following function computes the log likelihood using skew normal density:
  
  library (sn)
  LL.sn = function (x )
  {
    mu = x[1] ; sigma = x[2]; alpha = x[3]
    sum = 0
    for (i in 1:m){
      sum = sum + dsn(data2[i], location = mu, scale = sigma, shape = alpha, dp = NULL, log = TRUE)}
    return(- sum)
  }

• The following part plots of the fitted density curve of the truncated skew normal dis-
distribution:
par = fit.sn$par
mu = par[1]; sigma = par[2]; alpha = par[3]
c.sn = 1/(1-psn(0, location = mu, scale = sigma, shape = alpha, dp = NULL, log = FALSE))
a = min(data); b = max(data); x = seq(a, b, by=0.05)
n = length(x)
ysn = numeric(n)
for (i in 1:n)
    ysn[i] = c.sn * dsn(x[i], location=mu, scale=sigma, shape=alpha)
lines(x, ysn, main = "", type="l", lwd = 2, ylim = c(0,0.6), lty = 3, col="purple")

- The following part computes the AIC and BIC values of the fitted truncated skew normal density:
  k = 3 # number of parameters
  AIC.sn = 2 * (log(c.sn) + fit.sn$value) + 2*k  
  BIC.sn = 2* (log(c.sn) + fit.sn$value) + k*log(m)

- Change point detection using the SIC based method:
  library(sn)
  library(MASS)
  library(moments)
  ## This function computes the log likelihood assuming location and scale parameters are fixed at 0, 1 the degrees of freedom is fixed at 30, only the shape parameter is changing:
LLmu = function (x )
{
  mu = 0 ; sigma = 1 ; alpha = x; r = 30 ; sum = 0
  L = length (Data)
  for (i in 1:L)
    sum = sum + log (pdf.t(Data[i], mu, sigma, alpha, r ))
  return(-sum)
}

## This chunk reads the data and applies the required transformations:
V1= read.table("C:/Documents and Settings/Abeer/Desktop/chiweekly.txt", header = FALSE)$V1
n1 = length (V1)
n = n1 -1
Rt = numeric(n)
for (i in 1:n) Rt[i]= (V1[i+1]-V1[i]) / V1[i]
library(MASS)
truehist (Rt, prob=FALSE ,main="Chile's Weekly Return Rates", xlab="Weekly Return Rates", ylab="Frequency")
round ( summary(Rt), digits= 5)
skewness (Rt)
kurtosis (Rt)
## Iteration to compute SIC for all possible change point locations:
BIC = numeric (n-2)
xbar = mean (Rt); s = sd (Rt)
Rt.stand = (Rt - xbar)/s
Data = Rt.stand
truehist(Rt.stand, prob =TRUE, main= ”Histogram of the Chile’s Standardized Weekly Returns”, xlab= ”Standardized Weekly Rates”, ylab= ”Prob.”)

sk = skewness(Rt.stand)

fit = optim (sk, LLmu ,method = ”Brent”, lower=-20, upper=20 )

BIC[1] = 2 * (fit$value) + 4*log(n) stores BIC(n) under the null hypothesis.

for (J in 2: n-2) {
  L1 = L2 = 0
  Data = numeric (J)
  Data = Rt.stand [1: J]
  alph1 = skewness (Data, na.rm= TRUE)
  fit1 = optim(alph1, LLmu, method="Brent", lower=-20, upper=20)
  if(fit1$convergence == 0){
    L1 = fit1$value
    mu1 = fit1$par
  }
  Data = numeric (n-J)
  for(i in 1:n-J) Data[i] = Rt.stand [J+i]}
  alph2 = skewness(Data, na.rm = TRUE)
  fit2 = optim (alph2, LLmu, method = ”Brent”, lower=-20, upper=20)
  if(fit2$convergence == 0){
    L2 = fit2$value
    mu2 = fit2$par
  }
  BIC[J] = 2* (L1 + L2) + 5*log(n)
  if (L1*L2 ==0) J = J-1

  k.hat = which ( BIC == min(BIC))
  ifelse ( k.hat ==1, ”No further change points”, k.hat )
## Plot data Vs Time plot (c (1:n), Rt.stand, main="Chile’s Weekly Returns Data", xlab="Week #", ylab= "Standardized Weekly Return Rates",type ="l")
abline (v= k.hat, col= "red", lwd=2)
x = seq (2, n-2, by=1)
plot(x, BIC[x], main="Chile’s Weekly Returns Data", xlab= "Week #", ylab= " SIC")
abline(v= k.hat, col= "blue")

## This part detects the change point in the first segment [2: k.hat]
Data = Rt [1: k.hat]
n = length (Data)
BIC = numeric (n-2)
elt = mean (Data); s = sd (Data)
Rt.stand = (Data-elt)/s
sk = skewness (Rt.stand)
Data = Rt.stand
fit = optim (sk, LLmu, method = "Brent", lower=-20, upper=20 )
BIC[1] = 2* (fit$value) + 4*log(n)

for (J in 2:n-2) {
L1 = L2 = 0
Data = numeric (J)
Data = Rt.stand [1 :J]
alph1 = skewness (Data, na.rm=TRUE)
fit1 = optim(alph1, LLmu, method= "Brent", lower=-20, upper=20)
if(fit1$convergence == 0) {
L1 = fit1$value
}
mu1 = fit1$par

Data = numeric (n-J)

for(i in 1:n-J) Data[i] = Rt.stand [J+i] }

alph2 = skewness (Data, na.rm=TRUE)

fit2 = optim (alph2, LLmu, method = "Brent", lower=-20, upper=20)

if(fit2$convergence == 0) {
L2 = fit2$value
mu2 = fit2$par

BIC[J] = 2* (L1 + L2) + 5 * log(n)

if (L1*L2 ==0) J = J-1
}

k.hat1 = which ( BIC == min(BIC) )

ifelse (k.hat1==1, "No further change points", k.hat1 )

plot ( seq(1,k.hat,by=1), BIC[x], main="First Segment Chile’s Weekly Returns Data", xlab="Week #", ylab=" SIC")

abline (v= k.hat1, col="red", lwd=2)

## This part is to detect the change point in the second segment

Rt2 = Rt [k.hat: length(Rt)]

n = length (Rt2)

BIC = numeric (n-2)

xbar = mean(Rt2); s = sd(Rt2)

Rt.stand = (Rt-xbar)/s

Data = Rt.stand

sk = skewness (Data)

fit = optim (sk, LLmu, method = "Brent", lower=-20, upper=20 )
BIC[1] = 2* (fit$value) + 4*log(n)

for (J in 2: n-2) {
  L1 = L2 = 0
  Data = numeric (J)
  Data = Rt.stand [1: J]
  alph1 = skewness(Data, na.rm=TRUE)
  fit1 = optim(alph1, LLmu, method="Brent", lower=-20, upper=20)
  if(fit1$convergence == 0) {
    L1 = fit1$value
    mu1 = fit1$par
    Data = numeric (n-J)
    for (i in 1:n-J) Data[i] = Rt.stand [J+i] }
  alph2 = skewness (Data, na.rm=TRUE)
  fit2 = optim (alph2, LLmu, method = "Brent", lower=-20, upper=20)
  if(fit2$convergence == 0) {
    L2 = fit2$value
    mu2 = fit2$par
    BIC[J] = 2* (L1 + L2) + 5*log(n)
    if (L1*L2 ==0) J = J-1
  }
  k.hat2 = which (BIC == min(BIC)) + k.hat -1
  ifelse (k.hat2 == k.hat, "No further change points", k.hat2 )

plot( seq(k.hat,k.hat+110, by =1),BIC, main="Second Segment Chile’s Weekly Returns Data", xlab="Week #", ylab=" SIC")
abline (v= k.hat2, col= "red", lwd= 2)
cbind (k.hat1, k.hat, k.hat2) # # prints the detected change points
BIBLIOGRAPHY


