

Applications of The Normal Laplace and  
Generalized Normal Laplace Distributions

by

Fan Wu

BA. (Honors) University of Western Ontario 2005

M.Sc. University of Victoria 2008

A Thesis Submitted in Partial Fulfillment of the  
Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Mathematics and Statistics

© Fan Wu, 2008

University of Victoria

*All rights reserved. This thesis may not be reproduced in whole or in part, by  
photocopy or other means, without the permission of the author.*

Applications of The Normal Laplace and  
Generalized Normal Laplace Distributions

by

Fan Wu

BA. (Honors) University of Western Ontario 2005

M.Sc. University of Victoria 2008

Supervisory Committee

Dr. William J. Reed (Department of Mathematics and Statistics)

Supervisor

Dr. Julie Zhou (Department of Mathematics and Statistics)

Departmental Member

Dr. Farouk Nathoo (Department of Mathematics and Statistics)

Departmental Member

Dr. Judith A. Clarke (Department of Economics)

External Examiner

## Supervisory Committee

---

Dr. William J. Reed (Department of Mathematics and Statistics)

Supervisor

---

Dr. Julie Zhou (Department of Mathematics and Statistics)

Departmental Member

---

Dr. Farouk Nathoo (Department of Mathematics and Statistics)

Departmental Member

---

Dr. Judith A. Clarke (Department of Economics)

External Examiner

## Abstract

Two parametric models for income and financial return distributions are presented. There are the four-parameter *normal Laplace* (NL) and the five-parameter *generalized normal Laplace* (GNL) distributions. Their properties are discussed; furthermore, estimation of the parameters by the method of moments and maximum likelihood is presented. The performances of fitting the two models to nine empirical distributions of family income have been evaluated and compared against the four- and five-parameter *generalized beta2* (GB2) and *generalized beta* (GB) distributions which had been previously claimed as best-fitting four- and five- parameter models for income distribution. The results demonstrate that the NL distribution has better performance than the GB and GB2 distributions with the GNL distribution providing an even better fit. Limited application to data on financial log returns shows that the fit of the GNL is comparable to the well-known generalized hyperbolic distribution. However, the GNL suffers from a lack of closed-form expressions for its probability density and cumulative distribution functions, and fitting the distribution numerically is slow and not always reliable. The results of this thesis suggest a strong case for considering the GNL family as parametric models for income data and possibly for financial logarithmic returns.

## Table of Contents

Supervisory committee . . . . .	ii
Abstract . . . . .	iii
Table of Contents . . . . .	iv
List of Figures . . . . .	vi
Acknowledgments . . . . .	vii
List of abbreviations . . . . .	viii
1. Introduction . . . . .	1
1.1 Income Distributions . . . . .	2
1.2 Financial Return Distributions . . . . .	4
1.2.1 Generalized Hyperbolic Distribution . . . . .	6
2. The Normal-Laplace and Generalized Normal Laplace Distributions . . . . .	8
2.1 The Laplace Distribution . . . . .	8
2.2 Normal Laplace Distribution . . . . .	9
2.2.1 Properties . . . . .	10
2.3 Generalized Laplace Distribution . . . . .	11
2.4 Generalized Normal Laplace (GNL) Distribution . . . . .	13
2.5 Properties of the GNL Distribution . . . . .	18
2.5.1 Infinite Divisibility . . . . .	18
2.5.2 Mean, Variance and Cumulants . . . . .	19
2.6 Numerical determination of the pdf and cdf of the GNL . . . . .	22
2.6.1 Using the Representation as a Convolution . . . . .	23
2.6.2 Numerical Inversion of Characteristic Function . . . . .	24
2.6.3 Using the representation as a Normal mean-variance mixture . . . . .	25

3. Methods of Estimation . . . . .	27
3.1 Method of Moments . . . . .	27
3.2 Maximum Likelihood Estimation . . . . .	30
3.3 Nelder-Mead Method: Multi-dimensional Maximization Method . . . . .	34
4. Simulation Studies for The GNL distribution . . . . .	36
4.1 Simulating GNL Data . . . . .	36
4.2 Simulating GH Data . . . . .	40
5. Application to Income Data . . . . .	43
5.1 Description of the Income Data . . . . .	43
5.2 Income Distribution Results . . . . .	46
6. Application To Financial Data . . . . .	50
6.1 Description of the Stock Price Data . . . . .	50
6.2 Results of fitting GNL and GH . . . . .	50
7. Conclusions . . . . .	53
References . . . . .	55
Appendix . . . . .	60

## List of Figures

2.1	The effect of the parameters of GL and the comparison of GL and normal distributions. . . . .	13
2.2	A comparison of the pdf curves of GNL(0,1,1,1, $\rho$ ) and N (0,1) distributions.	14
2.3	The pdf curves of three GNL distributions (effect of $\mu$ ) . . . . .	15
2.4	The pdf curves of three GNL distributions (effect of $\sigma$ ) . . . . .	16
2.5	The pdf curves of three GNL distributions (effect of $\alpha$ ) . . . . .	17
2.6	The pdf curves of three GNL distributions (effect of $\beta$ ) . . . . .	17
2.7	The pdf curves of four GNL distributions (effect of $\rho$ ) . . . . .	18
2.8	The pdf curves and means of three GNL distributions (effect of $\alpha$ on mean) . . . . .	20
2.9	The pdf curves and means of three GNL distributions (effect of $\beta$ on mean) . . . . .	21
4.1	Q-Q plots for the simulated GNL data . . . . .	38
4.2	Q-Q plots for the simulated GNL data . . . . .	39
4.3	Q-Q plot for the GNL . . . . .	41
4.4	Q-Q plot for the GH . . . . .	42
6.1	The GNL fitted to IBM 05-06 . . . . .	51
6.2	The GH fitted to IBM 05-06 . . . . .	51
6.3	The GNL fitted to CitiGroup 05-06 . . . . .	52
6.4	The GH fitted to CitiGroup 05-06 . . . . .	52

## Acknowledgments

I would first like to thank my father Wu, Shaoqin, mother Chen, Yuhua and all my family members. Without their support, I could not have got to the point of writing this thesis. Dr. William J. Reed has guided me in the fields of statistics and finance, and provided many ideas, and offered infinite help and patience during my study at the University of Victoria. I owe him a lot. I would like to acknowledge my gratitude to my committee members, Dr. Julie Zhou and Dr. Farouk Nathoo who have been very helpful in assisting in the completion of the thesis. I would also like to acknowledge LIS. Without their data, this thesis could not have been completed.

## List of abbreviations

cdf.....	Cumulative Distribution Function	(1)
cgf.....	Cumulant Generating Function	(2)
ch.f.....	Characteristic Function	(3)
pdf.....	Probability Distribution Function	(4)
Dist'n.....	Distribution	(5)
GL.....	Generalized Laplace	(6)
GNL.....	Generalized Normal Laplace	(7)
GB2.....	Generalized Beta 2	(8)
GB.....	Generalized Beta	(9)
GH.....	Generalized Hyperbolic	(10)
GBM.....	Geometric Brownian Motion	(11)
LIS.....	Luxembourg Income Study	(12)
MLE.....	Maximum Likelihood Estimation	(13)
MME.....	Method of Moments Estimation	(14)
NL.....	Normal Laplace	(15)
Q-Q plot.....	Quantile-quantile plot	(16)
SSE.....	Sum of Squared Errors	(17)
SAE.....	Sum of Absolute Errors	(18)

## Chapter 1

### Introduction

A major aim of this thesis is to examine the performance of two new parametric probability distributions, the *normal Laplace* (NL) and *generalized normal Laplace* (GNL), in applications in economics and finance. In economics we consider fitting the four-parameter NL distribution and the five-parameter GNL distribution to data on income and earnings distributions. Reed and Jorgensen (2004) presented a number of examples of the fit of the NL distribution to various empirical size distributions and Reed (2004) gave examples of its fit to income distributions for four widely differing data sets. However to date no comparison of the fit of the NL (or GNL) with other proposed models has been conducted. In this thesis we perform such a comparison using nine different empirical income distributions. In the finance field we consider fitting the GNL distribution to logarithmic returns on financial assets. Reed (2004) has shown how a Lévy process, which he called Brownian-Laplace Motion whose increments follow the GNL distribution can be constructed and used for modelling stock-price dynamics; he obtained an option pricing formula for assets following such a process. The GNL distribution can exhibit skewness and excess kurtosis, properties present in high-frequency data of logarithmic returns. It therefore seems a good candidate model for use in option pricing. An aim of this thesis is to explore how well it fits to actual stock-price data, and to compare it with other proposed models.

## 1.1 Income Distributions

Many probability density functions have been proposed as parametric models for income distributions. The earliest model proposed was that of Pareto (1895). While fitting empirical distributions well in the upper tail, the eponymous *Pareto distribution* did not fit well the lower tail. Gibrat (1931) proposed the *Lognormal distribution* with two parameters based on a simple model for income evolution. This was further explored by Aitchinson and Brown (1969). Other two parameter models used have been the *gamma* (Ammon, 1925) and the *Weibull* (Bartels and van Metelel 1975) distributions. Since none of these two-parameter models provided completely satisfactory fits, various three-parameter models have been suggested. For example, Thurow (1970) used a three-parameter distribution, which he called the *beta distribution of the first kind*, and Amoroso (1924-25) and Taille (1981) applied the *generalized gamma distribution* with three parameters to model income distributions. Dagum (1977) introduced another three-parameter distribution, the *Dagum Type I distribution*, and two generalizations *Dagum Types II and III* (Dagum, 1977, 1980) as models for income distributions. Dagum reports that these families give a better fit to empirical income distributions than any of the previously considered functions, including the *Singh-Maddala (1976) distribution*. In the statistics literature, the Dagum, Bartels and the Singh-Maddala distributions are known under different names; they all belong to a classification system due to Burr (1942) (Kleiber, 1996).

Later studies, using generalization of these models, were used to find a better fit for income distribution data. McDonald (1984) proposed using two distributions which he termed the *generalized beta of the first and second kind* (GB1,GB2) with four parameters. These four-parameter generalized beta distributions include the beta distributions of the first kind and second kind (B1, B2), the gamma, and the

lognormal as special cases, and proved to provide better fits than previous models. The GB2 on the whole, provided a better fit than the GB1. (McDonald, 2002).

Subsequently, McDonald and Xu (1995) presented a new generalized five-parameter distribution which they called the *generalized beta (GB) distribution*, which nests all of the distributions that we have mentioned above as special cases. The GB is more flexible and of course having all of the others nested within it provided a better fit (McDonald 2002).

The five-parameter *Generalized Beta (GB) distribution* has probability density function (pdf):

$$f(y; a, b, c, p, q) = \frac{|a|y^{ap-1}(1 - (1 - c)(y/b)^a)^{q-1}}{b^{ap}B(p, q)(1 + c(y/b)^a)^{p+q}} \quad \text{for } 0 < y^a < b^a \quad (1.1)$$

and 0 otherwise, where  $0 \leq c \leq 1$ ,  $b$ ,  $p$ ,  $q$  are positive constants; and  $B(p, q)$  is the familiar beta function:  $B(p, q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$ . The *Generalized Beta Distribution of the second kind* (GB2) is a four-parameter distribution, which is a special case of the GB distribution when  $c=1$ . Many of the important properties and applications of the GB2 distribution can be found in McDonald and Xu (1995). The density function for the GB2 is

$$f(y; a, b, p, q) = \frac{|a|y^{ap-1}}{b^{ap}B(p, q)(1 + (y/b)^a)^{p+q}} \quad (1.2)$$

The parameters  $a$ ,  $p$ , and  $q$  influence the shape of the distribution, and  $b$  is a scale parameter. The cumulative distribution function for the GB2 is

$$F(y; a, b, p, q) = \frac{z^p {}_2F_1[p, 1 - q, 1 + p, z]}{pB(p, q)} \quad (1.3)$$

where  $z = (y/b)^a / (1 + (x/b)^a)$  and  ${}_2F_1[a, b, c, z]$  is Gauss' hypergeometric function described in detail by Abramowitz and Stegun(1972, p. 563).

## 1.2 Financial Return Distributions

We consider modeling logarithmic returns for financial assets in the form of a time series  $S_1, S_2, S_3, \dots$ , where  $S_n$  represents the closing asset price in period  $n$ . In the early development of the option-pricing theory (Black-Scholes), the asset price was assumed to follow geometric Brownian motion (GBM) a consequence of which was that  $S_t$  would be lognormally distributed and therefore, the logarithm of so called financial returns would be normally distributed, *i.e.*

$$\log \frac{S_{t+1}}{S_t} \stackrel{iid}{\sim} N(\mu, \sigma^2) \quad (1.4)$$

(where  $N(m, v)$  denotes a normal distribution with mean  $m$  and variance  $v$ ). The parameter  $\mu$  is the expected return and  $\sigma$  is the volatility of asset price. The option pricing theory of Black and Scholes (e.g. Cvitanic and Zapatero, 2004) also relies on several other important assumptions. For example, taxes, and transaction costs are excluded. Since the introduction of Black-Scholes option pricing, more detailed statistical analysis has revealed that real financial return distributions often depart from normality, especially when the reporting period is short. In this case, they are often skewed, and have excess kurtosis with longer tails than those of the normal distribution (see *e.g.* Rydberg, 2000).

New models for asset price evolution based on Lévy processes have been proposed. For such models the increments can exhibit both skewness and excess kurtosis. The mathematics of Lévy processes is somewhat esoteric and will not be discussed in this thesis. However, an important aspect is that given any infinitely divisible distribution, a Lévy process can be constructed with the marginal distribution of its increments following the given distribution.

Some examples of Lévy process which have been proposed are

- *The Gamma Process* (Ammon, 1895). In this process, the increments follow the Gamma distribution  $\text{Gamma}(a,b)$  with parameters  $a > 0$  and  $b > 0$  with probability density function(pdf) given by

$$f(x; a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-xb), \quad x > 0$$

This process is a pure jump process, with no continuous component.

- *The Generalized Inverse Gaussian Process* (GIG) (Seshadri, 1993). For this process the distribution of the increments has a pdf

$$f(x; \lambda, a, b) = \frac{(b/a)^\lambda}{2K_\lambda(ab)} x^{\lambda-1} \exp(-\frac{1}{2}(a^2 x^{-1} + b^2 x)), \quad x > 0$$

$$\text{where } K_\lambda(x) = \frac{1}{2} \int_0^\infty y^{\lambda-1} \exp(-\frac{1}{2}x(y + y^{-1})) dy$$

denotes the modified Bessel function of the third kind with index  $\lambda$ .

The *Inverse Gaussian (IG) Process* (Chhikara and Folks, 1989) is a special case of GIG, when  $\lambda = -1/2$ .

- *The Variance Gamma (Laplace) Process* (VG) (Madan and Seneta, 1990). For this process the increments follow the Variance Gamma distribution with the pdf

$$f(x; \alpha, \mu, \theta, \sigma) = \sqrt{\frac{\pi}{2}} \frac{\alpha^\alpha e^{(x-\mu)\theta/\sigma^2}}{\sigma \Gamma(\alpha)} \left( \frac{|x-\mu|}{\sqrt{\theta^2 + 2\alpha\sigma^2}} \right)^{\alpha-1/2} K_{\alpha-1/2} \left( \frac{|x-\mu|\sqrt{\theta^2 + 2\alpha\sigma^2}}{\sigma^2} \right)$$

- *The Meixner Process* (Schoutens and Teugels, 1998). For this process of the increments have pdf

$$f(x; \alpha, \beta, \delta) = \frac{(2 \cos(\beta/2))^{2\delta}}{2\alpha\pi\Gamma(2\delta)} \exp\left(\frac{bx}{a}\right) \left| \Gamma\left(\delta + \frac{ix}{\alpha}\right) \right|^2$$

- *The CGMY Process* (Carr, Geman, Madan and Yor, 2002). For this process the distribution of the increments has a characteristic function of the form

$$\phi(u; C, G, M, Y) = \exp(C\Gamma(-Y)((M - iu)^Y - M^Y + (G + iu)^Y - G^Y))$$

- *The Generalized Hyperbolic Process* (Eberlein and Hammerstein, 2002). For this process the distribution of the increments has a pdf

$$f(x; \lambda, \alpha, \beta, \delta, \mu) = a(\lambda, \alpha, \beta, \delta)(\delta^2 + (x - \mu)^2)^{(\lambda - \frac{1}{2})/2} \\ \times K_{\lambda - \frac{1}{2}}(\alpha\sqrt{\delta^2 + (x - \mu)^2}) \exp(\beta(x - \mu)),$$

where

$$a(\lambda, \alpha, \beta, \delta) = \frac{(\alpha^2 - \beta^2)^{\lambda/2}}{\sqrt{2\pi}\alpha^{\lambda - \frac{1}{2}}\delta^\lambda K_\lambda(\delta\sqrt{\alpha^2 - \beta^2})}$$

and  $K_\lambda$  the modified Bessel function of the third kind with index  $\lambda$ .

### 1.2.1 Generalized Hyperbolic Distribution

Barndorff-Nielsen (1977) introduced the four-parameter *hyperbolic distribution*, which he fitted to the size distribution of aeolian sand particles. Subsequently it has been fitted to size distributions in various fields such as physics, biology and agronomy. The generalized hyperbolic (GH) process used to model dynamics of logarithm stock price returns (Eberlein and Keller, 1995). In their work, they fitted GH distributions to German stock prices and the results were highly accurate. The five-parameter *generalized hyperbolic* (GH) distribution was introduced by Eberlein and Hammerstein (2002). In the early 90's, Blsild and Srensen (1992) developed a computer program, named HYP, to estimate the parameters of multivariate hyperbolic distributions by

maximum likelihood in up to three dimensions. From the pdf of the GH, the log-likelihood function for the independent observations  $x_i, i= 1, \dots, n$  is:

$$\begin{aligned} \ell_{GH}(\lambda, \alpha, \beta, \delta, \mu) = & n \log a(\lambda, \alpha, \beta, \delta) + \left(\frac{\lambda}{2} - \frac{1}{4}\right) \sum_{i=1}^n \log(\delta^2 + (x_i - \mu)^2) \\ & + \sum_{i=1}^n [\log K_{\lambda-\frac{1}{2}}(\alpha \sqrt{\delta^2 + (x_i - \mu)^2}) + \beta(x_i - \mu)] \end{aligned} \quad (1.5)$$

The parameters are as follows:  $\mu \in \Re$  is a location parameter,  $\alpha > 0$  determines the shape,  $0 \leq |\beta| < \alpha$  relates to the skewness and  $\delta > 0$  serves for scaling.  $\lambda \in \Re$  characterizes certain subclasses and influences considerably the size of mass contained in the tail.

A detailed description of the normal-Laplace (NL) and generalized normal-Laplace (GNL) distributions which are the main subject of this thesis and some of their properties are given in the next chapter. Chapter 3 deals with method of moments and maximum likelihood parameter estimation for the NL and GNL for both grouped and ungrouped data. Chapter 4 presents simulation studies for comparing the GNL and GH (generalized hyperbolic) distributions. Chapter 5 considers comparisons of the fit of the four-parameter NL and the five-parameter GNL distributions with the four- and five-parameter GB family for grouped income data. Chapter 6 considers comparisons of the fit of GNL distribution with GH distribution for ungrouped logarithm returns of stock price. Conclusions are given in Chapter 7.

## Chapter 2

# The Normal-Laplace and Generalized Normal Laplace Distributions

### 2.1 The Laplace Distribution

The classical Laplace distribution with mean zero and variance  $\sigma^2$  was introduced by Laplace in 1774 (see e.g. Kotz *et al.*, 2001). The distribution is symmetrical and leptokurtic, which means its shape is more peaked (has higher kurtosis) than that of the normal distribution. It has a characteristic function (ch.f)

$$\phi(t) = \frac{1}{1 + \frac{\sigma^2 t^2}{2}} \quad (2.1)$$

and pdf

$$f(x) = \frac{\sqrt{2}}{2\sigma} e^{-\sqrt{2}|x|/\sigma}, x \in \mathfrak{R}, \sigma > 0. \quad (2.2)$$

This distribution has been used for modelling data that have heavier tails than those of the normal distribution.

The skew-Laplace distribution (or asymmetric Laplace) is an asymmetric version of the Laplace distribution (Kotz *et al.*, 2001). Its pdf can be written

$$f(x) = \begin{cases} \frac{\alpha\beta}{\alpha+\beta} \exp^{-\alpha(x-\mu)} & x \geq \mu \\ \frac{\alpha\beta}{\alpha+\beta} \exp^{\beta(x-\mu)} & x < \mu \end{cases} \quad (2.3)$$

where  $\mu$  is a location parameter and the parameters  $\alpha$  and  $\beta$  influence right and left-tails, respectively. A value of  $\alpha$  greater than  $\beta$  results in less probability to the right side of  $\mu$  than to the left side; the opposite is of course true if  $\beta$  is greater than  $\alpha$ . If  $\alpha = \beta$ , the distribution is symmetrical.

## 2.2 Normal Laplace Distribution

The *normal Laplace* (NL) distribution (Reed and Jorgensen, 2004) is a relatively new distribution that belongs to the *generalized normal Laplace* (GNL) distribution family (Reed, 2000). It has been used to describe the distribution of incomes, particle sizes, oil-field sizes, city sizes, and other phenomena. It results from the convolution of independent normal and asymmetric Laplace components.

$$X \stackrel{d}{=} Z + W \quad (2.4)$$

where  $Z$  is a normally distributed random variable with mean  $\mu$  and variance  $\sigma^2$ , and  $W$  has the asymmetric Laplace distribution (2.3) with parameters  $\mu = 0$ ,  $\alpha$  and  $\beta$ . The cumulative distribution function (cdf) of the NL distribution can be showed to be (Reed and Jorgensen, 2004)

$$F(x) = \Phi\left(\frac{x - \mu}{\sigma}\right) - \phi\left(\frac{x - \mu}{\sigma}\right) \frac{\beta R(\alpha\sigma - (x - \mu)/\sigma) - \alpha R(\beta\sigma + (x - \mu)/\sigma)}{\alpha + \beta} \quad (2.5)$$

where  $\Phi$  and  $\phi$  are the cdf and pdf of a standard normal random variable and  $R$  is *Mills' ratio*:

$$R(z) = \frac{\Phi^c(z)}{\phi(z)} = \frac{1 - \Phi(z)}{\phi(z)} \quad (2.6)$$

The cdf above depends on four parameters:  $\mu \in \Re$  is a location parameter;  $\sigma > 0$  is the scale parameter for the normal component;  $\alpha > 0$  and  $\beta > 0$  are parameters controlling tail behaviour. Since the likelihood function for grouped data is expressed in terms of cdf, equation (2.4) above is very useful when fitting to data.

The probability density function (pdf) is

$$f(x) = \frac{\alpha\beta}{\alpha + \beta} \phi\left(\frac{x - \mu}{\sigma}\right) [R(\alpha\sigma - (x - \mu)/\sigma) + R(\beta\sigma + (x - \mu)/\sigma)] \quad (2.7)$$

Because an asymmetric Laplace distribution can be represented as a difference between independent exponential random variables (see e.g. Kotz *et al.*, 2001, p146) the normal-Laplace can be represented as

$$X \stackrel{d}{=} \mu + \sigma Z + E_1/\alpha - E_2/\beta \quad (2.8)$$

where  $Z$  denotes a standard normal random variable which is independent of two independent standard exponential random variables,  $E_1, E_2$ .

### 2.2.1 Properties

The following properties of the NL distribution are derived in Reed and Jorgensen (2004)

- Characteristic function (ch.f). From the expression (2.8) the ch.f of the NL distribution can be expressed as the product of the ch.fs of its normal and two exponential components,

$$\phi_{NL}(s) = \frac{\alpha\beta \exp(i\mu s - \sigma^2 s^2/2)}{(\alpha - is)(\beta + is)} \quad (2.9)$$

- Mean, variance and cumulants. From the ch.f (2.9), the mean, variance and cumulants can be determined. They are

$$E(X) = \mu + \frac{1}{\alpha} - \frac{1}{\beta}; \quad \text{var}(X) = \sigma^2 + \frac{1}{\alpha^2} + \frac{1}{\beta^2} \quad (2.10)$$

and when  $r > 2$ , the higher order cumulants are given as

$$\kappa_r = (r-1)! \left( \frac{1}{\alpha^r} + (-1)^r \frac{1}{\beta^r} \right) \quad (2.11)$$

- Two special limiting cases: when  $\alpha \rightarrow \infty$  and  $\beta \rightarrow \infty$ .

When  $\alpha = \infty$ , only the lower tail is fatter than the corresponding normal distribution, and the upper tail reduces to be the same as that of normal. The pdf (2.7) becomes

$$f_1(x) = \beta \phi\left(\frac{x-\mu}{\sigma}\right) R\left(\beta\sigma + \left(\frac{x-\mu}{\sigma}\right)\right) \quad (2.12)$$

Similarly, when  $\beta = \infty$ , the only upper tail of the distribution is fatter from normal; the lower tail behaves the same as that of a normal distribution. The pdf (2.7) reduces to

$$f_2(x) = \alpha \phi\left(\frac{x-\mu}{\sigma}\right) R\left(\alpha\sigma - \left(\frac{x-\mu}{\sigma}\right)\right) \quad (2.13)$$

- Representation as a mixture. The NL distribution can be represented as a mixture of the above two special limiting cases:

$$f_{NL}(x) = \frac{\alpha}{\alpha+\beta} f_1(x) + \frac{\beta}{\alpha+\beta} f_2(x) \quad (2.14)$$

where  $f_1$  and  $f_2$  denote the pdfs of the NL when  $\alpha = \infty$  and  $\beta = \infty$  respectively.

- When  $\alpha = \beta$ . The NL distribution becomes symmetric with the pdf as

$$f(x) = \frac{\alpha}{2} \phi\left(\frac{x-\mu}{\sigma}\right) [R(\alpha\sigma - (x-\mu)/\sigma) + R(\alpha\sigma + (x-\mu)/\sigma)] \quad (2.15)$$

### 2.3 Generalized Laplace Distribution

A generalization of the Laplace distribution known as the *Generalized Laplace* (GL) (see *e.g.* Kotz *et al.*, 2001) has four parameters,  $\theta, k \in \Re, \sigma, \tau \in \Re^+$ . ch.f

$$\phi(t) = \exp^{i\theta t} \left( \frac{1}{1 + i\frac{\sqrt{2}}{2}\sigma kt} \right)^\tau \left( \frac{1}{1 - i\frac{\sqrt{2}}{2}\sigma kt} \right)^\tau \quad (2.16)$$

Its probability density function (pdf) is:

$$f(x) = \frac{\sqrt{2}e^{\frac{\sqrt{2}}{2\sigma}(1/k-k)(x-\theta)}}{\sqrt{\pi}\sigma^{\tau+1/2}\Gamma(\tau)} \left(\frac{\sqrt{2}|x-\theta|}{k+1/k}\right)^{\tau-1/2} K_{\tau-1/2}\left(\frac{\sqrt{2}}{2\sigma}\left(\frac{1}{k}+k\right)|x-\theta|\right) \quad (2.17)$$

where  $K_\lambda$  is the modified Bessel function of third kind with index  $\lambda$ . There are some special cases associated with the distribution. For  $\tau = 1$ , we have an asymmetric Laplace, and for  $k = 1$  and  $\theta = 0$ , we obtain a symmetric Laplace distribution.

The pdf (2.17) can be written

$$f(x) = (\alpha\beta)^\tau \exp\left(\frac{\beta-\alpha}{2}x\right) \left(\frac{|x|}{\alpha+\beta}\right)^{\tau-1/2} K_{\tau-1/2}\left(\frac{\alpha+\beta}{2}|x|\right) \quad (2.18)$$

where  $\alpha$  and  $\beta$  describe the left and right-tail shapes, and have the same role as  $\alpha$  and  $\beta$  in Laplace distribution; and  $\tau$  is a parameter relating to the peakedness of the pdf. We shall denote such a distribution by  $GL(\alpha, \beta, \tau)$

Figure 2.1 (a) and (b) illustrate the effect of the parameters,  $\alpha$ ,  $\beta$  and  $\tau$ . Using the parameterization (2.18), Figure 2.1 (a) shows the effect of  $\tau$  on the shape of the distribution. The three curves are for  $GL(1, 1, \tau)$  with  $\tau = 0.8$  (red), 1 (blue), and 2 (yellow). Figure 2.1 (b) shows the effect of  $\alpha$  and  $\beta$  with  $GL(\alpha, \beta, 0.8)$  where  $\alpha = 3$  and  $\beta = 1$  (green), and  $\alpha = 1$  and  $\beta = 5$  (light blue). Figure 2.1 (c) shows the pdf curves for the  $GL(1, 1, 0.8)$  (Green),  $GL(1, 1, 1)$  and Normal (0, 1) (black dotted) distributions.

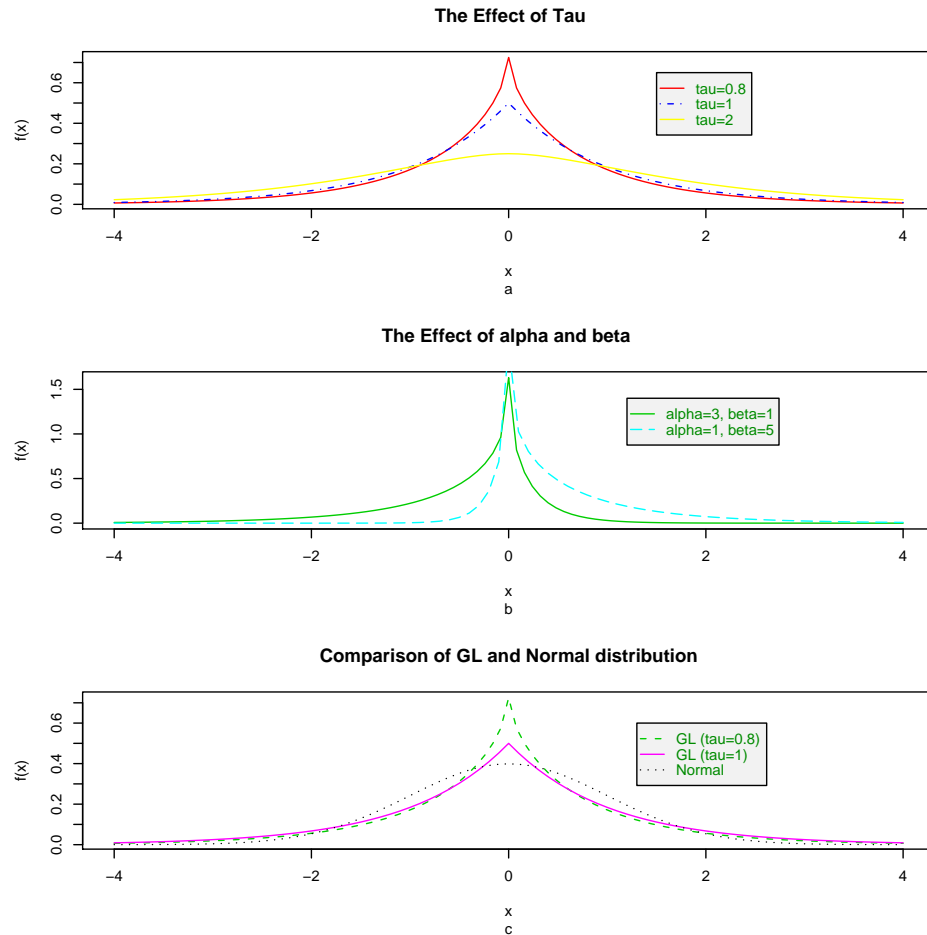


Figure 2.1: The effect of the parameters of GL and the comparison of GL and normal distributions.

## 2.4 Generalized Normal Laplace (GNL) Distribution

The *generalized normal Laplace* (GNL) distribution was introduced by Reed (2004) and has been used for modelling financial logarithmic price returns. A closed-form of the pdf of the GNL has not been found; however, it can be obtained from the convolution of independent normal and generalized Laplace distributions. The GNL

distribution is defined as a random variable  $X$  with ch.f

$$\phi_{GNL}(s) = \left[ \frac{\alpha\beta \exp(i\mu s - \sigma^2 s^2/2)}{(\alpha - is)(\beta + is)} \right]^\rho \quad (2.19)$$

where  $\mu \in \Re$  a location parameter,  $\sigma \in \Re^+$  is the scale parameter for the normal component,  $\alpha, \beta \in \Re^+$  are parameters influencing tail behavior and  $\rho \in \Re^+$  is a shape parameter ( $\rho$  corresponds to the parameter  $\tau$  in GL component, (2.18)).

Figure 2.2 compares the pdf curves of GNL with various  $\rho$  ( $=0.5, 0.9, 2$ ); the NL (or GNL(0.5, 0.9, 1)) and normal distributions (black dotted). The graph illustrates how the GNL distribution has fatter tails than those of normal distribution.

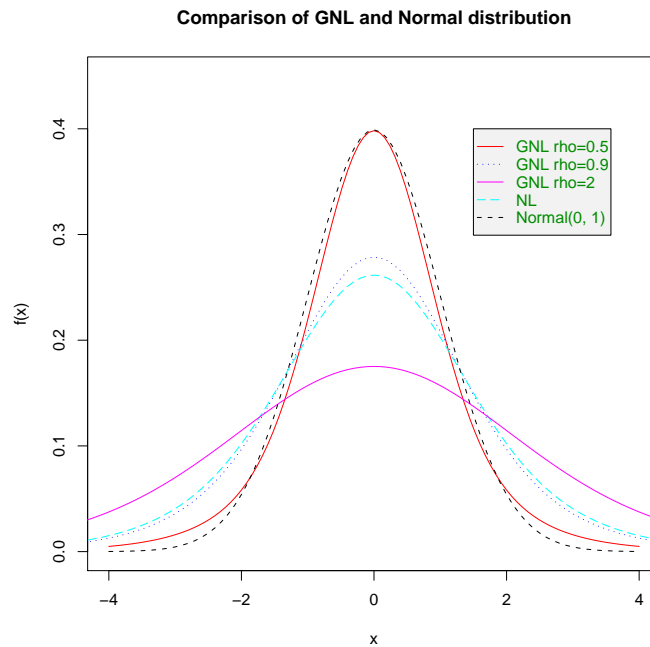


Figure 2.2: A comparison of the pdf curves of GNL(0,1,1,1,  $\rho$ ) and N (0,1) distributions.

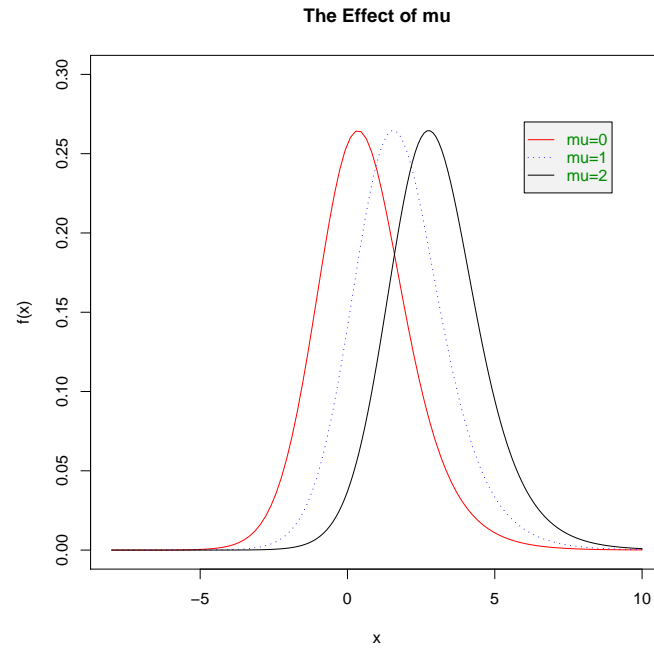


Figure 2.3: The pdf curves of three GNL distributions with  $\sigma^2=1$ ,  $\alpha=1$ ,  $\beta=2$  and  $\rho=1.2$ , when  $\mu=0$  (red),  $\mu=1$  (blue dotted) and  $\mu=2$  (black)

The role played by location parameter  $\mu$  is clearly shown in Figure 2.3: an increase in  $\mu$  moves the pdf curve of GNL distribution rightward horizontally.

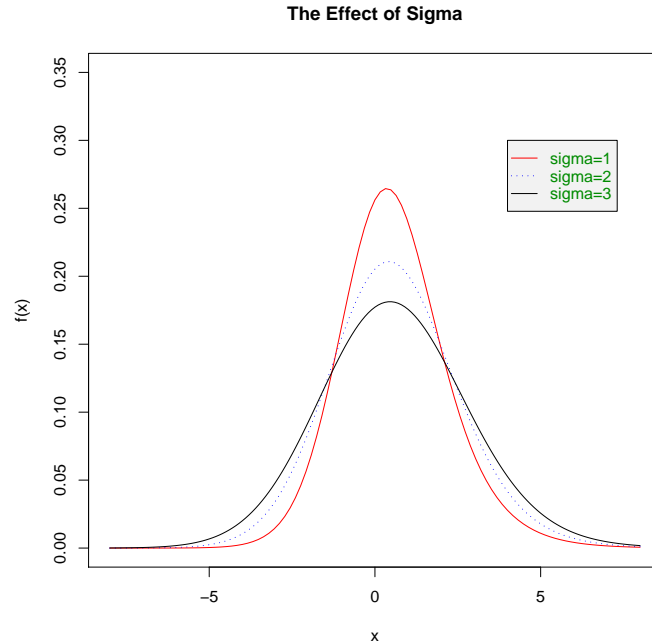


Figure 2.4: The pdf curves of three GNL distributions with  $\mu=0$ ,  $\alpha=1$ ,  $\beta=2$ , and  $\rho=1.2$ , when  $\sigma=1$  (red),  $\sigma=2$  (blue dotted) and  $\sigma=3$  (black)

Figure 2.4 shows the effect of the parameter  $\sigma$ ; with an increase in  $\sigma$ , the pdf curve becomes wider and flatter the same time.

The parameter,  $\alpha$ , affects the upper tail behavior of the GNL distribution: Figure 2.5 shows the change in the upper tail as  $\alpha$  increases. Small values of  $\alpha$  correspond to a fat upper tail. When  $\alpha = \infty$ , the upper tail of the distribution reduces to that of a normal distribution. Figure 2.6 shows similar behavior to the parameter  $\beta$  having effects on the lower tail.

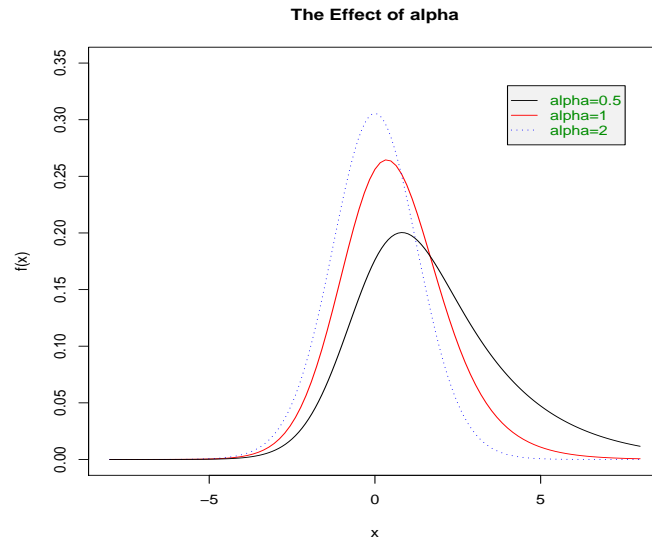


Figure 2.5: The pdf curves of three GNL distributions with  $\mu=0$ ,  $\sigma=1$ ,  $\beta=2$ , and  $\rho=1.2$ , when  $\alpha=1$  (red),  $\alpha=2$  (blue dotted) and  $\alpha=0.5$  (black)

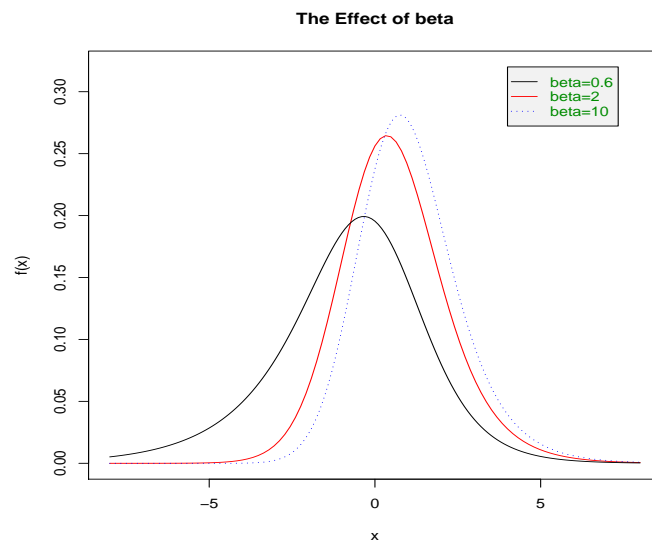


Figure 2.6: The pdf curves of three GNL distributions with  $\mu=0$ ,  $\sigma=1$ ,  $\alpha=1$ , and  $\rho=1.2$ , when  $\beta=2$  (red),  $\beta=10$  (blue dotted) and  $\beta=0.6$  (black)

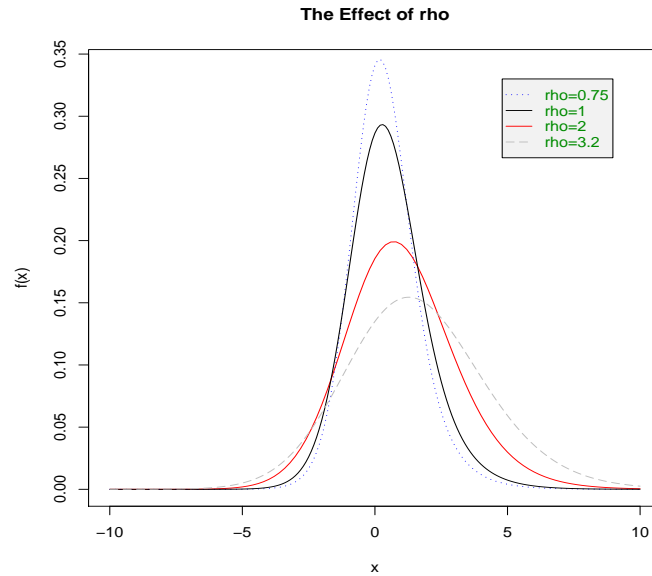


Figure 2.7: The pdf curves of four GNL distributions with  $\mu=0$ ,  $\sigma=1$ ,  $\alpha=1$ , and  $\beta=2$ , when  $\rho=2$ (red),  $\rho=0.75$  (blue dotted),  $\rho=1$  (black) and  $\rho=3.2$  (gray dash)

Figure 2.7 illustrates the effect of the parameter  $\rho$ . Increasing  $\rho$  both increase the mean and variance. It also changes the shape of the distribution with smaller values of  $\rho$  resulting in a distribution with sharp peak, thinner flanks and longer tails.

## 2.5 Properties of the GNL Distribution

### 2.5.1 Infinite Divisibility

Equation (2.19) demonstrates that the GNL is infinitely divisible<sup>1</sup>. As a consequence a Lévy process with increments following the GNL distribution can be constructed. Reed (2006) did this calling the resulting process Brownian-Laplace Motion.

<sup>1</sup>Suppose  $\phi(u)$  is the characteristic function of a distribution. If, for every positive integer  $n$ ,  $\phi^{1/n}(u)$  is also a characteristic function, then the distribution is *infinitely divisible*.

For such a process  $S_t$  the increments  $S_{w+t} - S_w$  have a characteristic function:

$$\left[ \frac{\alpha\beta \exp(i\mu s - \sigma^2 s^2/2)}{(\alpha - is)(\beta + is)} \right]^{\rho t} = [\phi_0(s)]^t \quad (2.20)$$

where  $\phi_0(s)$  is the characteristic function of a GNL variate of the form (2.19). It can be seen that the length  $t$  of the time increment affects only the exponent parameter  $\rho$  of the GNL distribution.

### 2.5.2 Mean, Variance and Cumulants

The cumulants  $\kappa_n$  of a distribution are defined as (Abramowitz and Stegun, 1972, p. 928)

$$\log(\phi(s)) = \sum_{n=1}^{\infty} \kappa_n \frac{(is)^n}{n!} \quad (2.21)$$

In particular, the first and second cumulants are the mean and variance of the distribution. For a GNL distribution

$$\begin{aligned} \log(\phi_{GNL}(s)) &= \rho\mu is + \rho\sigma^2 \frac{(is)^2}{2!} + \log\left(\frac{\alpha}{\alpha - is}\right)^\rho + \log\left(\frac{\beta}{\beta + is}\right)^\rho \\ &= \rho\mu is + \rho\sigma^2 \frac{(is)^2}{2!} + \rho \log \frac{\alpha}{\alpha - is} + \rho \log \frac{\beta}{\beta + is} \\ &= is\rho\left(\mu + \frac{1}{\alpha} - \frac{1}{\beta}\right) + \frac{1}{2!}(is)^2\rho\left(\sigma^2 + \frac{1}{\alpha^2} + \frac{1}{\beta^2}\right) + \frac{1}{3!}(is)^3\rho\left(\frac{2}{\alpha^3} - \frac{2}{\beta^3}\right) + \dots \end{aligned} \quad (2.22)$$

using the Maclaurin series expansions of  $\log \frac{\alpha}{\alpha - is}$  and  $\log \frac{\beta}{\beta + is}$ . We thus obtain the mean and variance

$$E(X) = \rho\left(\mu + \frac{1}{\alpha} - \frac{1}{\beta}\right); \quad \text{var}(X) = \rho\left(\sigma^2 + \frac{1}{\alpha^2} + \frac{1}{\beta^2}\right) \quad (2.23)$$

and the higher order cumulant functions ( $r > 2$ )

$$\kappa_r = \rho(r-1)! \left( \frac{1}{\alpha^r} + (-1)^r \frac{1}{\beta^r} \right) \quad (2.24)$$

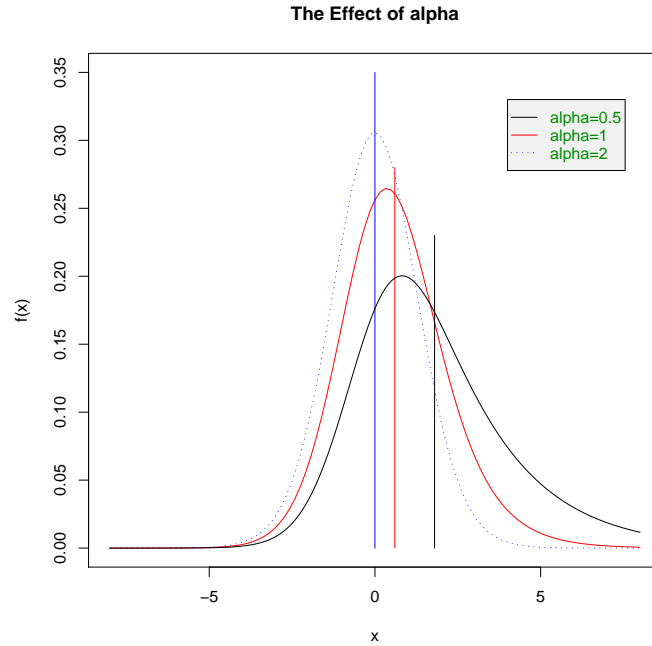


Figure 2.8: The pdf curves and means of three GNL distributions with  $\mu=0$ ,  $\sigma=1$ ,  $\beta=2$ ,  $\rho=1.2$ , when  $\alpha=1$  (red),  $\alpha=2$  (blue dotted) and  $\alpha=0.5$  (black)

Note the coefficients of the kurtosis and skewness

$$\begin{aligned} \frac{k_4}{k_2^2} &= \frac{6(\alpha^4 + \beta^4)}{\rho(\sigma^2\alpha^2\beta^2 + \alpha^2 + \beta^2)^2} \\ \frac{k_3}{k_2^{3/2}} &= \frac{2(\beta^3 - \alpha^3)}{\rho^{1/2}(\sigma^2\alpha^2\beta^2 + \alpha^2 + \beta^2)^2} \end{aligned} \quad (2.25)$$

The coefficients of the kurtosis and skewness involve all of the parameters, except  $\mu$  thus any changes in parameters in the GNL distribution will change the values of the kurtosis and skewness. When  $\rho$  increases, both of the kurtosis and skewness will decrease. As  $\rho \rightarrow \infty$ , both values converge to zero. Thus the kurtosis and skewness of the increments  $S_{w+t} - S_w$  of Brownian-Laplace motion decrease as the length  $t$  of the increment increases becoming zero in the limit (as  $t \rightarrow \infty$ ), as the distribution converges to normality. Such behaviour has been observed for logarithmic returns

on financial assets. From the expression for the skewness, it can be seen that  $\alpha$  and  $\beta$  determine in which direction the pdf will skew. If  $\alpha > \beta$ , the GNL distribution is skewed to the left, and vice versa. If  $\alpha = \beta$ , then the GNL distribution is symmetric.

The expression (2.23) indicates that  $\rho$ ,  $\mu$ ,  $\alpha$  and  $\beta$  will all have influence on the mean. An increase of  $\rho$ ,  $\mu$ , and  $\beta$  will result in an increase of the mean; while, parameter  $\alpha$  affects the mean in the opposite way. When  $\alpha$  increases, the mean decreases. The effects of  $\rho$  and  $\mu$  can be seen in Figure 2.3 and Figure 2.7: the pdf curves move rightward when  $\rho$  or  $\mu$  increases. Figure 2.8 and Figure 2.9, in which each colored vertical line in the Figures corresponds to the mean of the GNL distribution, demonstrate the negative impact of  $\alpha$  on the mean of GNL distribution: as the value of  $\alpha$  increases, the mean decreases. However the mean will increase with a rising  $\beta$ .

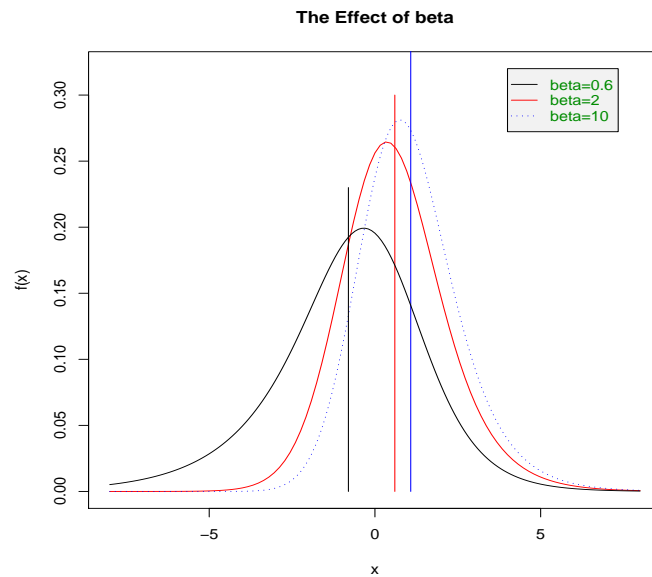


Figure 2.9: The pdf curves and means of three GNL distributions with  $\mu=0$ ,  $\sigma=1$ ,  $\alpha=1$  and  $\rho=1.2$ , when  $\beta=2$  (red),  $\beta=10$  (blue, dotted) and  $\beta=0.6$  (black)

The nature of the tails can be determined by the order of the poles of its characteristic (or moment generating) function (Reed, 2004), which are given in table 2.5.1

**Table 2.5.1** Description for the tails of the GNL distribution

<i>Limit</i>	pdf
$x \rightarrow \infty$	$f(x) \sim c_1 x^{\rho-1} e^{-\alpha x}$
$x \rightarrow -\infty$	$f(x) \sim c_2 (-x)^{\rho-1} e^{\beta x}$

where  $c_1$  and  $c_2$  are constants. The parameter  $\rho$  controls the thickness of the tails. For  $\rho < 1$ , both tails are fatter than the corresponding exponential distribution; for  $\rho = 1$ , they are like exponential tails; and for  $\rho > 1$ , they are thinner than those of exponential.

The GNL distributions are closed under linear transformation: *i.e.* if  $X \sim \text{GNL}(\mu, \sigma^2, \alpha, \beta, \rho)$ , then  $a + bX \sim \text{GNL}(b\mu + a/\rho, b^2\sigma^2, \alpha/b, \beta/b, \rho)$ , where  $a$  and  $b$  are constants.

When  $\mu = \sigma^2 = 0$ , the GNL distribution has ch.f

$$\left[ \frac{\alpha\beta}{(\alpha - is)(\beta + is)} \right]^\rho \quad (2.26)$$

which is that of the *generalized Laplace* distribution (2.18)

If  $\rho = 1$ , the GNL distribution reduces to the *normal Laplace* (NL) distribution with four parameters.

## 2.6 Numerical determination of the pdf and cdf of the GNL

In this section, we present three methods of numerically determining the pdf and cdf of the GNL distributions.

### 2.6.1 Using the Representation as a Convolution

The characteristic function (2.19) can be written

$$\phi_{GNL}(s) = \exp(\rho\mu is - \rho\sigma^2 s^2/2) \left[ \frac{\alpha}{\alpha - is} \right]^\rho \left[ \frac{\beta}{\beta + is} \right]^\rho \quad (2.27)$$

This is the product of the characteristic function of the normal distribution with parameters  $\mu$  and  $\sigma^2$  and that of generalized Laplace distribution (2.16). It follows that the GNL distribution is that of the convolution of normal  $N(\rho\mu, \rho\sigma^2)$  and  $GL(\alpha, \beta, \rho)$  distributions

$$X \stackrel{d}{=} W + U, \quad W \sim N(\rho\mu, \rho\sigma^2), \quad U \sim GL(\alpha, \beta, \rho) \quad (2.28)$$

Furthermore  $(\frac{\theta}{\theta - is})^\rho$  is the ch.f of a gamma random variable with shape parameter  $\rho$  and scale parameter  $\frac{1}{\theta}$ . Thus the last two terms of (2.27) are the ch.fs of (i) a gamma random variable with shape parameter  $\rho$  and scale parameter  $\frac{1}{\alpha}$  and (ii) the negative of a gamma random variable with shape parameter  $\rho$  and scale parameter  $\frac{1}{\beta}$ . It follows from (2.27) that a GNL random variable,  $X \sim GNL(\mu, \sigma^2, \alpha, \beta, \rho)$  can be represented as a convolution

$$X \stackrel{d}{=} \rho\mu + \sigma\sqrt{\rho}Z + \frac{1}{\alpha}G_1 - \frac{1}{\beta}G_2 \quad (2.29)$$

where  $Z, G_1$  and  $G_2$  are independent with  $Z \sim N(0,1)$  and  $G_1, G_2$  are gamma random variables with scale parameter  $\theta = 1$  and shape parameter  $\rho$ . *i.e* with pdf given by

$$\gamma(u) = \frac{1}{\Gamma(\rho)} u^{\rho-1} e^{-u}$$

Closed-form expressions for the pdf and cdf of the family of GNL distributions have not been found except when  $\rho = 1$ . However, the pdf (and cdf) can be obtained numerically using the convolution (2.28) as (2.29) to represent the pdf of the GNL

$$f(x) = \int_{-\infty}^{\infty} f_W(x - u) f_U(u) du \quad (2.30)$$

and the cdf of GNL is

$$F(x) = \int_{-\infty}^{\infty} F_W(x-u)f_U(u)du \quad (2.31)$$

where  $f_U(u)$  is the pdf of generalized Laplace distribution (2.18), and  $f_W(w)$  and  $F_W(w)$  are the pdf and cdf of a normal distribution with mean  $\rho\mu$  and variance  $\rho\sigma^2$ . The integrals (2.30) and (2.31) can be evaluated numerically to obtain the pdf and cdf of the GNL distribution.

### 2.6.2 Numerical Inversion of Characteristic Function

The ch.f of GNL (2.27) can be expressed

$$\phi_{GNL}(s) = r(s) \exp(i\theta(s)) \quad (2.32)$$

where  $r(s)$  and  $\theta(s)$  are the modulus and argument of the ch.f of the random variables of (2.27).

The ch.f can be inverted to obtain the pdf of GNL (see e.g. Knight and Satchell, 2001, p.285)

$$\begin{aligned} f_{GNL}(x) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-isx} \phi(s) ds \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} r(s) e^{i(\theta(s)-sx)} ds \\ &= \frac{1}{\pi} \int_0^{\infty} r(s) (\cos(\theta(s) - sx) + i \sin(\theta(s) - sx)) ds \\ &= \frac{1}{\pi} \int_0^{\infty} r(s) (\cos(\theta(s) - sx)) ds \end{aligned} \quad (2.33)$$

using the fact that  $f(x)$  is a real, so that the imaginary part of the integral will be zero.

The cdf of GNL can be obtained by inversion of the ch.f as (Shephard, 1991)

$$F_{GNL}(x) = \frac{1}{2} + \frac{1}{2\pi} \int_0^\infty \frac{e^{isx}\phi(-s) - e^{-isx}\phi(s)}{is} ds \quad (2.34)$$

Since  $\phi(s) = r(s)e^{i\theta(s)}$  and  $e^{isx}\phi(-s) - e^{-isx}\phi(s) = i2r(s)\sin(sx - \theta(s))$ , the cdf of GNL is

$$F_{GNL}(x) = \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \frac{r(s)}{s} \sin(sx - \theta(s)) ds \quad (2.35)$$

The integrals (2.33) and (2.35) can be evaluated numerically to obtain the pdf and cdf of GNL.

### 2.6.3 Using the representation as a Normal mean-variance mixture

The Normal variance-mean mixture representation of the GNL derives from it being the distribution at the state of the Brownian motion  $dx = \nu dt + \tau dw$  with initial state  $x_0 \sim N(\mu_0, \sigma_0^2)$  observed at a random time  $T$  independent of the Brownian motion with  $T \sim \text{Gamma}(\lambda, \rho)$ . By re-scaling time it is also the state of the Brownian motion  $dx = \frac{\nu}{\lambda} dt + \frac{\tau}{\sqrt{\lambda}} dw$  at time  $T' = \lambda T \sim \text{Gamma}(1, \rho)$ . For a fixed time  $t$  the state of this latter Brownian motion is

$$X(t) \sim N\left(\mu + \frac{\nu}{\lambda}t, \sigma^2 + \frac{\tau^2}{\lambda}t\right)$$

so that the state after gamma-distributed time is a “mean-variance” mixture of normal distributions with mixing parameter  $t$ .

Re-parameterizing, letting  $\frac{1}{\alpha} - \frac{1}{\beta} = \frac{\nu}{\lambda}$  and  $\frac{2}{\alpha\beta} = \frac{\tau^2}{\lambda}$  one gets

$$X(t) \sim N\left(\mu + \left(\frac{1}{\alpha} - \frac{1}{\beta}\right)t, \sigma^2 + \frac{2}{\alpha\beta}t\right)$$

so that, the pdf of GNL is

$$f_{GNL}(x) = \int_0^\infty \frac{1}{\sqrt{2\pi(\sigma^2 + \frac{2}{\alpha\beta}t)}} \exp\left(-\frac{(x - (\mu + (\frac{1}{\alpha} - \frac{1}{\beta})t))^2}{2(\sigma^2 + \frac{2}{\alpha\beta}t)}\right) g(t) dt \quad (2.36)$$

where  $g(t) = \frac{1}{\Gamma(\rho)} t^{\rho-1} e^{-t}$  is the pdf of a gamma distribution with scale-parameter 1 and shape parameter  $\rho$ . This integral can be evaluated numerically.

## Chapter 3

# Methods of Estimation

We consider estimation for both grouped and ungrouped data. In the applications the ungrouped data come from the logarithmic returns of stock prices, and the grouped data from household incomes data.

### Ungrouped Data

To fit the GNL model to ungrouped data (e.g. logarithmic returns), one can estimate model parameters using Maximum Likelihood Estimation (MLE), or the Method of Moments Estimation (MME). One could also consider Bayesian methods, but these will not be discussed in this thesis.

### 3.1 Method of Moments

Although the Method of Moments Estimation (MME) is less efficient than MLE, it is usually computationally simpler. MME is performed by solving a set of equations obtained by equating population moments to sample moments. The  $k^{th}$ -moment of a random variable is  $E(X^k)$ , and this must be expressed as a function of model parameters. The  $k^{th}$  sample moment is  $m_k = \frac{1}{n} \sum_{i=1}^n X_i^k$ . The MME estimates for i.i.d observation  $x_1, \dots, x_n$  are obtained by solving (for the parameters  $\underline{\theta}$ ) following system of equations

$$E_{\underline{\theta}}(X^k) = \frac{1}{n} \sum_{i=1}^n X_i^k \quad k = 1, \dots, p. \quad (3.1)$$

where  $\underline{\theta}$  is a  $p$ -vector of parameters.

The cumulants of a distribution are defined in terms of the cumulant generating

function (cgf)

$$K_X(t) = \log M_X(t) \quad (3.2)$$

where  $M_X(t)$  is the moment generating function

$$M_X(t) = E(e^{tx}) \quad (3.3)$$

The  $j^{\text{th}}$  cumulant  $\kappa_j$  is the coefficient of  $t^j/j!$  in the Taylor series expansion of  $K_x(t)$  i.e

$$\kappa_j = \left. \frac{d^j K_x(t)}{dt^j} \right|_{t=0} \quad (3.4)$$

Since there is a one-to-one relationship between cumulants and moments one can find method of moments estimates of parameters by solving simultaneously the  $p$  equations resulting from setting the first  $p$  cumulants of the distribution equal to their sample equivalents. i.e

$$\kappa_j(\theta) = k_j \quad j = 1, \dots, p \quad (3.5)$$

where  $k_j$  is the coefficient of  $t^j/j!$  the Taylor series expansion of the sample cgf i.e of

$$k_x(t) = \log \left[ \frac{1}{n} \sum_{i=1}^n e^{tx_i} \right] \quad (3.6)$$

i.e

$$k_j = \left. \frac{d^j k_x(t)}{dt^j} \right|_{t=0} \quad (3.7)$$

Sample cumulants are related to sample moments in the same way as population cumulants and moments are related. Precisely

$$\begin{aligned} k_1 &= m_1 \\ k_2 &= m_2 - m_1^2 \\ k_3 &= 2m_1^3 - 3m_1m_2 + m_3 \\ k_4 &= -6m_1^4 + 12m_1^2m_2 - 3m_2^2 - 4m_1m_3 + m_4 \\ k_5 &= 24m_1^5 - 60m_1^3m_2 + 20m_1^2m_3 - 10m_2m_3 + 5m_1(6m_2^2 - m_4) + m_5 \end{aligned} \quad (3.8)$$

where  $m_j$  is the  $j^{th}$  sample moment. The first few sample cumulants can thus be readily computed from the sample moments.

Reed (2004) determined the cumulants of the GNL as

$$\begin{aligned}\kappa_1 &= \rho\left(\mu + \frac{1}{\alpha} - \frac{1}{\beta}\right) \\ \kappa_2 &= \rho\left(\sigma^2 + \frac{1}{\alpha^2} + \frac{1}{\beta^2}\right) \\ \kappa_r &= \rho(r-1)!\left(\frac{1}{\alpha^r} + (-1)^r \frac{1}{\beta^r}\right) \quad r = 3, 4, \dots\end{aligned}\quad (3.9)$$

To find MMEs of the five parameters of the GNL thus involves solving for  $(\mu, \sigma^2, \alpha, \beta, \rho)$  simultaneously the five equations  $\kappa_1=k_1, \kappa_2=k_2, \dots, \kappa_5=k_5$ . With some simple algebra (Reed, 2004) this can be reduced to solving (for  $\alpha, \beta$ ) the pair of equations

$$12k_3(\alpha^{-5} - \beta^{-5}) = k_5(\alpha^{-3} - \beta^{-3}), \quad 4k_4(\alpha^{-5} - \beta^{-5}) = k_5(\alpha^{-4} + \beta^{-4}) \quad (3.10)$$

from which the corresponding solution values of the other parameters can be obtained as

$$\hat{\rho} = \frac{k_3}{2!} \frac{1}{\hat{\alpha}^{-3} - \hat{\beta}^{-3}}; \quad \hat{\sigma}^2 = \frac{k_2}{\hat{\rho}} - \hat{\alpha}^{-2} - \hat{\beta}^{-2} \quad \text{and} \quad \hat{\mu} = \frac{k_1}{\hat{\rho}} - \hat{\alpha}^{-1} + \hat{\beta}^{-1}$$

In the special case of a symmetric GNL distribution ( $\alpha = \beta$ ), the estimates of the four parameters by method of moments, can be found analytically. The equation (3.9) gives

$$\begin{aligned}\kappa_1 &= \rho\mu \\ \kappa_2 &= \rho\left(\sigma^2 + \frac{2}{\alpha^2}\right) \\ \kappa_3 &= 0 \\ \kappa_4 &= 3!\rho\left(\frac{2}{\alpha^4}\right) \\ \kappa_6 &= 5!\rho\left(\frac{2}{\alpha^6}\right)\end{aligned}\quad (3.11)$$

with higher odd-order cumulants all zero, i.e  $\kappa_3=\kappa_5=\kappa_7= \dots =0$

From the above system, the estimates are (Reed, 2004)

$$\hat{\alpha} = \hat{\beta} = \sqrt{20 \frac{k_4}{k_6}} ; \quad \hat{\rho} = \frac{100}{3} \frac{k_4^3}{k_6^2} ; \quad \hat{\sigma}^2 = \frac{k_2}{\hat{\rho}} - \frac{2}{\hat{\alpha}^2} \quad \text{and} \quad \hat{\mu} = \frac{k_1}{\hat{\rho}}$$

The parameter space for the GNL  $(\mu, \sigma^2, \alpha, \beta, \rho)$  distribution is  $\Re \otimes \Re_+^4$ . i.e the four-parameters  $\sigma^2, \alpha, \beta, \rho$  are constrained to be positive, while  $\mu$  can be any real number. This can cause problems for the method of moments, because sometimes the solution to the moment equations will fall outside of the parameter space. i.e result in an estimate in which some of  $\sigma^2, \alpha, \beta$  and  $\rho$  are negative.

Method of moments estimation can also be applied to the ordinary NL distribution (2.11) for which the third and fourth order cumulants are

$$\kappa_3 = 2\alpha^{-3} - 2\beta^{-3}; \quad \kappa_4 = 6\alpha^{-4} + 6\beta^{-4} \quad (3.12)$$

MMEs for  $\alpha$  and  $\beta$  can be found by solving numerically the pair of equations  $k_3 = \kappa_3$  and  $k_4 = \kappa_4$  and the corresponding estimates of  $\mu$  and  $\sigma$  can then be found as

$$\hat{\mu} = k_1 - \frac{1}{\hat{\alpha}} + \frac{1}{\hat{\beta}} \quad \text{and} \quad \hat{\sigma}^2 = k_2 - \left( \frac{1}{\hat{\alpha}^2} + \frac{1}{\hat{\beta}^2} \right)$$

## 3.2 Maximum Likelihood Estimation

Maximum likelihood estimation (MLE) is the “gold-standard” method for obtaining parameter estimates. The likelihood function is a mathematical expression obtained as an arbitrary constant times the probability of observing the given data regarded as a function of model parameters  $\underline{\theta}$ . Maximum likelihood (ML) estimates are obtained by maximizing the function with respect to  $\underline{\theta}$ . It is usually more convenient to maximize the log-likelihood function with respect to the model parameters. For independent identically distributed observations, the likelihood is the product of

the probability density (or mass) function  $f(x; \underline{\theta})$  evaluated at each of the observed data value. i.e

$$L(\underline{\theta}) = \prod_i f(x_i; \underline{\theta}) \quad (3.13)$$

and the log-likelihood is

$$\ell(\underline{\theta}) = \sum_i \log f(x_i; \underline{\theta}) \quad (3.14)$$

There is no closed-form expression for the probability density function (pdf) of the GNL distribution so one cannot obtain a closed-form for the likelihood or log-likelihood. One can however evaluate it numerically for given values of  $\underline{\theta}$  (and given data). In this thesis we consider three methods of performing this numerical calculation. They are (see section 2.6)

- (a) Convolution of normal and generalized Laplace pdfs
- (b) Inversion of the characteristic function
- (c) Using the representation of the GNL distribution as a normal mean-variance mixture.

To evaluate the log-likelihood for a single value of the parameters  $\underline{\theta}$ ,  $n$  numerical integrations (using method (a), (b) or (c)) must be conducted (where  $n$  is number of observations in the sample).

To find MLEs involves numerically maximizing the log-likelihood function. This has been performed using the R function `optim` in the `stats` package.

For the ordinary NL distribution a closed form of the pdf and hence of the log-likelihood exist. Precisely for independent observations  $y_1, y_2, \dots, y_n$  from  $NL(\mu, \sigma^2,$

$\alpha, \beta$ ) the log likelihood function is

$$\ell = n \log \alpha + n \log \beta - n \log(\alpha + \beta) + \sum_{i=1}^n \log[R(p_i) + R(q_i)] \quad (3.15)$$

where  $p_i = \alpha\sigma - (y_i - \mu)/\sigma$  and  $q_i = \beta\sigma + (y_i - \mu)/\sigma$ , and  $R$  is the *Mills' ratio* (2.6) of the complementary cumulative distribution function (cdf) to the pdf of a standard normal distribution.

This can be maximized analytically over  $\mu$  to obtain

$$\hat{\mu} = \bar{y} - \frac{1}{\alpha} + \frac{1}{\beta} \quad (3.16)$$

and a profile likelihood

$$\begin{aligned} \hat{\ell}(\alpha, \beta, \sigma^2) = & n \log \alpha + n \log \beta - n \log(\alpha + \beta) + \sum_{i=1}^n \phi\left(\frac{y_i - \bar{y} + 1/\alpha - 1/\beta}{\sigma^2}\right) + \\ & \sum_{i=1}^n \log\left[R\left(\alpha\sigma^2 - \frac{y_i - \bar{y} + 1/\alpha - 1/\beta}{\sigma^2}\right) + R\left(\beta\sigma^2 + \frac{y_i - \bar{y} + 1/\alpha - 1/\beta}{\sigma^2}\right)\right] \end{aligned} \quad (3.17)$$

This must be maximized numerically (e.g using the function `optim` in R) to obtain maximum likelihood estimates of parameters. Another approach is to use the EM-algorithm (Reed and Jorgensen 2004), although there seems to be little to be gained in terms of computation time.

### Measurement of the performance

Quantile-quantile (Q-Q) plots provide a way of visually assessing the fit a distribution to ungrouped data. In a Q-Q plot, if the resulting points lie roughly on the line of slope 1, then the compared distribution fits the data well. Q-Q plots are obtained by plotting the quantiles of the data of the empirical distribution against the theoretical quantiles using MLEs of the parameters. The empirical quantiles are just the sorted observations. The theoretical quantile  $Q_i$  corresponding to the  $i^{th}$  ordered observation is obtained by solving

$$F_{GNL}(Q_i) = p_i$$

where  $p_i = (i - 0.5)/n$ , therefore

$$Q_i = F_{GNL}^{-1}(p_i) \quad (3.18)$$

Unfortunately no closed-form exists for the inverse of the c.d.f of the GNL distribution, so equation (3.18) has to be solved numerically.

### Grouped data

We now consider the grouped data with boundaries  $0 < x_1 < x_2 \dots$ . Since grouped income data available from the Luxembourg Income Study (<http://www.lisproject.org>) are in the form of percentiles of the distribution, we consider the likelihood for such data is proportional to the joint distribution of the order statistics corresponding to the empirical percentiles. For example, if  $x_{(1)}, x_{(2)}, \dots, x_{(19)}$  correspond to 5th, 10th, ..., 95th percentiles of a sample of size  $N$ , then the log-likelihood is of the form

$$\ell(\underline{\theta}) = \sum_{i=1}^{19} \log f(\log x_{(i)}) + \frac{N}{20} \sum_{i=1}^{20} \log(P_i(\underline{\theta})) + C \quad (3.19)$$

where  $P_i(\underline{\theta}) = F(\log(x_{(i)}); \underline{\theta}) - F(\log(x_{(i-1)}); \underline{\theta})$ ;  $F()$  denotes the cumulative distribution function;  $\underline{\theta}$  is the parameter vector;  $x_{(i)}$  and  $x_{(i-1)}$  are the upper and lower bounds of the  $i^{th}$  of 20 data groups, and  $N$  is the total number of observations. Typically  $N$  will be a very large number, and the first part of the summation is relatively much smaller than that of second part. In previous studies of fitting income distribution it has been ignored, with simply the multinomial log likelihood (where all frequencies =  $N/20$ )

$$(N/20) \sum_{i=1}^{20} \log(P_i(\underline{\theta})) \quad (3.20)$$

being maximized.

### Goodness-of-Fit for grouped data

The sum of squared errors (SSE), sum of absolute errors (SAE), and chi-square ( $\chi^2$ ) goodness-of-fit statistic and the maximized log-likelihood are four measures used in previous studies to compare the fit of parametric income distribution models. The SSE, SAE and  $\chi^2$  are defined as

$$\text{SSE} = \sum_{i=1}^N \left\{ \frac{n_i}{N} - P_i(\hat{\theta}) \right\}^2 \quad (3.21)$$

$$\text{SAE} = \sum_{i=1}^N \left| \frac{n_i}{N} - P_i(\hat{\theta}) \right| \quad (3.22)$$

$$\chi^2 = N \sum_{i=1}^N \left[ \left\{ \frac{n_i}{N} - P_i(\hat{\theta}) \right\}^2 / P_i(\hat{\theta}) \right] \quad (3.23)$$

where  $\hat{\theta}$  denotes the estimated parameters. Note that one would not use the  $\chi^2$  statistic based on proportions to test for goodness of fit. We use this form of it simply to make comparisons with results of Bandourian *et al.* (2002), who used the  $\chi^2$  statistic in this form. In this thesis we compare the fit of the four-parameter NL with the best four-parameter fit obtained to date, that of the generalized beta (GB2) distribution (McDonald, 1984); and compare the five-parameter GNL with best five-parameter model obtained to date, that of the GB (McDonald and Xu, 1995).

### 3.3 Nelder-Mead Method: Multi-dimensional Maximization Method

The R function `optim` includes several methods of optimization. The Nelder-Mead method, which does not require derivatives, has been used in this thesis. The method

is simple, intuitive and relatively stable in approaching the optimum and can be applied to discontinuous problems. It is based on evaluating a function at the vertices of a simplex, then iteratively shrinking the simplex as better points are found and repeat the process until some desired bounds are obtained (Nelder and Mead, 1965). Ideally the optimum will not depend on the starting values. However if the likelihood possesses more than one local maximum the point to which the algorithm converges may depend on the starting value. To check whether this is the case optimization was run several times using different starting values.

## Chapter 4

### Simulation Studies for The GNL distribution

One way to examine the performance of an estimation procedure is to apply it to simulated data with a known distribution. In this chapter, we utilize GNL simulation to assess the results of estimation. In addition, we fit a generalized hyperbolic (GH) distribution to GNL simulated data and fit the GNL distribution to simulated GH data.

#### 4.1 Simulating GNL Data

Pseudo random variables from a GNL distribution can be simulated from (2.29) directly. This involves simulating random variables from the three independent distributions, namely the standard normal  $Z$  and two gamma distributions  $G_1, G_2$  with scale parameter  $\theta = 1$  and shape parameter  $\rho$ . A GNL random variable  $X \sim \text{GNL}(\mu, \sigma^2, \alpha, \beta, \rho)$  is then obtained as

$$X = \rho\mu + \sigma\sqrt{\rho}Z + \frac{1}{\alpha}G_1 - \frac{1}{\beta}G_2$$

Fitting to ungrouped data sometimes resulted in difficulties with multiple local maxima, with very similar values. As an alternative the data were grouped and the model fitted to grouped data. This seemed to eliminate difficulties with multiple maxima. Further research is needed to investigate the problem with multiple maxima. The process of the simulation was as follows:

- Generate 1,000 artificial GNL distributed observations, for given parameter values *e.g* generate 1000 (i.d) GNL (0.1, 0.4, 0.2, 0.3, 0.2) deviates.

- Group the observations into 20 equal-frequency intervals.
- Estimate the parameters using the MLE estimation method for grouped data.
- Compare results from the Q-Q plots.

The following table shows estimates obtained via numerical maximization for different starting values of the optimization routine. In addition, Q-Q plots of sample and fitted theoretical quantiles are given. In all cases shown the parameter values used in the simulation were  $\mu=0.1$ ,  $\sigma^2=0.4^2$ ,  $\alpha=0.2$ ,  $\beta=0.3$  and  $\rho=0.2$ .

**Table 4.1** GNL (using MLE) fitted to the simulated GNL data

GNL(starting values)	Max $\ell$	$\mu$	$\sigma^2$	$\alpha$	$\beta$	$\rho$
(0.1,0.4,0.9,1,0.2)	3021.452	0.0266	0.0379	0.2344	0.3441	0.2122
(-0.3,0.4,1,0.5,0.6)	3021.452	0.0266	0.0379	0.2344	0.3441	0.2122
(0.1,0.6,0.1,0.1,0.7)	3021.452	0.0266	0.0379	0.2344	0.3441	0.2122
(-1,0.9,0.3,0.4,0.5)	3021.452	0.0266	0.0379	0.2344	0.3441	0.2122

Table 4.1 indicates that the routine converges to the same maximum for different starting values. In addition, the MLEs are fairly close to the true values. The standard errors of the estimators are:  $\hat{\mu}$  (0.0423),  $\hat{\sigma}^2$  (0.0255),  $\hat{\alpha}$  (0.0248),  $\hat{\beta}$  (0.0275),  $\hat{\rho}$  (0.0161). Furthermore, the Q-Q plot in the Figure 4.1 shows a satisfactory fit. This and other similar Q-Q plots using simulated data were used as a reference for assessing the degree of deviation from a straight line that could be expected.

For some other choices of parameter values, e.g. GNL (1, 1, 4, 3, 2), multiple maxima resulted

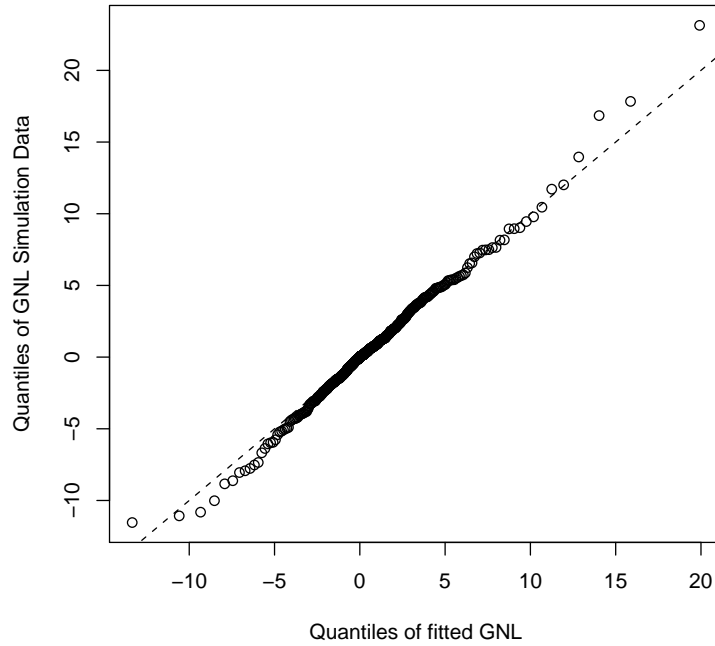


Figure 4.1: Q-Q plots for the simulated GNL data

**Table 4.2** GNL (using MLE) fitted to the simulated GNL data

GNL(starting values)	Max $\ell$	$\mu$	$\sigma^2$	$\alpha$	$\beta$	$\rho$
(1, 4, 5, 1, 2)	3013.791	1.5023	2.9406	17.1636	12.2965	1.5536
(-3, 4, 4, 2, 3)	3013.791	1.7452	3.4799	37.0888	37.3776	1.3167
(1, 2, 3, 6, 3)	3013.791	1.9770	3.9200	24.8859	20.0404	1.1682
(-2, 9, 3, 4, 0.5)	3013.791	1.9231	3.8741	28.8934	67.7220	1.1828

Table 4.2 indicates that there are not any set of estimates close to the true values (1, 1, 4, 3, 2). However, all of the max  $\ell$  are the same suggesting either multiple maxima or a very flat likelihood function. Also, the Q-Q plots from Figure 4.2 show

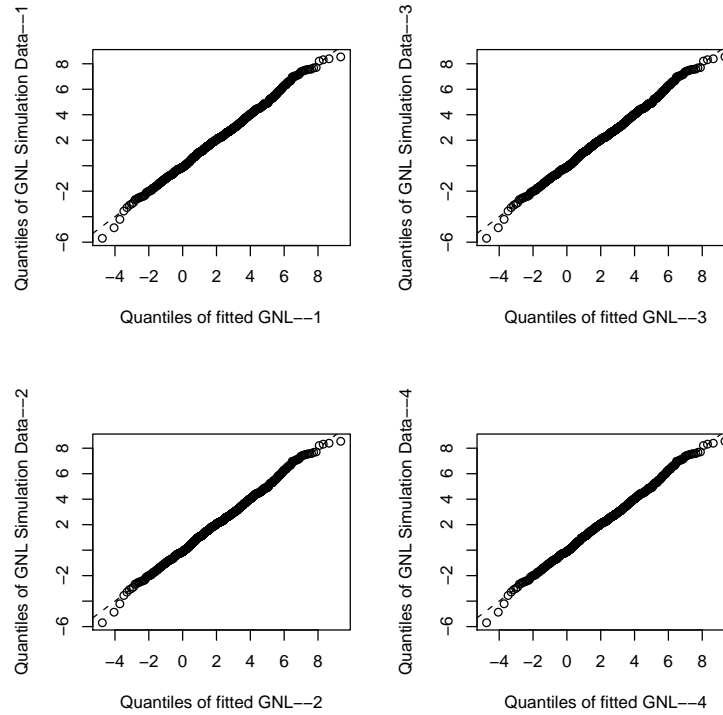


Figure 4.2: Q-Q plots for the simulated GNL data

a satisfactory fit.

One explanation for multiple maxima may be as follows. When  $\rho$  (or  $\alpha$  and  $\beta$ ) are large the GNL distribution is close to a normal. When  $\rho$  (or  $\alpha$  and  $\beta$ )  $\rightarrow \infty$ ,  $\text{GNL} \rightarrow$  normal, so for large  $\rho$  there is virtually no information about the tail parameters  $\alpha$  and  $\beta$ , leading to large variances in their estimates. Also  $\hat{\rho}$  will be confounded with  $\hat{\mu}$  and  $\hat{\sigma}^2$ . For large  $\alpha, \beta$  the GNL will also be close to normal and in this case there will be little information about them and also  $\hat{\rho}$  will be confounded with  $\hat{\mu}$  and  $\hat{\sigma}^2$ . Thus in such cases one would expect a manifold in parameter space over which the likelihood changes very little. The different local maxima to which the optimization routine converged could be explained as small deviations from the flat likelihood sur-

face, caused by numerical effect (roundoff and other numerical error).

## 4.2 Simulating GH Data

A common method for simulating generalized hyperbolic variables is to use the normal variance-mean mixture structure. Using the generalized inverse Gaussian (GIG) Rydberg(2000) (sec 1.2) as the mixing distribution. The algorithm is as follows

- Sample  $Y$  from  $GIG(\lambda, \chi, \psi)$ .
- Sample  $Z$  from  $N(0, 1)$  standard normal.
- Return  $X = \mu + \beta Y + \sqrt{Y}Z$ .

Simulation from the GIG-distribution is not straightforward. There are two different algorithms (Atkinson (1982) and Dagpunar (1989)) that have been used. Both of the algorithms have been implemented by Dr. David Scott in the R package. ([http : //www.stat.auckland.ac.nz/ dscott/](http://www.stat.auckland.ac.nz/~dscott/))

We can compare how well the GNL distribution fits to simulated GH data and vice versa. Applying the above method, we generate 1000 observations from GH with the parameters satisfying the following conditions:  $\alpha > 0$ ,  $0 < |\beta| < \alpha$ ,  $\mu \in \mathfrak{R}$ ,  $\lambda \in \mathfrak{R}$  and  $\delta > 0$ , *i.e.* GH(1.2 , 3, 1, 1, 1). We then fit the GNL.

Table 4.3 GNL (using MME) fitted to the simulated GH data

GNL	likelihood-value	$\mu$	$\sigma^2$	$\alpha$	$\beta$	$\rho$
Estimators	1172.531	0.5371	0.0462	1.9866	4.6477	1.9377

Table 4.4 GH (using MLE) fitted to the simulated GNL data

GH	$\max \ell$	$\lambda$	$\alpha$	$\beta$	$\delta$	$\mu$
Estimators	2138.59	-0.3898	1.1378	-0.0151	4.7486	-0.0160

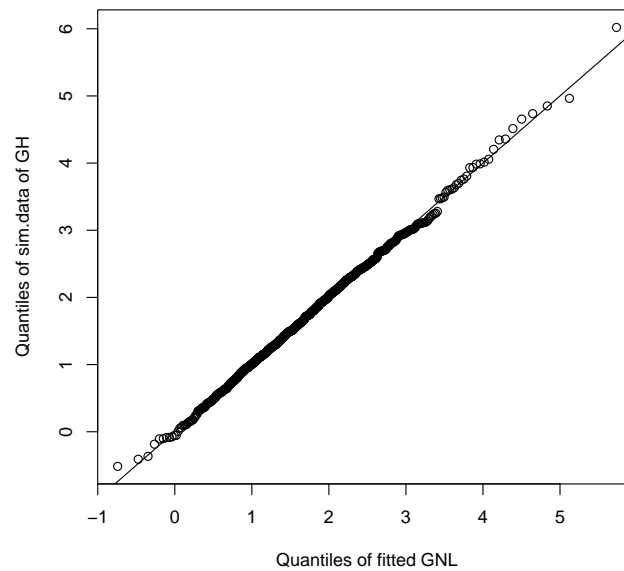


Figure 4.3: Q-Q plot for the GNL

Fig. 4.3 is the quantile-quantile plot for the quantiles of simulated GH data fitted and the quantiles of the GNL distribution. From the figure, we can see that most of the quantile points fall approximately along the 45-degree line demonstrating the GNL fitting the GH data very well.

Next we generate 1000 observations from  $\text{GNL}(0, 1.5, 1, 1, 1)$  and fit the GH Fig. 4.4 shows a Q-Q plot of the fit.

From this very limited study it seems these two five-parameter distributions can be similar in shape at least for the parameter values chosen.

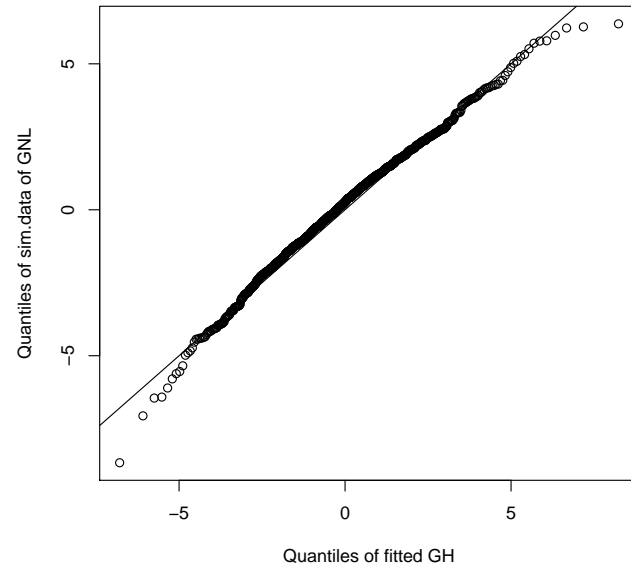


Figure 4.4: Q-Q plot for the GH

Further research, both analytical and numerical, needs to be carried out to explore the similarities and differences between the GNL and GH distributions.

## Chapter 5

### Application to Income Data

This chapter will discuss the application of NL and GNL models and their comparison against other four and five-parameter models when fitting to grouped income data. Currently, the generalized beta (GB2) distribution (McDonald, 1984) and the generalized beta or GB (McDonald and Xu, 1995) are being claimed as the best fitting four and five-parameter models for income data (Bandourian *et al.*, 2002). Consequently, we compare the performance of the NL with GB2 and of the GNL and GB.

#### 5.1 Description of the Income Data

The household income data are obtained from the Luxembourg Income Study (LIS) database, including 30 countries from four continents: Europe, Asia, America and Oceania, and spanning up to 30 years. The term income distribution spans many kinds of ‘income’ (e.g. earnings is one type of income). Following the LIS definitions ‘earnings’ is the total of gross wages and salaries; total gross income is the sum of earnings, factor income and market income, and disposable income is the difference between total gross income and total tax. Only positive earnings data (see table 5.2) are used in this thesis. Hence, if an observation value is equal to or smaller than zero, it is rejected. Due to privacy laws, income data is considered confidential information; thus, only grouped income data is available. The samples involve twenty-equal probability intervals, corresponding to the 5<sup>th</sup> through 95<sup>th</sup> percentiles obtained from the LIS database using a SAS program.

All datasets are in local currency amounts. The following table (Table 5.1) lists the countries, dates and summary statistics.

**Table 5.1** Statistics descriptions of the data

<i>Variable</i>	No.of Observations	Mean	Std Dev	Skewness	Kurtosis
Australia 1994	4619	45039.15	31412.61	2.19	11.76
Belgium 1997	1640	1662629.77	1047288.54	1.87	8.73
Canada 1987	8667	33810.93	23842.05	1.40	4.67
Canada 1997	25108	47965.58	37066.56	3.41	53.78
Italy 2000	5276	42810.01	33969.87	6.35	99.34
Mexico 2000	9239	60601.89	84932.97	7.99	132.79
Taiwan 2000	12301	857641.73	590643.72	2.57	18.79
UK 1999	15199	28180.91	28921.34	10.97	267.73
USA 1997	39815	49412.87	50520.11	3.94	24.07

**Table 5.2.**<sup>a</sup> Definitions of Income Variables

Define	VARIABLE DEFINITION
+	Gross wages and Salaries
+	Farm self-employment income
+	Non-farm self-employment income
=	<b>Total Earning</b>
+	Cash property incom
+	Private occupational pensions
+	Public sector pensions
+	Social Retirement benefits
+	Child or family allowances
+	Unemployment compensation
+	Sick pay; Accident pay; Disability pay; Maternity pay
+	Military/vet/war benefits
+	Other social insurance
+	Means-tested cash benefits
+	Near-cash benefits
+	Alimony or Child Support
+	Other regular private income and Other cash income
=	<b>Total Gross Income</b>
-	Mandatory employee contribution for sled-employed
-	Mandatory employee contribution
-	Income tax
=	<b>Disposable Income</b>

<sup>a</sup>Luxembourg Income Study website <http://www.lisproject.org>

## 5.2 Income Distribution Results

The following tables contain the goodness-of-fit statistics (the sum of squared errors (SSE), the sum of absolute errors (SAE), the Pearson goodness-of-fit statistic  $\chi^2$  and the maximized log-likelihood) for the two four-parameter distributions (NL and GB2), and for the two five-parameter distributions (GNL and GB).

### Australia 1994

Dist'n	SSE	SAE	$\chi^2$	$\max \ell$
NL	0.0019	0.1341	151.97	-13914.7
GB2	0.0024	0.1437	195.24	-13936.2
GNL	0.0007	0.0982	66.4	-13869.90
GB	0.0023	0.1353	192.1	-13934.76

### Belgium 1997

Dist'n	SSE	SAE	$\chi^2$	$\max \ell$
NL	0.0008	0.1027	47.89	-8729.5
GB2	0.0009	0.1078	54.11	-8732.8
GNL	0.0005	0.0787	29.1	-8720.01
GB	0.0009	0.1096	53.0	-8732.24

**Canada 1987**

Dist'n	SSE	SAE	$\chi^2$	$\max \ell$
NL	0.0007	0.0865	128.90	-26026.4
GB2	0.0010	0.1037	191.17	-26056.6
GNL	0.0004	0.0612	66.8	-25996.47
GB	0.0010	0.1027	189.6	-26055.89

**Canada 1997**

Dist'n	SSE	SAE	$\chi^2$	$\max \ell$
NL	0.0002	0.0600	118.51	-75275.9
GB2	0.0005	0.0791	228.46	-75330.8
GNL	0.0002	0.04647	84.3	-75258.68
GB	0.0004	0.0757	219.5	-75326.30

**Italy 2000**

Dist'n	SSE	SAE	$\chi^2$	$\max \ell$
NL	0.0044	0.2345	586.97	-16063.4
GB2	0.0046	0.2418	628.74	-16077.1
GNL	0.0042	0.2233	557.0	-16051.21
GB	0.0046	0.2394	624.7	-16076.45

**Mexico 2000**

Dist'n	SSE	SAE	$\chi^2$	$\max l$
NL	0.0006	0.0778	102.51	-27997.7
GB2	0.0007	0.0892	127.12	-28010.1
GNL	0.0003	0.0630	68.2	-27980.66
GB	0.0007	0.0837	123.7	-28008.31

**Taiwan 2000**

Dist'n	SSE	SAE	$\chi^2$	$\max l$
NL	0.0022	0.1270	457.71	-37084.8
GB2	0.0024	0.1408	495.64	-37104.3
GNL	0.0002	0.0484	50.0	-36875.50
GB	0.0023	0.1322	470.4	-37091.40

**UK 99**

Dist'n	SSE	SAE	$\chi^2$	$\max l$
NL	0.0002	0.0491	64.270	-45564.9
GB2	0.0005	0.0696	133.07	-45599.8
GNL	0.0001	0.0406	42.11	-45553.29
GB	0.0004	0.0685	128.9	-45597.59

## US 97

Dist'n	SSE	SAE	$\chi^2$	$\max \ell$
NL	0.0003	0.0566	266.60	-119406.2
GB2	0.0005	0.0738	418.61	-119480.1
GNL	0.0003	0.0548	218.47	-119384.2
GB	0.0005	0.0719	414.63	-119478.2

In some cases there were multiple local maxima to the log-likelihood function (e.g. fitting the GNL to Australia 1994, Belgium 1997, Canada 1987, Canada 1997, Taiwan 2000 and UK 1999). However, they resulted in identical values for the maximized log-likelihood.

While the numerical results for GB2 and GB are comparable to those of Bandourian *et al.* (2002), there are some slight differences. These differences may be explained by one or more of the following reasons: (i) end-point differences - we used 0 and  $\infty$  for the lower and upper boundaries of the smallest and largest class; (ii) differences in retrieval of data from LIS - we used SAS 9.13; (iii) possibility of multiple local maxima of the likelihood function for GB2 and GB.

We will discuss the results of NL and GB2 distributions first and subsequently, those of GNL and GB distributions. As can be seen from the tables, in all cases the four-parameter NL has better performance than GB2 does; and indeed is better than the five-parameter GB. The GNL performs notably better than all of the other distributions. One thing to be emphasized is the nine cases reported are the only ones to which model fitting was done. They were not selected because the NL and GNL fitted well. We expect that if the models are fitted to other datasets of Bandourian *et al.*(2002), the results would be similar.

## Chapter 6

# Application To Financial Data

### 6.1 Description of the Stock Price Data

Empirical studies seem to suggest that the distribution of logarithmic returns tend to be heavy-tailed. The logarithmic return of a stock is defined as

$$r_t = \log(P_{t+1}) - \log(P_t) = \log\left(\frac{P_{t+1}}{P_t}\right) \quad (6.1)$$

where  $P_t$  is the price of a financial security at time  $t$ . To assess the fit of the GNL distribution to empirical logarithmic returns, and to compare the fit with the GH, which is now widely used in derivative-pricing models (Eberlein and Hammerstein, 2002) both distributions were fitted to sets of financial data obtained from the website <http://finance.yahoo.com/>. Two sets of data IBM and CitiGroup stock price were used using a 2-year period, 3 January, 2005 to 29 December, 2006. This resulted in 503 observations in each time series. The adjusted closing price (close price adjusted for dividends and splits) was used to calculate the logarithmic returns. As discussed in section 4.1, for large scale parameter  $\rho$  of the GNL there is virtually no information about the tail parameters  $\alpha$  and  $\beta$  and also  $\hat{\rho}$  will be confounded with  $\hat{\mu}$  and  $\hat{\sigma}^2$ . In such cases, the likelihood changes very little with the different parameters. Therefore, MME would be used for the estimation of GNL in this chapter.

### 6.2 Results of fitting GNL and GH

The following tables show the estimates of GNL and GH when fitted to the log returns of IBM and CitiGroup stock price, 2005-2006.

The GNL (using MME) fitted to IBM and CitiGroup 05-06

GNL	likelihood value	$\mu$	$\sigma^2$	$\alpha$	$\beta$	$\rho$
IBM	-1628.16	0.0098	0.0008	77.4011	45.0120	0.0716
CitiGroup	-1708.61	0.0110	0.0010	90.1476	73.0707	0.0527

The GH fitted to IBM and CitiGroup 05-06

GH	$\max \ell$	$\lambda$	$\alpha$	$\beta$	$\delta$	$\mu$
IBM	-1632.486	-2.2675	0.0717	0.0717	0.1583	$2.57 * 10^{-5}$
CitiGroup	-1711.712	-2.3839	30.4370	11.5155	0.0142	-0.00033

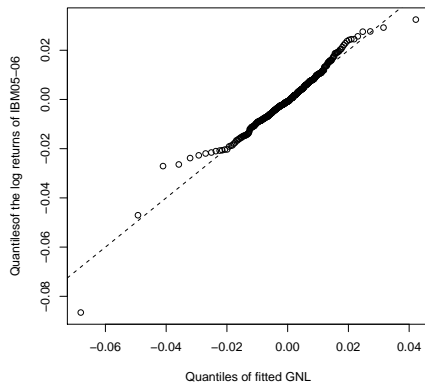


Figure 6.1: The GNL fitted to IBM 05-06

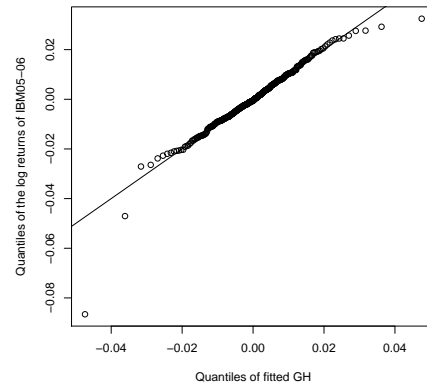


Figure 6.2: The GH fitted to IBM 05-06

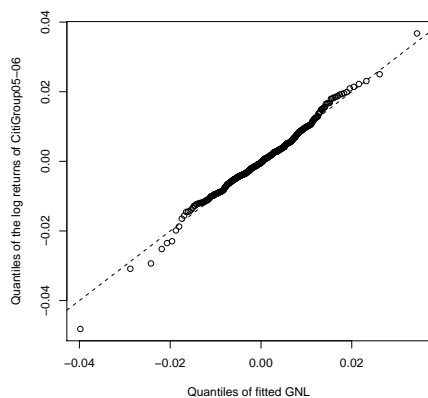


Figure 6.3: The GNL fitted to CitiGroup  
05-06

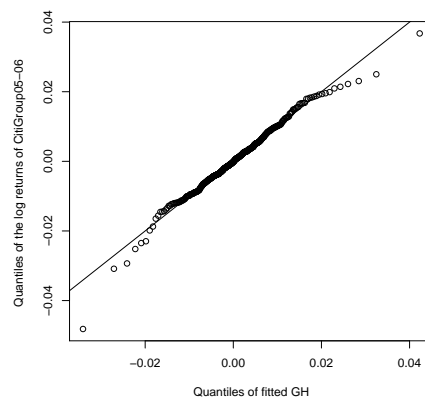


Figure 6.4: The GH fitted to CitiGroup  
05-06

Q-Q plots for fitting the GNL and GH to IBM 2005-2006 indicate that neither fit very well in the lower tail; For the CitiGroup 2005-2006, the Q-Q plot for the GNL appears more satisfactory than for the GH. Also, calculated after estimating by MME, the log-likelihood values for IBM and Citigroup 2005-2006 for the GNL distribution are smaller than the maximum log-likelihood values for the GH. As known, the MLE is a better estimation than the MME. If we can solve the MLE multiple maxima problem for the GNL distribution, then we should get smaller maximum log-likelihood value which it would be an even better result than the GH. While the fits of neither the GH nor the GNL are perfect, the results suggest that the fit of the GNL is at least as good as for the more well-known GH.

## Chapter 7

### Conclusions

Two families of heavy-tailed distributions: the normal-Laplace (NL) and its relative distribution the generalized normal-Laplace (GNL) distribution are the focus of this thesis. We have summarized the major properties of the GNL and NL distributions. The thesis gives prominence to parameter estimation for the five-parameter GNL and four-parameter NL distributions to grouped income data by maximum likelihood. The results of fitting four theoretical distributions to nine income datasets clearly indicate that the NL fits better than what has previously been claimed as the best-fitting four-parameter model for income distribution namely the GB2 (McDonald, 1984). Indeed it fits better than what has been previously claimed as the best fitting five-parameter distribution the GB (McDonald and Xu, 1995). The five-parameter GNL distribution provides an even better fit.

Even though there are no closed form expressions for the pdf and cdf of the GNL, they can be obtained numerically from the convolution of GL and Normal distributions. However, ML estimation for the GNL is not always satisfactory, there sometimes being problems with multiple maxima. Furthermore it requires considerable computing time.

The similarity of the GNL and GH distribution was demonstrated using simulation in which in turn the GNL and GH were fitted to samples generated from the other. Also the two models were fitted to two-year log returns of IBM and CitiGroup stock prices and the results indicate that GNL and GH distributions provide similar

good fits to the data. These results suggest that the GNL may provide a model at least as good as the more widely known GH for fitting to financial data. Option pricing for European call options for assets following Brownian-Laplace motion, whose increments have a GNL distribution, has been developed by Reed (2007). More work in this area could lead to the GNL providing a viable alternative to the GH in option-pricing theory. However some difficulties remain mainly to do with maximum likelihood estimation of the GNL parameters.

As we have discussed in this thesis, since there is no closed-form for the pdf or cdf of the GNL distribution, evaluating the likelihood function involves numerical integration for each data point whichever of three methods is used. This makes numerical maximization of the likelihood function rather slow. In addition there appear to be problems with flat likelihoods and multiple maximum, especially when the parameters  $\alpha$ ,  $\beta$  and  $\rho$  are large. The application in this thesis is a beginning of research in fitting the GNL and its family. We have some other areas for further research:

- More work on ML estimation of the GNL distribution for ungrouped data.
- Using methods such as the EM algorithm for fitting GNL distribution.
- Using analytic and numerical methods for finding similarities and differences between the GNL and GH distributions.
- The application of the GNL to credit risk, risk management and portfolio optimization problems as has been conducted using the GH distribution.

## References

- Abramowitz, M. and Stegun, I.A., 1972.** Handbook of Mathematical Functions. Dover Publications, New York.
- Aitchinson, J., and Brown, J.A.C., 1969.** The lognormal distribution with special reference to its uses in economics, Cambridge University Press.
- Ammon, O., 1895.** Die Gesellschaftsordnung und ihre Natürlichen Grundlagen, Jena.
- Amoroso, L. , 1924-1925.** Ricerche Intorno alla Curva dei Redditi, Annali di Matematica. Pura ed Applicata, Series 4-21.2:123-157.
- Atkinson, A. C., 1982.** The simulation of generalized inverse gaussian and hyperbolic random variables. SIAM Journal of Scientific & Statistical Computing, 3:502-515.
- Bandourian, R., McDonald, J. B. and Turley, R. S., 2002.** A comparison of parametric models of income distribution across countries and over time. Luxembourg Income Study Working Paper, No. 305.
- Barndorff-Nielsen, O., 1977.** Exponentially decreasing distributions for the logarithm of particle size, Proceedings of the Royal Society of London A 353 (1977), 401-419.
- Barndorff-Nielsen, O., Kent J. and Srensen, M., 1982.** Normal Variance-Mean Mixtures and  $z$  Distributions. International Statistical Review, 50, 145-159.
- Blisild, P and Srensen, M.K., 1992.** 'Hyp' - A computer program for analyzing by means of the hyperbolic distribution. Department of Theoretical Statistics, Research Report No.300.

**Bartels, C.P.A. and van Metelen, H., 1975.** Alternative Probability Density Functions of Income, Vrije University Amsterdam: Research memorandum 29, 30 pages.

**Burr, I. W., 1942.** Cumulative frequency functions. *Ann. Math. Statist.* 13, 215-32.

**Carr, P., H. Geman, D. Madan and M. Yor, 2002.** The Fine Structure of Asset Returns: An Empirical Investigation, *Journal of Business*, 75,305-332.

**Chhikara, R. S. and Folks, J. L., 1989.** The Inverse Gaussian Distribution, New York: Dekker.

**Kleiber, Carlos, 1996.** Dagum vs. Singh-Maddala income distributions. *Economics Letters*, 53, 265-268.

**Dastrup, S.R., McDonald, J. B. and Hartshorn, Rachel, 2005.** The Impact of Taxes and Transfer Payments on the Distribution of Income: a Parametric Comparison. Luxembourg Income Study Working Paper, No. 401.

**Dagum, C., 1977.** A new model for personal income distribution: specification and estimation. *Economie Appliquée*, 30, 413-437.

**Dagum, C., 1980.** The generation and distribution of income, the Lorenz curve and the Gini ratio. *Economie Appliquée*, 33, 327-367.

**Dagpunar, J. S., 1989.** An easily implemented generalized inverse gaussian generator. *Communications in Statistics, Simulations* 18:703C71

**Eberlein, E. and Hammerstein, Ernst August v., 2002.** Generalized Hyperbolic and Inverse Gaussian Distributions: Limiting Cases and Approximation of Processes, University of Freiburg.

**Eberlein, E. and U. Keller, 1995.** Hyperbolic distributions in finance. *Bernoulli*, 1, 281-299.

- Gibrat, R., 1931.** Les Inegalites Economiques. Sirely, Paris.
- Cvitanic, Jaksa and Zapatero, Fernando, 2004.** Introduction to the Economics and Mathematics of Financial Markets. Cambridge Massachusetts: MIT Press.
- Kleiber, Christian, 1996.** Dagum vs. Singh-Maddala income distributions. Economics Letters, 53, 265-268.
- Knight, John and Satchell, Stephen, 2001.** Return Distributions in Finance, 285-293.
- Kotz, S., T. J. Kozubowski and K. Podgórski, 2001.** The Laplace Distribution and Generalizations. Birkhäuser, Boston.
- Madan, Dilip, and Seneta, Eugene, 1990.** The V.G. Model for Share Returns. Journal of Business, 63, 511-524.
- McDonald, J.B. and Ransom, Michael R., 1979.** Functional Forms, Estimation Techniques and the Distribution of Income. Econometrica, Vol 47, 1513-1526.
- McDonald, J.B., 1984.** Some generalized functions for the size distribution of income. Econometrica, 52, 647-663.
- McDonald, J.B. and Xu, Y. J., 1995.** A generalization of the beta distribution with applications. Journal of Econometrics, 66, 133-152.
- Nelder, J. A. and Mead, R., 1965.** A Simplex Method for Function Minimization. Computer Journal, 7, 308-313.
- Pareto, V. 1895.** "La Legge della Domanda", Giornale degli Economisti, Gennaio:59-68.
- Reed, W. J., 2003.** The Pareto law of incomes - an explanation and an extension. Physica A, 319, 579-597.

**Reed, W. J., 2004.** The Normal-Laplace Distribution and its Relatives. In Order Statistics and Inference, N.Balakrishna et al. (eds) Birkhauser.

**Reed, W. J., 2007.** Brownian-Laplace motion and its use in financial modelling. Commun. Stat. - Theory & Methods, 36, 473:484.

**Reed, W. J. and M. Jorgensen, 2004.** The double Pareto-lognormal distribution - A new parametric model for size distribution. Commun. Stat. - Theory & Methods, 33, 1733-1753

**Rydberg, Tina Hviid, 2000.** Realistic Statistical Modelling of Financial Data. J. Nuffld College, Oxford, United Kingdom.

**Seshadri, V., 1993.** The Inverse Gaussian Distribution. Clarendon Press.

**Singh, S.K, and G.S, Maddala, 1976.** A function for the size distribution of income. Econometrica, 44, 963-970.

**Shephard, N.G., 1991.** From Characteristic Functions to Distribution Function: A Simple Framework For The Theory. Cambridge University Press.

**Schoutens, W. and Teugels, J.L. 1998.** Levy processes, polynomials and martingales. Communications in Statistics - Stochastic Models, 14, 335-349.

**Taille, C., Patil, G.P and Balderssari, B., 1981.** Ordering within the Generalized Gamma Family of Income Distributions. Statistical Distributions in Scientific Work, Vol 6.

**Thurow, L.C., 1970.** Analyzing the American Income Distribution. American Economics Association, 60, 261-269

**Waller, Lance A., Turnbull, Bruce W. and Hardin, J. Michael, 1995.** Obtaining Distribution Functions by Numerical Inversion of Characteristic Functions With Application. The American Statistician, Vol 49., No.4.

**Wolff, Sebastian, 2004.** A local and globalized, constrained and simple bounded Nelder-Mead method.

# Appendix

## 1. SAS Code for Obtaining the LIS Income Data

```
* user =
* password =
* package =
* project =
%LET cc = CA ;
%LET yy = 97 ;
%MACRO Prep ;
  DATA prep          (DROP=d5 hweight);
  SET &&&cc.&&yy.h (KEEP=country hweight d4 d5 earning);
  * avoid double counting of household records ;
  IF earning <= 0    THEN DELETE;
  IF d5    =      3    THEN DELETE;
  * set equivalence scale as square root of number of persons;
  ey      =    earning  ;
  * create person weight as hweight times number of persons;
  wt      = hweight*d4 ;
RUN;
* Get the median dpi & the mean equivalised income ;
PROC UNIVARIATE DATA=prep NOPRINT;
  VAR    ey earning ;
  WEIGHT wt ;
  OUTPUT OUT=temp MEAN=aveey MEDIAN=medey meearning ;
RUN ;
DATA _NULL_;
  SET temp;
  CALL SYMPUT("m",meearning);
  CALL SYMPUT("a",aveey);
RUN;
DATA start (DROP= botlin toplin);
  SET prep  ;
  * bottomcoding at one percent equivalized mean;
  botlin = 0.01 * &a;
  IF ey  < botlin THEN ey = botlin ;
  * topcoding at ten times unequivalized median ;
  toplin = 10 * &m;
  IF earning > toplin THEN ey = (toplin) ;
RUN ;
PROC SORT DATA=start;
  BY country ey ;
RUN ;
%MEND Prep ;
```

```

%MACRO quantile ;
PROC UNIVARIATE DATA=start NOPRINT;
  BY    country ;
  VAR   ey      ;
  WEIGHT wt      ;
  OUTPUT OUT     = tmp
         PCTLPTS = 5 to 95 by 5
         PCTLPRE = dec ;
RUN ;
DATA   &cc.&yy._2 (KEEP=country d5 d10 d15 d20 d25 d30 d35 d40 d45) ;
MERGE start (KEEP=country)
      tmp ;
      d5  = dec5 ;
      d10 = dec10;
      d15 = dec15;
      d20 = dec20;
      d25 = dec25;
      d30 = dec30;
      d35 = dec35;
      d40 = dec40;
      d45 = dec45;
      IF _N_ = 1 ;
RUN;
%MEND quantile ;
%MACRO tabulate ;
      PROC TABULATE DATA=&dataset
                    FORMAT=&formgen
                    FORMCHAR=&bordure
                    NOSEPS;
          CLASS &classe ;
          VAR &variable ;
          TABLE &table / CONDENSE
                 PRINTMISS
                 BOX="&box"
                 RTS=&rts ;
          KEYLABEL &lab ;
      RUN ;
%MEND tabulate ;
%MACRO show ;
DATA   &cc.&yy._KF ;
MERGE &cc.&yy._2;
  BY country      ;
ATTRIB
  d5  label='P5' format=10.4

```

```

d10 label='P10' format=10.4
d15 label='P15' format=10.4
d20 label='P20' format=10.4
d25 label='P25' format=10.4
d30 label='P30' format=10.4
d35 label='P35' format=10.4
d40 label='P40' format=10.4
d45 label='P45' format=10.4 ;

RUN ;

%LET dataset= &cc.&yy._2 ;
%LET formgen = 15.4 ;
%LET bordure = '|---|+|---|+|---|+|---|---|+|---|+|---|+|---|' ;
%LET classe = country ;
%LET variable = d5 d10 d15 d20 d25 d30 d35 d40 d45 ;
%LET table = country, d5 d10 d15 d20 d25 d30 d35 d40 d45;
%LET box = 'Percentile ' ;
%LET rts = 40 ;
%LET lab = sum = '' ;
%TABULATE
%MEND show ;
&&&cc.&yy.f ;
%PREP
%QUANTILE
%SHOW

```

## R code for numerically computing the pdf of GNL

### 2. Using the Representation as a Convolution

#### # PDF of gen. Laplace

```
GL.pdf<-function (x,alpha,beta,tau)
{# PDF of gen. laplace in terms of my parameters (Kotz et al., p. 189)
  l.box1<-((tau/2+1/4)*log(alpha*beta)+(beta-alpha)*x/2)
  l.box1<-l.box1-lgamma(tau)-log(pi)/2 -(tau/2-1/4)*log(2)
  box1<-exp(l.box1)
  l.box2<-((tau-1/2)*(log(2*alpha*beta)/2 +log(abs(x))-log(alpha+beta))
  l.box3<-log(besselK(abs(x)*(alpha+beta)/2,abs(tau-1/2),expon.scaled =TRUE))-abs(x)*(alpha+beta)/2
box2.3<-ifelse(tau>0.5,(1/2)*gamma(tau-1/2)*(4*sqrt(2*alpha*beta)*(alpha+beta)^(-2))^(tau-1/2)
,gamma(0.5-tau)*2^(-2*tau)*(sqrt(2*alpha*beta))^(tau-0.5)*(abs(x))^(2*tau-1))
  box0<-box1*box2.3
  out<-ifelse(x==0,box0,exp(l.box1+l.box2+l.box3))
  ifelse(is.finite(out)==TRUE,out,1e8)}
```

#### # CDF of gen Laplace

```
GL.cdf<-function (x,alpha,beta,tau)
{
int.fn<-function (y)
  { supply(y,FUN=GL.pdf,alpha=alpha,beta=beta,tau=tau) }
temp1<-ifelse(x<0,integrate(int.fn,low=-Inf,upp=x,rel.tol =1e-10,abs.tol =
1e-10)$value,integrate(int.fn,low=-Inf,upp=0)$value+integrate(int.fn,low=0,upp=x,rel.tol =1e-10,abs.tol =
1e-10)$value)
  temp1 }
```

#### # CDF of GNL Convolution Form

```
GNL.cdf<-function (x,mu,sigma2,alpha,beta,rho)
{
# Computes cdf of GNL by convolution of GL and normal
  sigma<-sqrt(sigma2)
  temp.fn<-function(t){
    out<-GL.pdf(t,alpha,beta,rho)*pnorm(x-t,mu*rho,sqrt(sigma2*rho))
    out }
  int.fn<-function(y) supply(y,FUN=temp.fn)
  temp1<-integrate(int.fn,lower=-Inf,upper=0,stop.on.error=F,rel.tol =1e-10, abs.tol =
1e-10,subdivisions=100000)
  temp2<-integrate(int.fn,lower=0,upper=Inf,stop.on.error=F,rel.tol =1e-10, abs.tol =
1e-10,subdivisions=100000)
  temp1$value + temp2$value
}
```

**# PDF of GNL Convolution Form**

```

GNL.pdf<-function(x,mu,sigma2,alpha,beta,rho)
{# Computes pdf of GNL by convolution of GL and normal
  sigma<-sqrt(sigma2)
  temp.fn<-function(t){
    out<-GL.pdf(t,alpha,beta,rho)*dnorm(x-t,mu*rho,sqrt(sigma2*rho))
    out      }
  int.fn<-function(y) sapply(y,FUN=temp.fn)
  temp1<-integrate(int.fn,lower=-Inf,upper=0,stop.on.error=F,rel.tol      =1e-8,      abs.tol      =
1e-8,subdivisions=100000)
  temp2<-integrate(int.fn,lower=0,upper=Inf,stop.on.error=F,rel.tol      =1e-8,      abs.tol      =
1e-8,subdivisions=100000)
  temp1<-ifelse(temp1$message=="OK",temp1$value,1e4)
  temp2<-ifelse(temp2$message=="OK",temp2$value,1e4)
  temp1 + temp2}

```

**#Income log-Likelyhood Function##**

```

llik.GNL.grouped<-function(par,groupdata)
{# Log-lik. for fitting GNL to " Grouped data"
  interval<-log(groupdata[,1])
  freq<-groupdata[,2]
  mu<-par[1]
  sigma2<-par[2]
  alpha<-par[3]
  beta<-par[4]
  rho<-par[5]
  prob<-numeric(length(interval))
  prob[1]<-GNL.cdf(interval[1],mu=mu,sigma2=sigma2,alpha=alpha,beta=beta,rho=rho)-GNL.cdf(0,mu=mu,sigm
a2=sigma2,alpha=alpha,beta=beta,rho=rho)
  for(i in 2:length(interval)){
    if(i ==length(interval)) prob[length(interval)]<-1-sum(prob[1:length(interval)-1])
    else
  prob[i]<-GNL.cdf(interval[i],mu=mu,sigma2=sigma2,alpha=alpha,beta=beta,rho=rho)-GNL.cdf(interval[i-1],mu
=mu,sigma2=sigma2,alpha=alpha,beta=beta,rho=rho)
  }
  sum(log(prob)*freq)}
obj.GNL.grouped<-function(par,groupdata)
  {-llik.GNL.grouped(par,groupdata) }
obj.GNL.grouped.new<-function(par,groupdata)
  {      # sigma^2, alpha, beta, rho >0, mu real #
  par.new<-c(par[1],par[2]^2,par[3]^2,par[4]^2,par[5]^2)
  obj.GNL.grouped(par.new,groupdata)
  }

```

### 3. Inversion of Characteristic Function and Normal Mean-variance Mixture

#### #PDF and CDF of the GNL: Numerical Inversion of Characteristic Function

```
GNL.pdf.in<-function (x,mu,sigma2,alpha,beta,rho){
  y<-x-rho*mu
  cf.fn<-function(s){
    cplex<-complex(1,0,1)
    temp1<-alpha*beta*exp(-sigma2*s^2/2)
    temp2<-(alpha-cplex*s)*(beta+cplex*s)
    out<-(temp1/temp2)^rho
    out}
  temp.fn<-function(s){
    (Mod(cf.fn(s))*cos(Arg(cf.fn(s))-s*y))}
  int.fn<-function(t){sapply(t,FUN=temp.fn)}
  te<-integrate(int.fn,lower=0,upper=Inf,rel.tol=1e-10,subdivisions=1000000)
  temp3<-ifelse(te$message=="OK", te$value/pi,NA)
  temp3
}

GNL.cdf.in<-function (x,mu,sigma2,alpha,beta,rho){
  y<-x-rho*mu
  #c.f. for mu=0 GNL
  cf.fn<-function(s){
    cplex<-complex(1,0,1)
    temp1<-alpha*beta*exp(-sigma2*s^2/2)
    temp2<-(alpha-cplex*s)*(beta+cplex*s)
    out<-(temp1/temp2)^rho
    out}
  temp.fn<-function(s){
    (Mod(cf.fn(s))/(s*pi))*sin(s*y-Arg(cf.fn(s)))}
  int.fn<-function(t){sapply(t,FUN=temp.fn)}
  te<-integrate(int.fn,lower=0,upper=Inf,rel.tol=1e-10,subdivisions=1000000)
  temp3<-ifelse(te$message=="OK", te$value,1e4)
  temp3+0.5
}
```

#### #PDF for the GNL: Normal mean-variance mixture

```
GNL.pdf.NV<-function (x,mu,sigma2,alpha,beta,rho){
  temp.fn<-function(t0){
    out<-dnorm(x,mu+(1/alpha-1/beta)*t0,sqrt(sigma2+
t0*(2/(alpha*beta))),log=T)+dgamma(t0,shape=rho,rate=1,scale=1,log=T)
    exp(out)}
  int.fn<-function(y)
sapply(y,FUN=temp.fn)
```

```
temp1<-integrate(int.fn,lower=0,upper=Inf,rel.tol=1e-8,abs.tol=1e-8,subdivisions=100000)
temp1$value }
```

### **#Ungrouped Data- log-Likelyhood Function##**

```
llik.GNL<-function(par,data)
{# Log-lik. for fitting GNL to "data"
mu<-par[1]
sigma2<-par[2]
alpha<-par[3]
beta<-par[4]
rho<-par[5]
temp<-sapply(data,FUN=GNL.pdf,mu=mu,sigma2=sigma2,alpha=alpha,beta=beta,rho=rho)
sum(log(temp))}
obj.GNL<-function(par,data)
{ -llik.GNL(par,data)}
obj.1.GNL<-function(par,data)
{ # neg. log.lik using pars. sqrt(sigma2),sqrt(alpha) etc.
par.1<-c(par[1],par[2]^2,par[3]^2,par[4]^2,par[5]^2)
obj.GNL(par.1,data)}
```

### **# Equation for computing p-quantile of GNL**

```
quant.GNL.eq<-function(x,p,mu,sigma2,alpha,beta,rho)
{# Equation defining p-quantile of GNL
GNL.cdf(x,mu,sigma2,alpha,beta,rho) - p }
```

#### **##### for one P-quantile only####**

```
quant.1.GNL<- function(p,mu,sigma2,alpha,beta,rho)
{out<-uniroot(quant.GNL.eq,low=-100,upp=100,p=p,mu=mu,sigma2=sigma2,alpha=alpha,beta=beta,rho=rho,tol=1e-9)
ifelse(out$root>-5,out$root,0)
}
```

#### **##### for a vector of P-quantiles####**

```
quant.all.GNL<-function(p.vr,mu,sigma2,alpha,beta,rho)
{sapply(p.vr,FUN=quant.1.GNL,mu=mu,sigma2=sigma2,alpha=alpha,beta=beta,rho=rho)
}
```

### **#NL cdf##**

```
NL.cdf<-function(x,mu,sigma2,alpha,beta){
sigma<-sqrt(sigma2)
temp1<-pnorm((x-mu)/sigma)
temp2<-(-beta*millsratio(alpha*sigma-(x-mu)/sigma)-alpha*millsratio(beta*sigma+(x-mu)/sigma))/(alpha+beta)
temp3<-dnorm((x-mu)/sigma)*temp2
temp1-temp3}
```