Université de Montréal

The Double Pareto-Lognormal Distribution and its applications in actuarial science and finance

par

Chuan Chuan Zhang

Département de mathématiques et de statistique
Faculté des arts et des sciences

Mémoire présenté à la Faculté des études supérieures
en vue de l'obtention du grade de
Maître ès sciences (M.Sc.)
en Mathématiques

Orientation actuariat

février 2015

Université de Montréal
Faculté des études supérieures

Ce mémoire intitulé

The Double Pareto-Lognormal Distribution and its applications in actuarial science and finance

présenté par

Chuan Chuan Zhang

a été évalué par un jury composé des personnes suivantes :

Martin Bilodeau
(président-rapporteur)

Louis G. Doray
(directeur de recherche)

Maciej Augustyniak
(membre du jury)

Mémoire accepté le:
Le 17 janvier 2015

SOMMAIRE

Le but de ce mémoire de maîtrise est de décrire les propriétés de la loi double Pareto-lognormale, de montrer comment on peut introduire des variables explicatives dans le modèle et de présenter son large potentiel d'applications dans le domaine de la science actuarielle et de la finance.

Tout d'abord, nous donnons la définition de la loi double Pareto-lognormale et présentons certaines de ses propriétés basées sur les travaux de Reed et Jorgensen (2004). Les paramètres peuvent être estimés en utilisant la méthode des moments ou le maximum de vraisemblance. Ensuite, nous ajoutons une variable explicative à notre modèle. La procédure d'estimation des paramètres de ce modèle est également discutée. Troisièmement, des applications numériques de notre modèle sont illustrées et quelques tests statistiques utiles sont effectués.

Mots-clés : Loi normale-Laplace, loi double Pareto-lognormale, estimation du maximum de vraisemblance, transformation de Box-Cox, variables explicatives, test d'ajustement.

SUMMARY

The purpose of this Master's thesis is to describe the double Pareto-lognormal distribution, show how the model can be extended by introducing explanatory variables in the model and present its large potential of applications in actuarial science and finance.

First, we give the definition of the double Pareto-lognormal distribution and present some of its properties based on the work of Reed and Jorgensen (2004). The parameters could be estimated by using the method of moments or maximum likelihood. Next, we add an explanatory variable to our model. The procedure of estimation for this model is also discussed. Finally, some numerical applications of our model are illustrated and some useful statistical tests are conducted.

Keywords: Normal-Laplace distribution, double Pareto-lognormal distribution, maximum likelihood estimation, Box-Cox transformation, explanatory variables, goodness-of-fit test.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to express my deep gratitude to Professor Louis G. Doray, my research supervisor, for his patient guidance, enthusiastic encouragement and useful critiques of this work. My grateful thanks are also extended to all the teachers and staff of the Department of Mathematics and Statistics for their extraordinary work that allowed me to complete my studies. I also wish to thank my parents for their support and encouragement throughout my studies. Finally, I would like to dedicate this thesis to the loving memory of my grandmother, who passed away last summer.

INTRODUCTION

The presence of extreme values in a sample is well-documented in many fields such as in insurance, finance, hydrology and geography, etc. The Gaussian model (normal distribution) might be relevant for the centre of a distribution, but not for the extreme values. Thus, the statistical analysis of extremes was developed for fitting parametric models to samples with extreme events and it is also key to many risk management problems related to insurance, reinsurance and finance. There are two standard parametric distributions in the field of extreme value analysis: one is the generalized extreme value distribution which is designed for a sample of extreme outcomes, and the other one is the generalized Pareto distribution which plays an important role in modeling a sample of excesses over a high threshold (see Reiss and Thomas, 2007).

Reed and Jorgensen (2004) introduce a new distribution named the double Pareto-lognormal distribution which exhibits Paretian (power law) behavior in both tails. This distribution has proved to be very useful in modeling the size distributions of various phenomena possibly with extreme events in a wide range of areas such as economics, finance and casualty and property insurance.

In this thesis, we intend to show how the double Pareto-lognormal distribution is derived and discuss its properties based on Reed and Jorgensen's work; we will then try to extend the model by including explanatory variables and show its potential applications in insurance and finance with examples.

This thesis consists of five chapters.

In Chapter 1, we derive the normal-Laplace distribution based on the definition of Reed (2004) and present its properties. Then the double Pareto-lognormal distribution can be defined as an exponentiated normal-Laplace distributed random variable; its properties such as its moment generating function, cumulative distribution function and hazard rate, are also studied.

Chapter 2 discusses the parametric estimation of the double Pareto-lognormal distribution. The method of moments and maximum likelihood will be employed to estimate the parameters.

We will show how to include explanatory variables into our model in Chapter 3; a transformation will also be used to deal with this explanatory variable. Moreover, we will discuss the procedure to estimate the parameters in our new model with covariates.

In Chapter 4, we apply the double Pareto-lognormal distribution to real financial and insurance data. The first application is to fit the original double Pareto-lognormal distribution to stock price returns; the second one is modeling the Danish fire loss claims with the extended model including an explanatory variable. The chi-square test will be conducted in order to test the goodness-of-fit of our model.

Finally, some conclusions are drawn in Chapter 5.

Chapter 1

THE DOUBLE PARETO-LOGNORMAL DISTRIBUTION

The double Pareto-lognormal (dPlN) distribution is defined as an exponentiated normal-Laplace random variable and provides a useful parametric form for modelling size distributions. In this chapter, the normal-Laplace (NL) distribution which results from convolving independent normally distributed and Laplace distributed components, will be defined and its properties such as its cumulative distribution function, its tail behaviour and its moments will also be presented. Then the density function of the dPlN will be derived from that of the NL distribution, and some of its properties will also be discussed. However, we begin with presenting some important statistical tools.

1.1. Moment generating function and cumulants

Several statistical tools are indispensable to our analysis; in this section, we will introduce three such useful statistical concepts: the moment generating function, the characteristic function and the cumulants of a distribution.

Definition 1.1.1. *The moment generating function $M_X(t)$ of a continuous random variable X with density $f(x)$ is defined by*

$$M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x) dx$$

for all real values of t for which the integral converges absolutely.

Note that $M_X(0)$ always exists and is equal to 1.

For a positive integer k, the kth raw moment $E(X^k)$ may be found by evaluating the kth derivative of $M_X(t)$ at 0,

$$\begin{aligned}
M_X^{(k)}(t) &= \frac{d^k}{dt^k} M_X(t) \\
&= \frac{d^k}{dt^k} E(e^{tX}) \\
&= E\left[\frac{d^k(e^{tX})}{dt^k}\right] \\
&= E(X^k e^{tX}),
\end{aligned}$$

which yields

$$M_X^{(k)}(0) = E(X^k) \text{ denoted by } \mu'_k. \tag{1.1.1}$$

The following example shows how to calculate the moment generating function of the normal distribution.

Example 1. *We first compute the moment generating function of a standard normal random variable $Z \sim N(0,1)$ with probability density function (pdf) given by*

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Then

$$\begin{aligned}
M_Z(t) &= E(e^{tZ}) \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tz} e^{-z^2/2} dz \\
&= e^{t^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(z-t)^2/2} dz \\
&= e^{t^2/2}.
\end{aligned}$$

Note that $X = \mu + \sigma Z$ will have a normal distribution with mean μ and variance σ^2 whenever Z follows a standard normal distribution. Hence, the moment *generating function of X is given by*

$$\begin{aligned} M_X(t) &= E(e^{tX}) \\ &= E(e^{t(\mu+\sigma Z)}) \\ &= e^{\mu t}E(e^{t\sigma Z}) \\ &= e^{\mu t}M_Z(\sigma t) \\ &= e^{\mu t}e^{(\sigma t)^2/2} \\ &= \exp(\mu t + \sigma^2 t^2/2). \end{aligned}$$

Theorem 1.1.1. *The moment generating function (mgf) of the sum of independent random variables equals the product of the individual moment generating functions.*

PROOF. Let $X_1, X_2, ..., X_n$ be independent random variables with moment generating functions $M_{X_1}(t), M_{X_2}(t), ..., M_{X_n}(t)$. Then the mgf of $X_1 + X_2 + ... + X_n$ is given by

$$\begin{aligned} M_{X_1+X_2+...+X_n}(t) &= E[e^{t(X_1+X_2+...+X_n)}] \\ &= E(e^{tX_1}e^{tX_2}...e^{tX_n}) \\ &= E(e^{tX_1})E(e^{tX_2})...E(e^{tX_n}) \\ &= M_{X_1}(t)M_{X_2}(t)...M_{X_n}(t) \end{aligned}$$

□

Definition 1.1.2. *(Klugman et al. 2008) The characteristic function of a random variable X is*

$$\phi_X(t) = E(e^{itX}) = E(\cos tX + i \sin tX), \quad -\infty < t < \infty$$

where $i = \sqrt{-1}$.

We introduce the characteristic function because it exists for all distributions, while the moment generating function does not always exist. It is useful to note the relation between these two functions, $\phi_X(t) = M_X(it)$, if $M_X(t)$ exists.

The cumulant generating function $K_X(t)$ is defined as the log of the mgf $M_X(t)$, provided it exists.

Definition 1.1.3. *(Groparu-Cojocaru, 2007) The nth cumulant of a random variable X, denoted κ_n, is defined as the coefficient of the Taylor's series expansion of the cumulant generating function $K_X(t)$ about the origin*

$$K_X(t) = \log M_X(t) = \sum_n \kappa_n t^n/n!. \tag{1.1.2}$$

Obviously, κ_n can be found directly by the nth derivative of $K_X(t)$ at 0, i.e. $\kappa_n = K_X^{(n)}(0)$.

Proposition 1.1.1. *Let X be a random variable, and let us assume that its mgf $M_X(t)$ exists in a neighborhood of 0, then $\kappa_1 = E(X)$ and $\kappa_2 = Var(X)$.*

PROOF.

$$\kappa_1 = \frac{d}{dt}\log M_X(t)\bigg|_{t=0} = \frac{M'_X(t)}{M_X(t)}\bigg|_{t=0} = M'_X(0), \text{ as } M_X(0) = 1.$$

Therefore, $\kappa_1 = E(X)$.

$$Var(X) = E(X^2) - E(X)^2 = M''_X(0) - M'_X(0)^2.$$

Then, by using $M_X(0) = 1$ we have

$$\kappa_2 = \frac{d^2}{dt^2}\log M_X(t)\bigg|_{t=0} = \frac{M''_X(t)M_X(t) - M'_X(t)^2}{M_X(t)^2}\bigg|_{t=0} = M''_X(0) - M'_X(0)^2.$$

So, $\kappa_2 = Var(X)$. □

Kendall and Stuart (1987) showed that the nth cumulant can also be calculated by the first n raw moments. The first five cumulants in terms of raw moments are

$\kappa_1 = \mu'_1$,

$\kappa_2 = \mu'_2 - \mu_1^{'2}$,

$\kappa_3 = \mu'_3 - 3\mu'_2\mu'_1 + 2\mu_1^{'3}$,

$\kappa_4 = \mu'_4 - 4\mu'_3\mu'_1 - 3\mu_2^{'2} + 12\mu'_2\mu_1^{'2} - 6\mu_1^{'4}$,

$\kappa_5 = \mu'_5 - 5\mu'_4\mu'_1 - 10\mu'_3\mu'_2 + 20\mu'_3\mu_1^{'2} + 30\mu_2^{'2}\mu'_1 - 60\mu'_2\mu_1^{'3} + 24\mu_1^{'5}$,

where μ'_n denotes the nth raw moments. We may use this method to find sample cumulants by using sample raw moments.

The cumulants provide an alternative to the moments of the distribution. In some cases, for example for the normal-Laplace distribution, cumulants may be much easier to compute than the moments. We may also use cumulants instead of moments to estimate parameters; this will be discussed in future sections.

Definition 1.1.4. *(Infinitely divisible distribution) The distribution of a real-valued random variable X is infinitely divisible if for every $n \in \mathbb{N}_+$, there exists a sequence of independent, identically distributed variables $(X_1, X_2, ..., X_n)$ such that $X_1 + X_2 + ... + X_n$ has the same distribution as X.*

1.2. The Normal-Laplace distribution

Since most of the results concerning the double Pareto-lognormal distribution are derived using the normal-Laplace distribution, we begin by presenting results for this distribution. In this section, the normal-Laplace distribution is derived from the definition of Reed (2004), and its properties are presented.

1.2.1. Genesis and definitions

Reed (2004) showed that the normal-Laplace distribution (NL) can be defined as the convolution of a normal (N) distribution and an asymmetric Laplace (L) distribution, i.e. $Y \sim NL(\mu, \sigma^2, \alpha, \beta)$ can be represented as

$$Y \stackrel{d}{=} Z + W \tag{1.2.1}$$

where Z and W are independent random variables with $Z \sim N(\mu, \sigma^2)$ and W following an asymmetric Laplace distribution with probability density function (pdf)

$$f_W(w) = \begin{cases} \frac{\alpha\beta}{\alpha+\beta} e^{\beta w}, & \text{for } w \le 0 \\ \frac{\alpha\beta}{\alpha+\beta} e^{-\alpha w}, & \text{for } w > 0. \end{cases} \tag{1.2.2}$$

where $\alpha > 0$ and $\beta > 0$.
The cumulative distribution function (cdf) of W can be easily shown to be

$$F_W(w) = \begin{cases} \frac{\alpha}{\alpha+\beta} e^{\beta w}, & \text{for } w \le 0 \\ 1 - \frac{\beta}{\alpha+\beta} e^{-\alpha w}, & \text{for } w > 0. \end{cases} \tag{1.2.3}$$

Note that if $\alpha = \beta$, W follows a symmetric Laplace distribution with pdf

$$f_W(w) = \begin{cases} \frac{\alpha}{2} e^{\alpha w}, & \text{for } w \le 0 \\ \frac{\alpha}{2} e^{-\alpha w}, & \text{for } w > 0. \end{cases}$$

Its cdf will be

$$F_W(w) = \begin{cases} \frac{1}{2} e^{\alpha w}, & \text{for } w \le 0 \\ 1 - \frac{1}{2} e^{-\alpha w}, & \text{for } w > 0. \end{cases}$$

Definition 1.2.1. *The (convolved or folded) sum of two independent random variables $U = X + Y$ has the probability density $f(u)$ given by the convolution integrals*

$$f(u) = \int_{-\infty}^{+\infty} f_X(x) f_Y(u-x)\, dx = \int_{-\infty}^{+\infty} f_Y(y) f_X(u-y)\, dy, \tag{1.2.4}$$

where X and Y have the probability density functions $f_X(x)$ and $f_Y(y)$ respectively.

The probability density function of Z is

$$f_Z(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2}. \tag{1.2.5}$$

We may derive the pdf of the normal-Laplace distribution from (1.2.2) and (1.2.5) by using the convolution integral (1.2.4), that is,

$$g(y) = \int_{-\infty}^{+\infty} f_Z(z) f_W(y-z)\, dz. \tag{1.2.6}$$

Furthermore, since $f_W(w)$ takes two different forms according to the value of w (see 1.2.2), the pdf of the normal-Laplace $f(y)$ can be obtained by

$$g(y) = g_1(y) + g_2(y) \tag{1.2.7}$$

where $g_1(y)$ and $g_2(y)$ are defined as

$$g_1(y) = \int_{y}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2} \frac{\alpha\beta}{\alpha+\beta} e^{\beta(y-z)}\, dz, \text{ if } y \le z \tag{1.2.8}$$

and

$$g_2(y) = \int_{-\infty}^{y} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2} \frac{\alpha\beta}{\alpha+\beta} e^{-\alpha(y-z)}\, dz, \text{ if } y > z \tag{1.2.9}$$

The calculation of each part of $g(y)$ could be relatively complicated. Reed (2004) managed to express it in terms of normal distribution related functions.

Let us first evaluate the term $g_1(y)$. By multiplying and dividing the pdf of a standard normal distribution, we may write (1.2.8) as

$$\begin{aligned}
g_1(y) &= \frac{\alpha\beta}{\alpha+\beta} \int_{y}^{+\infty} \frac{1}{\sigma} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2} e^{\beta(y-z)}\, dz \\
&= \frac{\alpha\beta}{\alpha+\beta} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} \frac{\int_{y}^{+\infty} \frac{1}{\sigma} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2} e^{\beta(y-z)}\, dz}{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}} \\
&= \frac{\alpha\beta}{\alpha+\beta} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} \frac{\int_{y}^{+\infty} \frac{1}{\sigma} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2 + \beta(y-z)}\, dz}{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}}.
\end{aligned}$$

Multiply the denominator and numerator of the right-hand side by $e^{-\frac{(\mu-\beta\sigma^2)^2}{2\sigma^2}}$, then we obtain

$$g_1(y) = \frac{\alpha\beta}{\alpha+\beta}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}\frac{\int_y^{+\infty}\frac{1}{\sigma}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2-\beta z}e^{-\frac{(\mu-\beta\sigma^2)^2}{2\sigma^2}}dz}{\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2-\beta y}e^{-\frac{(\mu-\beta\sigma^2)^2}{2\sigma^2}}}$$

$$= \frac{\alpha\beta}{\alpha+\beta}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}\frac{\int_y^{+\infty}\frac{1}{\sigma}\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2-2z(\mu-\beta\sigma^2)+(\mu-\beta\sigma^2)^2}{2\sigma^2}}dz}{\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2-2y(\mu-\beta\sigma^2)+(\mu-\beta\sigma^2)^2}{2\sigma^2}}}$$

$$= \frac{\alpha\beta}{\alpha+\beta}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}\frac{\int_y^{+\infty}\frac{1}{\sigma}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left[\frac{z-(\mu-\beta\sigma^2)}{\sigma}\right]^2}dz}{\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left[\frac{y-(\mu-\beta\sigma^2)}{\sigma}\right]^2}}. \qquad (1.2.10)$$

Let $\phi(x)$ denote the pdf of the standard normal distribution $X \sim N(0,1)$

$$\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}.$$

Then

$$\phi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2},$$

and

$$\phi\left[\frac{x-(\mu-\beta\sigma^2)}{\sigma}\right] = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left[\frac{x-(\mu-\beta\sigma^2)}{\sigma}\right]^2}.$$

In addition, let $\Phi\left(\frac{x-\mu}{\sigma}\right)$ denote the cdf of the normal distribution $X \sim N(\mu,\sigma^2)$, therefore the cdf of the normal distribution $X \sim N(\mu-\beta\sigma^2,\sigma^2)$ can be denoted by $\Phi\left[\frac{x-(\mu-\beta\sigma^2)}{\sigma}\right]$, i.e.

$$\Phi\left[\frac{x-(\mu-\beta\sigma^2)}{\sigma}\right] = \int_{-\infty}^{x}\frac{1}{\sigma}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left[\frac{t-(\mu-\beta\sigma^2)}{\sigma}\right]^2}dt.$$

Thus, we may rewrite (1.2.10) as

$$g_1(y) = \frac{\alpha\beta}{\alpha+\beta}\phi\left(\frac{y-\mu}{\sigma}\right)\frac{1-\Phi\left[\frac{y-(\mu-\beta\sigma^2)}{\sigma}\right]}{\phi\left[\frac{y-(\mu-\beta\sigma^2)}{\sigma}\right]}. \qquad (1.2.11)$$

The second term $g_2(y)$ (1.2.9) can be found in a similar way, we have

$$\begin{aligned} g_2(y) &= \int_{-\infty}^{y} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2} \frac{\alpha\beta}{\alpha+\beta} e^{-\alpha(y-z)}\, dz \\ &= \frac{\alpha\beta}{\alpha+\beta} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} \frac{\int_{-\infty}^{y} \frac{1}{\sigma}\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2} e^{-\alpha(y-z)}\, dz}{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}} \\ &= \frac{\alpha\beta}{\alpha+\beta} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} \frac{\int_{-\infty}^{y} \frac{1}{\sigma}\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2-\alpha(y-z)}\, dz}{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}}. \end{aligned}$$

Multiplying the denominator and numerator of the right-hand side by $e^{-\frac{(\mu+\alpha\sigma^2)^2}{2\sigma^2}}$, we obtain

$$\begin{aligned} g_2(y) &= \frac{\alpha\beta}{\alpha+\beta} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} \frac{\int_{-\infty}^{y} \frac{1}{\sigma}\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2+\alpha z} e^{-\frac{(\mu+\alpha\sigma^2)^2}{2\sigma^2}}\, dz}{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2+\alpha y} e^{-\frac{(\mu+\alpha\sigma^2)^2}{2\sigma^2}}} \\ &= \frac{\alpha\beta}{\alpha+\beta} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} \frac{\int_{-\infty}^{y} \frac{1}{\sigma}\frac{1}{\sqrt{2\pi}} e^{-\frac{z^2-2z(\mu+\alpha\sigma^2)+(\mu+\alpha\sigma^2)^2}{2\sigma^2}}\, dz}{\frac{1}{\sqrt{2\pi}} e^{-\frac{y^2-2y(\mu+\alpha\sigma^2)+(\mu+\alpha\sigma^2)^2}{2\sigma^2}}}. \end{aligned}$$

Let $z = -t$, then we have $dz = -dt$.

$$\begin{aligned} g_2(y) &= \frac{\alpha\beta}{\alpha+\beta} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} \frac{\int_{+\infty}^{-y} -\frac{1}{\sigma}\frac{1}{\sqrt{2\pi}} e^{-\frac{(-t)^2+2t(\mu+\alpha\sigma^2)+(\mu+\alpha\sigma^2)^2}{2\sigma^2}}\, dt}{\frac{1}{\sqrt{2\pi}} e^{-\frac{y^2-2y(\mu+\alpha\sigma^2)+(\mu+\alpha\sigma^2)^2}{2\sigma^2}}} \\ &= \frac{\alpha\beta}{\alpha+\beta} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} \frac{\int_{-y}^{+\infty} \frac{1}{\sigma}\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{t+(\mu+\alpha\sigma^2)}{\sigma}\right]^2}\, dt}{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{y-(\mu+\alpha\sigma^2)}{\sigma}\right]^2}} \\ &= \frac{\alpha\beta}{\alpha+\beta} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} \frac{\int_{-y}^{+\infty} \frac{1}{\sigma}\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{t+(\mu+\alpha\sigma^2)}{\sigma}\right]^2}\, dt}{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{(\mu+\alpha\sigma^2)-y}{\sigma}\right]^2}}. \end{aligned} \tag{1.2.12}$$

We may also express (1.2.12) in terms of the cdf and the pdf of a standard normal distribution,

$$g_2(y) = \frac{\alpha\beta}{\alpha+\beta}\phi\left(\frac{y-\mu}{\sigma}\right) \frac{1-\Phi\left[\frac{(\mu+\alpha\sigma^2)-y}{\sigma}\right]}{\phi\left[\frac{(\mu+\alpha\sigma^2)-y}{\sigma}\right]}. \tag{1.2.13}$$

From (1.2.7), the pdf of a normal-Laplace random variable can be obtained by adding the terms (1.2.11) and (1.2.13)

$$g(y) = \frac{\alpha\beta}{\alpha+\beta}\phi\left(\frac{y-\mu}{\sigma}\right)\left\{\frac{1-\Phi\left[\frac{(\mu+\alpha\sigma^2)-y}{\sigma}\right]}{\phi\left[\frac{(\mu+\alpha\sigma^2)-y}{\sigma}\right]} + \frac{1-\Phi\left[\frac{y-(\mu-\beta\sigma^2)}{\sigma}\right]}{\phi\left[\frac{y-(\mu-\beta\sigma^2)}{\sigma}\right]}\right\}. \quad (1.2.14)$$

Moreover, we shall write $Y \sim NL(\alpha, \beta, \mu, \sigma^2)$ to indicate that Y follows a NL distribution.

Reed (2004) proposes to express (1.2.14) by using the *Mills ratio* $R(z)$, which is defined by

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

where Φ^c is the complementary cumulative distribution function of the standard normal random variable. The complementary cumulative distribution function is also called the survival function and denoted by $S(z)$.

Recall that the Mills ratio is also related to the hazard rate $h(z)$ which is defined as

$$h(z) = \frac{f(z)}{S(z)},$$

so that

$$R(z) = \frac{1}{h(z)}.$$

A convenient way to express (1.2.14) in terms of $R(z)$ is

$$g(y) = \frac{\alpha\beta}{\alpha+\beta}\phi\left(\frac{y-\mu}{\sigma}\right)[R(\alpha\sigma-(y-\mu)/\sigma) + R(\beta\sigma+(y-\mu)/\sigma)]. \quad (1.2.15)$$

Alternatively, since an asymmetric Laplace distribution can be represented as a difference between two independent exponential distributions (Kotz et al., 2001), the normal-Laplace distribution can also be derived based on the following decomposition

$$Y \stackrel{d}{=} \mu + \sigma Z + E_1/\alpha - E_2/\beta, \quad (1.2.16)$$

where Z denotes a standard normal random variable, independent of two independent standard exponential random variables, E_1 and E_2, with probability density function $f(x) = e^{-x}$, $x > 0$.

This definition is useful to calculate the moment generating function of the NL distribution or to simulate NL random variables.

1.2.2. Properties

Reed (2004) proved some important properties of the general $NL(\alpha, \beta, \mu, \sigma^2)$ distribution. However, some of them were not shown in detail. We will prove them here.

(1) *Cumulative distribution function*

As the normal-Laplace random variable Y results from the convolution of independent normally distributed Z and Laplace distributed W, i.e. $Y = Z + W$, its cdf can be found in the following way:
Assume that for a particular value of z, the conditional probability of $Y \leq y$ given $Z = z$ can be written as

$$Pr(Y \leq y|Z = z) = Pr(W \leq y - z) = F_W(y - z).$$

Integrating over z, we have

$$Pr(Y \leq y) = \int_{-\infty}^{+\infty} Pr(Y \leq y|Z = z) f_Z(z) dz.$$

Thus, substituting from the previous step, the cdf of Y can be calculated with

$$G_Y(y) = \int_{-\infty}^{+\infty} F_W(y - z) f_Z(z) dz. \tag{1.2.17}$$

Considering the cdf (1.2.3) of the asymmetric Laplace distribution, the cdf of the NL distribution is calculated in two parts, $y \leq z$ and $y > z$:

$$G_Y(y) = G_1(y) + G_2(y),$$

where

$$G_1(y) = \int_{y}^{+\infty} \frac{\alpha}{\alpha + \beta} e^{\beta(y-z)} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2} dz, \text{ if } y \leq z \tag{1.2.18}$$

and

$$G_2(y) = \int_{-\infty}^{y} \left(1 - \frac{\beta}{\alpha + \beta} e^{-\alpha(y-z)}\right) \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2} dz, \text{ if } y > z. \tag{1.2.19}$$

We notice that $G_1(y) = \beta^{-1} g_1(y)$, from (1.2.11) we obtain directly

$$G_1(y) = \frac{\alpha}{\alpha + \beta} \phi\left(\frac{y - \mu}{\sigma}\right) \frac{1 - \Phi\left[\frac{y-(\mu-\beta\sigma^2)}{\sigma}\right]}{\phi\left[\frac{y-(\mu-\beta\sigma^2)}{\sigma}\right]}.$$

If we write it under the Mills ratio form, we get

$$G_1(y) = \frac{\alpha}{\alpha+\beta}\phi\left(\frac{y-\mu}{\sigma}\right) R(\beta\sigma + (y-\mu)/\sigma). \tag{1.2.20}$$

Let us now deal with the second part $F_2(w)$; the first term could be written as the cdf of a normal distribution, that is,

$$\begin{aligned} G_2(y) &= \int_{-\infty}^{y} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2} - \frac{\beta}{\alpha+\beta} e^{-\alpha(y-z)} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2} dz \\ &= \Phi\left(\frac{y-\mu}{\sigma}\right) - \frac{\beta}{\alpha+\beta} \int_{-\infty}^{y} e^{-\alpha(y-z)} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2} dz \\ &= \Phi\left(\frac{y-\mu}{\sigma}\right) - \alpha^{-1} g_2(y). \end{aligned}$$

The second term of the above expression can be simplified by using (1.2.12), so we obtain

$$G_2(y) = \Phi\left(\frac{y-\mu}{\sigma}\right) - \frac{\beta}{\alpha+\beta}\phi\left(\frac{y-\mu}{\sigma}\right) \frac{1-\Phi\left[\frac{(\mu+\alpha\sigma^2)-y}{\sigma}\right]}{\phi\left[\frac{(\mu+\alpha\sigma^2)-y}{\sigma}\right]},$$

and we can also write

$$G_2(y) = \Phi\left(\frac{y-\mu}{\sigma}\right) - \frac{\beta}{\alpha+\beta}\phi\left(\frac{y-\mu}{\sigma}\right) R(\alpha\sigma - (y-\mu)/\sigma). \tag{1.2.21}$$

Combining the two parts (1.2.20 and 1.2.21) of $G_Y(y)$, we obtain the cdf of the normal-Laplace distribution

$$G_Y(y) = \Phi\left(\frac{y-\mu}{\sigma}\right) - \phi\left(\frac{y-\mu}{\sigma}\right) \frac{\beta R(\alpha\sigma - (y-\mu)/\sigma) - \alpha R(\beta\sigma + (y-\mu)/\sigma)}{\alpha+\beta}. \tag{1.2.22}$$

(2) *Two special limiting cases*

As $\alpha \to \infty$, the distribution exhibits a fatter tail than the normal distribution only in the lower tail and the pdf is

$$g(y) = \beta\phi\left(\frac{y-\mu}{\sigma}\right) R(\beta\sigma + (y-\mu)/\sigma).$$

As $\beta \to \infty$, the distribution has a fatter tail than the normal only in the upper tail. The pdf becomes

$$g(y) = \alpha\phi\left(\frac{y-\mu}{\sigma}\right) R(\alpha\sigma - (y-\mu)/\sigma).$$

For example, when $\beta \to \infty$, from (1.2.16) the asymmetric Laplace distribution becomes an exponential distribution with parameter α, so $Y \stackrel{d}{=} \mu + \sigma Z + E_1/\alpha$

which explains why Y has a fatter tail than the normal distribution.

(3) *Moment generating function (mgf)*

From the representation (1.2.1) we may derive the mgf of $NL(\alpha, \beta, \mu, \sigma^2)$ as the product of the mgfs of its normal and Laplace components.

The mgf of $Z \sim N(\mu, \sigma^2)$ is given by

$$M_Z(t) = \exp(\mu t + \sigma^2 t^2/2). \tag{1.2.23}$$

Furthermore, from (1.2.2) the mgf of an asymmetric Laplace distribution can be calculated as

$$\begin{aligned}
M_W(t) &= E(e^{tw}) \\
&= \frac{\alpha\beta}{\alpha+\beta}\int_0^\infty e^{tw}e^{-\alpha w}\,dw + \frac{\alpha\beta}{\alpha+\beta}\int_{-\infty}^0 e^{tw}e^{\beta w}\,dw \\
&= \frac{\alpha\beta}{\alpha+\beta}\int_0^\infty e^{tw-\alpha w}\,dw + \frac{\alpha\beta}{\alpha+\beta}\int_{-\infty}^0 e^{tw+\beta w}\,dw \\
&= \frac{\alpha\beta}{\alpha+\beta}\int_0^\infty e^{tw-\alpha w}\,dw + \frac{\alpha\beta}{\alpha+\beta}\left.\frac{e^{tw+\beta w}}{t+\beta}\right|_{-\infty}^0 \\
&= \frac{\alpha\beta}{\alpha+\beta}\int_0^\infty e^{(t-\alpha)w}\,dw + \frac{\alpha\beta}{(\alpha+\beta)(\beta+t)}.
\end{aligned}$$

Let $w = -s$, then we have $dw = -ds$. Thus,

$$\begin{aligned}
M_W(t) &= \frac{\alpha\beta}{\alpha+\beta}\int_0^{-\infty} -e^{-(t-\alpha)s}\,ds + \frac{\alpha\beta}{(\alpha+\beta)(\beta+t)} \\
&= \frac{\alpha\beta}{\alpha+\beta}\int_{-\infty}^0 e^{(\alpha-t)s}\,ds + \frac{\alpha\beta}{(\alpha+\beta)(\beta+t)} \\
&= \frac{\alpha\beta}{\alpha+\beta}\left.\frac{e^{(\alpha-t)s}}{\alpha-t}\right|_{-\infty}^0 + \frac{\alpha\beta}{(\alpha+\beta)(\beta+t)} \\
&= \frac{\alpha\beta}{(\alpha+\beta)(\alpha-t)} + \frac{\alpha\beta}{(\alpha+\beta)(\beta+t)} \\
&= \frac{\alpha\beta}{(\alpha-t)(\beta+t)},
\end{aligned} \tag{1.2.24}$$

where $-\beta < t < \alpha$.

By theorem 1.1.1, the mgf of the normal-Laplace distribution is obtained as the product of (1.2.23) and (1.2.24),

$$M_Y(t) = \frac{\alpha\beta\exp(\mu t + \sigma^2 t^2/2)}{(\alpha-t)(\beta+t)}, \quad -\beta < t < \alpha. \tag{1.2.25}$$

(4) *Mean, variance and cumulants*

From the mgf, we can derive the mean and the variance of the NL distribution by evaluating the first two cumulants κ_1 and κ_2. First, the cumulant generation function of Y is equal to

$$K_Y(t) = \log M_Y(t) = \log\alpha + \log\beta + \mu t + \sigma^2 t^2/2 - \log(\alpha - t) - \log(\beta + t), -\beta < t < \alpha.$$

Then,

$$\frac{d}{dt}\log M_Y(t) = \mu + \sigma^2 t + 1/(\alpha - t) - 1/(\beta + t).$$

By proposition 1.1.1,

$$E(Y) = \kappa_1 = \frac{d}{dt}\log M_Y(t)\bigg|_{t=0} = \mu + 1/\alpha - 1/\beta.$$

Similarly, we have

$$\frac{d^2}{dt^2}\log M_Y(t) = \sigma^2 + 1/(\alpha - t)^2 + 1/(\beta + t)^2.$$

The variance of Y can be expressed as

$$Var(Y) = \kappa_2 = \frac{d^2}{dt^2}\log M_Y(t)\bigg|_{t=0} = \sigma^2 + 1/\alpha^2 + 1/\beta^2.$$

In order to find higher order cumulants, let us rewrite

$$\log M_Y(t) = \mu t + \sigma^2 t^2/2 - \log(1 - t/\alpha) - \log(1 + t/\beta), \quad -\beta < t < \alpha;$$

By using the Taylor's series expansion of

$$-\log(1 - t/\alpha) = \sum_{n=1}^{\infty}(n-1)!t^n/\alpha^n n!$$

and

$$-\log(1 + t/\beta) = \sum_{n=1}^{\infty}(-1)^n(n-1)!t^n/\beta^n n!,$$

we obtain the higher order cumulants of the NL distribution for $n > 2$,

$$\kappa_n = (n-1)!(\alpha^{-n} + (-1)^n\beta^{-n}), \qquad n > 2. \tag{1.2.26}$$

Particularly, the third and fourth order cumulants are

$$\kappa_3 = 2/\alpha^3 - 2/\beta^3 \text{ and } \kappa_4 = 6/\alpha^4 + 6/\beta^4.$$

(5) *Closure under a linear transformation*

Reed (2007) defined the generalized normal-Laplace (GNL) distribution as the distribution of a random variable with characteristic function

$$\phi(t) = \left[\frac{\alpha\beta\exp(\mu it - \sigma^2 t^2/2)}{(\alpha - it)(\beta + is)}\right]^{\rho},$$

where α, β, ρ and σ are positive parameters and $-\infty < \mu < \infty$.

He showed that the normal-Laplace distribution is a special case of the generalized normal-Laplace, $GNL(\alpha, \beta, \mu, \sigma^2, \rho)$, distribution with $\rho = 1$, and that the family of GNL distributions is closed under linear transformations. Therefore, the NL distribution also satisfies this property. Precisely, if $Y \sim NL(\alpha, \beta, \mu, \sigma^2)$ and $a > 0$ and b is any constant, according to the moment generating function of the NL distribution (1.2.25), the mgf of $aY + b$ can be written as,

$$M_{aY+b}(t) = E[e^{(aY+b)t}] = e^{bt} M_Y(at) = e^{bt} \frac{\alpha\beta e^{\mu at + \sigma^2 a^2 t^2/2}}{(\alpha - at)(\beta + at)}.$$

Dividing the terms $\alpha - at$ and $\beta + at$ in denominator and α and β in numerator by a, we obtain,

$$M_{aY+b}(t) = \frac{(\alpha/a)(\beta/a)e^{(\mu a+b)t + \sigma^2 a^2 t^2/2}}{(\alpha/a - t)(\beta/a + t)}.$$

Therefore,

$$aY + b \sim NL(\alpha/a, \beta/a, a\mu + b, a^2\sigma^2).$$

For $a < 0$, in order to make sure that the first two parameters are positive, we need to rewrite,

$$M_{aY+b}(t) = \frac{(-\alpha/a)(-\beta/a)e^{(\mu a+b)t + \sigma^2 a^2 t^2/2}}{(t - \alpha/a)(-\beta/a - t)} = \frac{(-\alpha/a)(-\beta/a)e^{(\mu a+b)t + \sigma^2 a^2 t^2/2}}{(-\beta/a - t)(-\alpha/a + t)}.$$

So,

$$aY + b \sim NL(-\beta/a, -\alpha/a, a\mu + b, a^2\sigma^2).$$

(6) *The NL distribution is infinitely divisible*

Reed and Jorgensen (2004) rewrite the mgf of Y as

$$M_Y(t) = \left[\exp\left(\frac{\mu}{n}t + \frac{\sigma^2}{2n}t^2\right)\left(\frac{\alpha}{\alpha - t}\right)^{1/n}\left(\frac{\beta}{\beta + t}\right)^{1/n}\right]^n$$

for any integer $n > 0$; note that the term in square brackets is the mgf of a random variable formed as $Z + G_1 - G_2$, where Z, G_1 and G_2 are independent and $Z \sim N\left(\frac{\mu}{n}, \frac{\sigma^2}{n}\right)$ and G_1 and G_2 have gamma distributions $\Gamma(1/n, \alpha)$ and $\Gamma(1/n, \beta)$ respectively. As Y can be expressed as an independent sum of n GNL random variables for arbitrary integer n, Y is infinitely divisible.

(7) *The symmetric NL distribution*

If we set $\alpha = \beta$, from (1.2.15), we have

$$\begin{aligned} g(\mu + x) &= \frac{\alpha}{2}\phi\left(\frac{x}{\sigma}\right)[R(\alpha\sigma - x/\sigma) + R(\alpha\sigma + x/\sigma)] \\ &= \frac{\alpha}{2}\phi\left(-\frac{x}{\sigma}\right)[R(\alpha\sigma + x/\sigma) + R(\alpha\sigma - x/\sigma)] = g(\mu - x). \end{aligned}$$

Therefore, the pdf of the NL distribution is symmetric about the line $x = \mu$. The pdf and cdf of the symmetric NL distribution become respectively

$$g(y) = \frac{\alpha}{2}\phi\left(\frac{y-\mu}{\sigma}\right)[R(\alpha\sigma - (y-\mu)/\sigma) + R(\alpha\sigma + (y-\mu)/\sigma)]$$

and

$$G(y) = \Phi\left(\frac{y-\mu}{\sigma}\right) - \phi\left(\frac{y-\mu}{\sigma}\right)\frac{R(\alpha\sigma - (y-\mu)/\sigma) - R(\alpha\sigma + (y-\mu)/\sigma)}{2}.$$

1.3. The double Pareto-lognormal distribution

Since the double Pareto-lognormal (dPlN) distribution is related to the normal-Laplace distribution, we will derive the probability density function of the dPlN distribution from the NL distributionin in this section. Moreover, some of its properties will also be presented.

1.3.1. Definition of the double Pareto-lognormal distribution

The double Pareto-lognormal distribution is related to the normal-Laplace distribution in the same way as the lognormal is related to the normal, i.e. a random variable X follows the double Pareto-lognormal distribution if $\log X \sim NL(\alpha, \beta, \mu, \sigma^2)$. Therefore, the pdf of X can be obtained from the pdf of the normal-Laplace distribution (1.2.7),

$$f(x) = \left|\frac{\partial \log x}{\partial x}\right| g(\log x) = \frac{1}{x}g(\log x) = \frac{1}{x}g_1(\log x) + \frac{1}{x}g_2(\log x), \; x \geq 0. \quad (1.3.1)$$

From $g_1(y)$, the first component of $f(x)$ can be derived as follows

$$\frac{1}{x}g_1(\log x) = \frac{\alpha\beta}{\alpha+\beta}\frac{1}{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\log x-\mu}{\sigma}\right)^2}\frac{\int_{\log x}^{+\infty}\frac{1}{\sigma}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left[\frac{z-(\mu-\beta\sigma^2)}{\sigma}\right]^2}dz}{\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left[\frac{\log x-(\mu-\beta\sigma^2)}{\sigma}\right]^2}}$$
$$= \frac{\alpha\beta}{\alpha+\beta}\frac{1}{x}e^{-\frac{1}{2}\left(\frac{\log x-\mu}{\sigma}\right)^2}\frac{1-\Phi\left[\frac{\log x-(\mu-\beta\sigma^2)}{\sigma}\right]}{e^{-\frac{1}{2}\left[\frac{\log x-(\mu-\beta\sigma^2)}{\sigma}\right]^2}}$$
$$= \frac{\alpha\beta}{\alpha+\beta}\frac{1}{x}e^{\frac{1}{2}\left[\frac{\log x-(\mu-\beta\sigma^2)}{\sigma}\right]^2-\frac{1}{2}\left(\frac{\log x-\mu}{\sigma}\right)^2}\Phi^c\left[\frac{\log x-(\mu-\beta\sigma^2)}{\sigma}\right]$$
$$= \frac{\alpha\beta}{\alpha+\beta}\frac{1}{x}e^{\beta\log x-\beta\mu+\frac{1}{2}\beta^2\sigma^2}\Phi^c\left(\frac{\log x-\mu+\beta\sigma^2}{\sigma}\right)$$
$$= \frac{\alpha\beta}{\alpha+\beta}x^{\beta-1}e^{-\beta\mu+\frac{1}{2}\beta^2\sigma^2}\Phi^c\left(\frac{\log x-\mu+\beta\sigma^2}{\sigma}\right). \tag{1.3.2}$$

Similarly, from $g_2(y)$,

$$\frac{1}{x}g_2(\log x) = \frac{\alpha\beta}{\alpha+\beta}\frac{1}{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\log x-\mu}{\sigma}\right)^2}\frac{\int_{-\log x}^{+\infty}\frac{1}{\sigma}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left[\frac{t+(\mu+\alpha\sigma^2)}{\sigma}\right]^2}dt}{\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left[\frac{(\mu+\alpha\sigma^2)-\log x}{\sigma}\right]^2}}$$

Let $t=-z$, then we have $dt=-dz$.

$$= \frac{\alpha\beta}{\alpha+\beta}\frac{1}{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\log x-\mu}{\sigma}\right)^2}\frac{\int_{\log x}^{-\infty}-\frac{1}{\sigma}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left[\frac{-z+(\mu+\alpha\sigma^2)}{\sigma}\right]^2}dz}{\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left[\frac{(\mu+\alpha\sigma^2)-\log x}{\sigma}\right]^2}}$$
$$= \frac{\alpha\beta}{\alpha+\beta}\frac{1}{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\log x-\mu}{\sigma}\right)^2}\frac{\int_{-\infty}^{\log x}\frac{1}{\sigma}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left[\frac{z-(\mu+\alpha\sigma^2)}{\sigma}\right]^2}dz}{\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left[\frac{(\mu+\alpha\sigma^2)-\log x}{\sigma}\right]^2}}$$
$$= \frac{\alpha\beta}{\alpha+\beta}\frac{1}{x}e^{-\frac{1}{2}\left(\frac{\log x-\mu}{\sigma}\right)^2}\frac{\Phi\left[\frac{\log x-(\mu+\alpha\sigma^2)}{\sigma}\right]}{e^{-\frac{1}{2}\left[\frac{(\mu+\alpha\sigma^2)-\log x}{\sigma}\right]^2}}$$
$$= \frac{\alpha\beta}{\alpha+\beta}\frac{1}{x}e^{\frac{1}{2}\left[\frac{\mu-\log x+\alpha\sigma^2}{\sigma}\right]^2-\frac{1}{2}\left(\frac{\log x-\mu}{\sigma}\right)^2}\Phi\left[\frac{\log x-(\mu+\alpha\sigma^2)}{\sigma}\right]$$
$$= \frac{\alpha\beta}{\alpha+\beta}\frac{1}{x}e^{\frac{2(\mu-\log x)\alpha\sigma^2+\alpha^2\sigma^4}{2\sigma^2}}\Phi\left[\frac{\log x-(\mu+\alpha\sigma^2)}{\sigma}\right]$$
$$= \frac{\alpha\beta}{\alpha+\beta}\frac{1}{x}e^{-\alpha\log x+\alpha\mu+\frac{1}{2}\alpha^2\sigma^2}\Phi\left(\frac{\log x-\mu-\alpha\sigma^2}{\sigma}\right)$$

$$= \frac{\alpha\beta}{\alpha+\beta} x^{-\alpha-1} e^{\alpha\mu+\frac{1}{2}\alpha^2\sigma^2} \Phi\left(\frac{\log x - \mu - \alpha\sigma^2}{\sigma}\right). \tag{1.3.3}$$

By adding the two components (1.3.2) and (1.3.3) of (1.3.1), we obtain

$$\begin{aligned} f(x) = \frac{\alpha\beta}{\alpha+\beta} \Bigg[& x^{-\alpha-1} e^{\alpha\mu+\frac{1}{2}\alpha^2\sigma^2} \Phi\left(\frac{\log x - \mu - \alpha\sigma^2}{\sigma}\right) + \\ & x^{\beta-1} e^{-\beta\mu+\frac{1}{2}\beta^2\sigma^2} \Phi^c\left(\frac{\log x - \mu + \beta\sigma^2}{\sigma}\right)\Bigg]. \end{aligned} \tag{1.3.4}$$

Reed and Jorgensen (2004) gave a convenient way to express the previous pdf, as follows:

$$\begin{aligned} f(x) = \frac{\alpha\beta}{\alpha+\beta} \Bigg[& x^{-\alpha-1} A(\alpha,\mu,\sigma) \Phi\left(\frac{\log x - \mu - \alpha\sigma^2}{\sigma}\right) + \\ & x^{\beta-1} A(-\beta,\mu,\sigma) \Phi^c\left(\frac{\log x - \mu + \beta\sigma^2}{\sigma}\right)\Bigg], \end{aligned} \tag{1.3.5}$$

where

$$A(\theta,\mu,\sigma) = \exp(\theta\mu + \theta^2\sigma^2/2).$$

This distribution is defined as the *double Pareto-lognormal distribution* which is written as

$$X \sim dPlN(\alpha,\beta,\mu,\sigma^2)$$

to indicate that a random variable X follows this distribution with the four parameters α, β, μ and σ^2, where α, β, $\sigma^2>0$ and $\mu \in \mathcal{R}$. From (1.2.16) a $dPlN(\alpha,\beta,\mu,\sigma^2)$ random variable can be represented as

$$X = e^Y \stackrel{d}{=} e^{\mu+\sigma Z} \frac{e^{E_1/\alpha}}{e^{E_2/\beta}}. \tag{1.3.6}$$

We may find the distribution of a random variable of the form $V = e^{E/\theta}$, where E is a standard exponential variable,

$$f_V(v) = \frac{\theta}{v} f_E(\theta \log v) = \frac{\theta}{v} e^{-\theta \log v} = \theta v^{-\theta-1}.$$

Clearly V follows a Pareto distribution, hence (1.3.6) can be rewritten as

$$X \stackrel{d}{=} UV_1/V_2, \tag{1.3.7}$$

where U, V_1 and V_2 are independent, with U lognormal, i.e. $\log U \sim N(\mu,\sigma^2)$, and with V_1 and V_2 following Pareto distributions with parameters α and β respectively, with pdf

$$f(v) = \theta v^{-\theta-1}, \quad v > 1$$

where $\theta = \alpha$ or $\theta = \beta$ respectively.

Alternatively, Reed and Jorgensen (2004) write

$$X \stackrel{d}{=} UQ,$$

where Q is the ratio of the above Pareto random variables, so that Q has pdf (see Reed and Jorgensen 2004)

$$f_Q(q) = \begin{cases} \frac{\alpha\beta}{\alpha+\beta} q^{\beta-1}, & \text{for } 0 < q \leq 1 \\ \frac{\alpha\beta}{\alpha+\beta} q^{-\alpha-1}, & \text{for } q > 1. \end{cases} \tag{1.3.8}$$

The distribution with pdf (1.3.8) is called a double Pareto distribution. Therefore the reason why the distribution of X is named double Pareto-lognormal distribution is that a such distribution results from the product of independent double Pareto and lognormal components.

To generate pseudo-random variables from the $dPlN(\alpha, \beta, \mu, \sigma^2)$ distribution, one can exponentiate the pseudo-random variables generated from $NL(\alpha, \beta, \mu, \sigma^2)$ using (1.2.16).

1.3.2. Properties

Based on the work of Reed and Jorgensen (2004), the double Pareto-lognormal distribution has the following properties.

(1) *Cumulative distribution function*

The cdf of $X \sim dPlN(\alpha, \beta, \mu, \sigma^2)$ can be written as $F_X(x) = G_Y(\log x)$ or $F_X(x) = G_1(\log x) + G_2(\log x)$, where the cdf of the NL distribution G_Y is given by (1.2.22) and

$$G_1(\log x) = \int_{\log x}^{+\infty} \frac{\alpha}{\alpha+\beta} e^{\beta(\log x - z)} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2} dz, \text{ if } \log x \leq z \tag{1.3.9}$$

and

$$G_2(\log x) = \int_{-\infty}^{\log x} \left(1 - \frac{\beta}{\alpha+\beta} e^{-\alpha(\log x - z)}\right) \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2} dz, \text{ if } \log x > z. \tag{1.3.10}$$

The calculation of $G_1(\log x)$ is similar to $g_1(\log x)$ in section (1.3.1); after moving the constant from the integral, we can easily obtain

$$\begin{aligned} G_1(\log x) &= \frac{\alpha}{\alpha+\beta}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\log x-\mu}{\sigma}\right)^2}\frac{\int_{\log x}^{+\infty}\frac{1}{\sigma}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left[\frac{z-(\mu-\beta\sigma^2)}{\sigma}\right]^2}dz}{\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left[\frac{\log x-(\mu-\beta\sigma^2)}{\sigma}\right]^2}} \\ &= \frac{\alpha}{\alpha+\beta}x^{\beta}e^{-\beta\mu+\frac{1}{2}\beta^2\sigma^2}\Phi^c\left(\frac{\log x-\mu+\beta\sigma^2}{\sigma}\right). \end{aligned} \tag{1.3.11}$$

The first term of $G_2(\log x)$ will be the cdf a log-normal distribution and the second term can be easily found in the same way as $g_2(\log x)$, that is

$$\begin{aligned} G_2(\log x) &= \Phi\left(\frac{\log x-\mu}{\sigma}\right) - \frac{\beta}{\alpha+\beta}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\log x-\mu}{\sigma}\right)^2}\frac{\int_{-\log x}^{+\infty}\frac{1}{\sigma}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left[\frac{t+(\mu+\alpha\sigma^2)}{\sigma}\right]^2}dt}{\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left[\frac{(\mu+\alpha\sigma^2)-\log x}{\sigma}\right]^2}} \\ &= \Phi\left(\frac{\log x-\mu}{\sigma}\right) - \frac{\beta}{\alpha+\beta}x^{-\alpha}e^{\alpha\mu+\frac{1}{2}\alpha^2\sigma^2}\Phi\left(\frac{\log x-\mu-\alpha\sigma^2}{\sigma}\right). \end{aligned} \tag{1.3.12}$$

Then the cdf of $X \sim dPlN(\alpha, \beta, \mu, \sigma^2)$ can be expressed by combining $G_1(\log x)$ and $G_2(\log x)$,

$$\begin{aligned} F_X(x) = \Phi\left(\frac{\log x-\mu}{\sigma}\right) - \frac{1}{\alpha+\beta}\Bigg[&\beta x^{-\alpha}A(\alpha,\mu,\sigma)\Phi\left(\frac{\log x-\mu-\alpha\sigma^2}{\sigma}\right) + \\ &\alpha x^{\beta}A(-\beta,\mu,\sigma)\Phi^c\left(\frac{\log x-\mu+\beta\sigma^2}{\sigma}\right)\Bigg], \end{aligned} \tag{1.3.13}$$

where

$$A(\theta, \mu, \sigma) = \exp(\theta\mu + \theta^2\sigma^2/2).$$

(2) *The limiting forms when $\alpha \to \infty$ or $\beta \to \infty$*

As $\alpha \to \infty$, the pdf of the dPlN distribution has the limiting form

$$f_1(x) = \beta x^{\beta-1}A(-\beta,\mu,\sigma)\Phi^c\left(\frac{\log x-\mu+\beta\sigma^2}{\sigma}\right). \tag{1.3.14}$$

As $\beta \to \infty$, the pdf becomes

$$f_2(x) = \alpha x^{-\alpha-1}A(\alpha,\mu,\sigma)\Phi\left(\frac{\log x-\mu-\alpha\sigma^2}{\sigma}\right). \tag{1.3.15}$$

Colombi (1990) considered this distribution, which he called the Pareto-lognormal, as a model for income distributions.

Clearly the $dPlN(\alpha, \beta, \mu, \sigma^2)$ pdf (1.3.5) can be represented as a mixture of the above pdfs

$$f(x) = \frac{\alpha}{\alpha+\beta} f_1(x) + \frac{\beta}{\alpha+\beta} f_2(x).$$

(3) *Power-law tail behaviour*

The $dPlN(\alpha, \beta, \mu, \sigma^2)$ distribution exhibits power-law in both tails.

$$\text{If } x \to \infty,\; f(x) \sim \alpha A(\alpha, \mu, \sigma) x^{-\alpha-1}, \text{ and if } x \to 0,\; f(x) \sim \beta A(-\beta, \mu, \sigma) x^{\beta-1}.$$

The cdf $F_X(x)$ and the survival function $S_X(x) = 1 - F_X(x)$ also exhibit power-law tail behaviour:

$$\text{If } x \to \infty,\; S_X(x) \sim \alpha A(\alpha, \mu, \sigma) x^{-\alpha}, \text{ and if } x \to 0,\; F_X(x) \sim \beta A(-\beta, \mu, \sigma) x^{\beta}.$$

However, the limiting case $f_1(x)$ exhibits only lower-tail power-law behaviour: as $x \to 0$, $f_1(x) \sim \beta A(-\beta, \mu, \sigma) x^{\beta-1}$; the pdf $f_2(x)$ exhibits only upper-tail power-law behaviour: as $x \to \infty$, $f_2(x) \sim \alpha A(\alpha, \mu, \sigma) x^{-\alpha-1}$.

(4) *Hazard rate.*

Also known as the force of mortality, it is denoted $h_X(x)$ and defined as the ratio of the density $f_X(x)$ and the survival function $S_X(x)$. That is,

$$\begin{aligned} h_X(x) &= \frac{f_X(x)}{S_X(x)} \\ &= \frac{\frac{\alpha\beta}{\alpha+\beta}\left[x^{-\alpha-1} A(\alpha, \mu, \sigma) \Phi\left(\frac{\log x - \mu - \alpha\sigma^2}{\sigma}\right) + x^{\beta-1} A(-\beta, \mu, \sigma) \Phi^c\left(\frac{\log x - \mu + \beta\sigma^2}{\sigma}\right)\right]}{1 - F_X(x)}, \end{aligned} \tag{1.3.16}$$

where $F_X(x)$ is expressed as (1.3.13).
According to the power-law tail behaviour of the dPlN distribution, as $x \to \infty$,

$$h_X(x) \sim \frac{\alpha A(\alpha, \mu, \sigma) x^{-\alpha-1}}{\alpha A(\alpha, \mu, \sigma) x^{-\alpha}} = \frac{1}{x}.$$

Then as $x \to \infty$, $h_X(x) \to 0$.

(5) *Mean, variance and moments.*

Recall that if $Y = \log X$ follows a NL distribution, then $X = e^Y$ follows a dPlN distribution. The kth raw moment of the dPlN distribution can be easily obtained by (1.2.25) the moment generating function of the NL distribution.

$$\mu'_k = E(X^k) = E[(e^Y)^k] = M_Y(k) = \frac{\alpha\beta \exp(\mu k + \sigma^2 k^2/2)}{(\alpha - k)(\beta + k)}. \tag{1.3.17}$$

Note that μ'_k exists only for $k < \alpha$. Since the kth raw moment does not always exist, the moment generating function of the dPlN distribution does not exist. If $k = 1$, the mean (for $\alpha > 1$) can be expressed as

$$E(X) = \frac{\alpha\beta \exp(\mu + \sigma^2/2)}{(\alpha - 1)(\beta + 1)}, \tag{1.3.18}$$

while the variance (for $\alpha > 2$) is

$$\begin{aligned} Var(X) = E(X^2) - E(X)^2 &= \frac{\alpha\beta e^{2\mu+2\sigma^2}}{(\alpha-2)(\beta+2)} - \frac{\alpha^2\beta^2 e^{2\mu+\sigma^2}}{(\alpha-1)^2(\beta+1)^2} \\ &= \frac{\alpha\beta e^{2\mu+\sigma^2}}{(\alpha-1)^2(\beta+1)^2}\left[\frac{(\alpha-1)^2(\beta+1)^2 e^{\sigma^2}}{(\alpha-2)(\beta+2)} - \alpha\beta\right], \end{aligned} \tag{1.3.19}$$

and the coefficient of variation (for $\alpha > 2$) is

$$\begin{aligned} CV = \frac{\sqrt{Var(X)}}{E(X)} &= \frac{\frac{(\alpha\beta)^{1/2} e^{\mu+\sigma^2/2}}{(\alpha-1)(\beta+1)}\left[\frac{(\alpha-1)^2(\beta+1)^2 e^{\sigma^2}}{(\alpha-2)(\beta+2)} - \alpha\beta\right]^{1/2}}{\frac{\alpha\beta e^{\mu+\sigma^2/2}}{(\alpha-1)(\beta+1)}} \\ &= \left[\frac{(\alpha-1)^2(\beta+1)^2 e^{\sigma^2}}{\alpha\beta(\alpha-2)(\beta+2)} - 1\right]^{1/2}. \end{aligned} \tag{1.3.20}$$

The coefficient of variation is independent of μ, increases with σ^2 and decreases with α and β.

(6) *Closure under power-law transformation*

The dPlN family of distributions is closed under the power-law transformation, i.e. if $X \sim dPlN(\alpha, \beta, \mu, \sigma^2)$, then for constants $c > 0, d \in \mathbb{R}$, cX^d will also follow a dPlN distribution. To prove this fact, we only need to show that $\log(cX^d)$ follows a normal-Laplace distribution. First of all,

$$\log(cX^d) = \log c + d \log X.$$

Since $\log c \in \mathbb{R}$, we must have $c > 0$. Its moment generating function can be expressed as,

$$M_{\log c + d\log X}(t) = E[e^{(\log c + d \log x)t}] = e^{t \log c} M_{\log X}(dt).$$

We know that $Y = \log X \sim NL(\alpha, \beta, \mu, \sigma^2)$. Then, from the mgf of the NL distribution (1.2.25),

$$M_{\log c + d\log X}(t) = e^{t\log c} M_Y(dt) = e^{t \log c} \frac{\alpha\beta e^{\mu dt + \sigma^2 d^2 t^2/2}}{(\alpha - dt)(\beta + dt)},$$

then dividing both of the numerator and the denominator by d^2, if $d > 0$, we have

$$M_{\log(cX^d)}(t) = \frac{(\alpha/d)(\beta/d)e^{(\mu d+\log c)t+\sigma^2 d^2 t^2/2}}{(\alpha/d - t)(\beta/d + t)}.$$

So,

$$\log(cX^d) \sim NL(\alpha/d, \beta/d, d\mu + \log c, d^2\sigma^2).$$

Clearly,

$$cX^d \sim dPlN(\alpha/d, \beta/d, d\mu + \log c, d^2\sigma^2). \tag{1.3.21}$$

If $d < 0$, in order to make sure that the first two parameters are positive, we need to rewrite the moment generating function of $\log(cX^d)$ as,

$$M_{\log(cX^d)}(t) = \frac{(-\alpha/d)(-\beta/d)e^{(\mu d+\log c)t+\sigma^2 d^2 t^2/2}}{(-\alpha/d + t)(-\beta/d - t)} = \frac{(-\alpha/d)(-\beta/d)e^{(\mu d+\log c)t+\sigma^2 d^2 t^2/2}}{(-\beta/d - t)(-\alpha/d + t)}.$$

Thus,

$$\log(cX^d) \sim NL(-\beta/d, -\alpha/d, d\mu + \log c, d^2\sigma^2).$$

We obtain

$$cX^d \sim dPlN(-\beta/d, -\alpha/d, d\mu + \log c, d^2\sigma^2). \tag{1.3.22}$$

(7) *Application in finance*

Let i_j denote the daily stock price return and i denote compound daily return over the n-day period, we have

$$1 + i = \prod_{j=1}^{n} (1 + i_j)^{1/n}.$$

Then the logarithmic stock price return can be expressed as

$$\log(1 + i) = \frac{1}{n} \sum_{j=1}^{n} \log(1 + i_j).$$

If $1 + i_j \sim dPlN(\alpha, \beta, \mu_j, \sigma_j^2)$ and $1 + i_j$ are independent random variables with common α and β, then

$$\log(1 + i_j) \sim NL(\alpha, \beta, \mu_j, \sigma_j^2).$$

As mentioned in section 1.2.2, we have

$$\log(1 + i_j) \sim GNL(\alpha, \beta, \mu_j, \sigma_j^2, 1).$$

Since the sum of n independent $GNL(\alpha, \beta, \mu_j, \sigma_j^2, \rho)$ random variables also follows a GNL distribution (see Groparu-Cojocaru and Doray, 2013) as

$$GNL(\alpha, \beta, \sum_{j=1}^{n} \mu_j/n, \sum_{j=1}^{n} \sigma_j^2/n, n\rho).$$

We write the sum of the logarithmic daily returns as

$$\sum_{j=1}^{n} \log(1+i_j) \sim GNL(\alpha, \beta, \sum_{j=1}^{n} \mu_j/n, \sum_{j=1}^{n} \sigma_j^2/n, n).$$

The GNL distribution is closed under linear transformation (Reed, 2007), i.e. if $W \sim GNL(\alpha, \beta, \mu, \sigma^2, \rho)$ then for constants $a, b > 0$, $a+bW \sim GNL(\alpha/b, \beta/b, b\mu + a/\rho, b^2\sigma^2, \rho)$.

Thus, we obtain

$$\log(1+i) = \frac{1}{n}\sum_{j=1}^{n} \log(1+i_j) \sim GNL(n\alpha, n\beta, \sum_{j=1}^{n} \mu_j/n^2, \sum_{j=1}^{n} \sigma_j^2/n^3, n).$$

Chapter 2

ESTIMATION OF PARAMETERS

In this chapter, we propose two methods to estimate the parameters of the double Pareto-lognormal distribution. One is the method of moments which is relatively easy to implement but tends to give poor results. The other method is the maximum likelihood estimation which is more difficult to use but has superior statistical properties and is considerably more flexible.

2.1. Method of moments

For this method, we assume that all n observations are independent and from the same parametric distribution. In particular, let the cumulative distribution function be given by

$$F(x|\theta), \quad \theta^T = (\theta_1, \theta_2, ..., \theta_p),$$

where θ^T is the transpose of θ. That is, θ is a column vector containing the p parameters to be estimated. Furthermore, let $\mu'_k(\theta) = E(X^k)$ be the kth raw moment, and let us assume the kth moment exists. For a sample of n independent observations from this random variable, let $\hat{\mu}'_k = \frac{1}{n}\sum_{j=1}^{n} x_j^k$ be the empirical estimate of the kth moment.

Definition 2.1.1. *(Klugman et al. 2008) Suppose $\mu_k(\theta)$ exists, a method-of-moments estimate of θ is any solution of the p equations*

$$\mu'_k(\theta) = \hat{\mu}'_k, \quad k = 1, 2, ..., p.$$

The motivation for this estimator is that it produces a model that has the same first p raw moments as the data (as represented by the empirical distribution). The traditional definition of the method of moments uses positive integers for the moments. Arbitrary negative or fractional moments could also be used.

In the case of the double Pareto-lognormal distribution, in order to calculate the estimates of the four parameters $\hat{\alpha}$, $\hat{\beta}$, $\hat{\mu}$ and $\hat{\sigma}^2$, we need to solve the equations using the first four raw moments given by (1.3.17):

$$\mu_1' = \frac{\alpha\beta\exp(\mu + \sigma^2/2)}{(\alpha-1)(\beta+1)},$$

$$\mu_2' = \frac{\alpha\beta\exp(2\mu + 2\sigma^2)}{(\alpha-2)(\beta+2)},$$

$$\mu_3' = \frac{\alpha\beta\exp(3\mu + 9\sigma^2/2)}{(\alpha-3)(\beta+3)},$$

$$\mu_4' = \frac{\alpha\beta\exp(4\mu + 8\sigma^2)}{(\alpha-4)(\beta+4)}.$$

Unfortunately, we cannot find the method of moments estimates of the four parameters analytically. If the given data are assumed to be from the dPlN distribution, one could, in principle, obtain these estimators numerically. However, if the raw moments of order α or greater do not exist, for example, if $\alpha \leq 4$ or $\alpha = 1, 2, 3$, then the direct use of the dPlN moments may result in poor estimates. Therefore having first log-transformed the data and using the normal-Laplace distribution to find the method of moments estimates is recommended.

We can use the sample cumulants instead of the sample raw moments to compute the method of moments estimates, because these two methods will result in the same estimators. In the case of the log-transformed data following a symmetric normal-Laplace distribution ($\alpha = \beta$), we need to calculate the estimates of three parameters $\hat{\alpha}$=$\hat{\beta}$, $\hat{\mu}$ and $\hat{\sigma}^2$, therefore the sample cumulants κ_1, κ_2 and κ_4 will be set equal to their theoretical counterparts. Note that the third cumulant is equal to zero in this case.
If X follows a symmetric normal-Laplace distribution, consider the following equations:

$$\kappa_1 = E(X) = \mu \tag{2.1.1}$$

$$\kappa_2 = Var(X) = \sigma^2 + 2\alpha^{-2} \tag{2.1.2}$$

$$\kappa_4 = 12\alpha^{-4}. \tag{2.1.3}$$

Using the first equation, we obtain

$$\mu = \kappa_1.$$

From equation (2.1.3) and the fact that $\alpha > 0$, we have

$$\alpha = \left(\frac{12}{\kappa_4}\right)^{1/4}.$$

Replace α in equation (2.1.2) by the previous expression

$$\kappa_2 = \sigma^2 + \left(\frac{12}{\kappa_4}\right)^{-1/2},$$

then

$$\sigma^2 = \kappa_2 - \left(\frac{12}{\kappa_4}\right)^{-1/2}.$$

Thus, the estimates of the three parameters are

$$\hat{\alpha} = \left(\frac{12}{\kappa_4}\right)^{1/4}$$

$$\hat{\mu} = \kappa_1$$

and

$$\hat{\sigma}^2 = \kappa_2 - \sqrt{\frac{\kappa_4}{12}},$$

where κ_1, κ_2 and κ_4 are sample cumulants obtained from the log-transformed data.

For the general case, the log-transformed data following a normal-Laplace distribution with four parameters to estimate, the first four sample cumulants must be set equal to their theoretical counterparts, that is to say,

$$\kappa_1 = \mu + \alpha^{-1} - \beta^{-1} \tag{2.1.4}$$

$$\kappa_2 = \sigma^2 + \alpha^{-2} + \beta^{-2} \tag{2.1.5}$$

$$\kappa_3 = 2\alpha^{-3} - 2\beta^{-3} \tag{2.1.6}$$

$$\kappa_4 = 6\alpha^{-4} + 6\beta^{-4}. \tag{2.1.7}$$

In order to find $\hat{\alpha}$ and $\hat{\beta}$, one only needs to solve the equations (2.1.6) and (2.1.7). Then $\hat{\mu}$ and $\hat{\sigma}^2$ can be obtained from the first two equations, as

$$\hat{\mu} = \kappa_1 - \hat{\alpha}^{-1} + \hat{\beta}^{-1} \text{ and } \hat{\sigma}^2 = \kappa_2 - \hat{\alpha}^{-2} - \hat{\beta}^{-2}.$$

There is no analytical solution for method of moments estimates in this case and numerical methods must be used. Reed and Jorgensen (2004) recommended the use of the method of moments only to get starting values in the iterative procedure for finding maximum likelihood estimates. The method of maximum likelihood will be introduced in the next section.

2.2. Method of maximum likelihood

Estimation by the method of moments is often easy to do, but these estimators tend to perform poorly mainly because they use few features of the data, rather than the entire set of observations. It is important to use as much information as possible when the population has a heavy right tail.

There are a variety of estimators based on individual data points. All of them are implemented by setting an objective function and then determining the parameter values that optimize that function. The only one used here is the maximum likelihood estimator.

To define our maximum likelihood estimator, let the data set consist of n events $A_1, ..., A_n$, where A_i is whatever was observed for the ith observation. Further assume that the event A_i results from observing the random variable X_i. The random variables $X_1, ...X_n$ are assumed identically independently distributed. And their distribution depends on the same parameter vector θ.

Definition 2.2.1. *(Klugman et al. 2008) The likelihood function is*

$$L(\theta) = \prod_{i=1}^{n} \Pr(X_i \in A_i|\theta)$$

and the maximum likelihood estimate of θ is the vector that maximizes the likelihood function.

However it is often easier to maximize the logarithm of the likelihood function, the log-likelihood function denoted as $l(\theta) = \log L(\theta)$.

When there is no truncation and no censoring, and the value of each observation is recorded, the likelihood function and the log-likelihood function can be written as:

$$L(\theta) = \prod_{i=1}^{n} f_{X_i}(x_i|\theta), \qquad l(\theta) = \sum_{i=1}^{n} \log f_{X_i}(x_i|\theta).$$

There is no guarantee that the function has a maximum at eligible parameter values. Often, it is not possible to analytically maximize the likelihood function (by setting partial derivatives equal to zero). Numerical approaches are usually needed.

Given independent and identically distributed observations assumed to be from the $dPlN(\alpha, \beta, \mu, \sigma^2)$, one could either fit the dPlN to data $x_1, x_2, ..., x_n$ or fit the NL to $y_1 = \log x_1, ..., y_n = \log x_n$. The maximum likelihood estimates (MLEs) are the same in both cases. If we use the dPlN pdf, the likelihood function is

$$L = \prod_{i=1}^{n} \frac{\alpha\beta}{\alpha+\beta} \left[x_i^{-\alpha-1} e^{\alpha\mu+\frac{1}{2}\alpha^2\sigma^2} \Phi\left(\frac{\log x_i - \mu - \alpha\sigma^2}{\sigma}\right) + x_i^{\beta-1} e^{-\beta\mu+\frac{1}{2}\beta^2\sigma^2} \Phi^c\left(\frac{\log x_i - \mu + \beta\sigma^2}{\sigma}\right)\right].$$

So the log-likelihood function is of the form

$$l = n \log \alpha\beta - n\log(\alpha+\beta) + \sum_{i=1}^{n} \log \left[x_i^{-\alpha-1} e^{\alpha\mu+\frac{1}{2}\alpha^2\sigma^2} \Phi\left(\frac{\log x_i - \mu - \alpha\sigma^2}{\sigma}\right) + x_i^{\beta-1} e^{-\beta\mu+\frac{1}{2}\beta^2\sigma^2} \Phi^c\left(\frac{\log x_i - \mu + \beta\sigma^2}{\sigma}\right)\right]. \tag{2.2.1}$$

The partial derivative with respect to certain parameters requires the derivative of the cumulative distribution function of the normal distribution. The resulting equation cannot be solved analytically. However the above function can be maximized numerically using the method of moments estimates as initial values.

More generally, when the log-transformed data is fitted to a normal-Laplace distribution, the likelihood function is

$$L = \prod_{i=1}^{n} \frac{\alpha\beta}{\alpha+\beta} \phi\left(\frac{y_i - \mu}{\sigma}\right) [R(\alpha\sigma - (y_i-\mu)/\sigma) + R(\beta\sigma + (y_i-\mu)/\sigma)],$$

which yields

$$\begin{aligned} l = n\log\alpha + n\log\beta - n\log(\alpha+\beta) + \sum_{i=1}^{n} \log\phi\left(\frac{y_i-\mu}{\sigma}\right) \\ + \sum_{i=1}^{n} \log[R(\alpha\sigma - (y_i-\mu)/\sigma) + R(\beta\sigma + (y_i-\mu)/\sigma)]. \end{aligned} \tag{2.2.2}$$

This is also a complicated function, one may maximize it numerically by using the method of moments estimates (e.g. with the first four NL cumulants) as starting values.

In general, it is not easy to determine the variance of complicated estimators such as the maximum likelihood estimator. However, it is possible to approximate the variance of the maximum likelihood estimator using the observed information matrix.

Theorem 2.2.1. *(Klugman et al. 2008) Assume that the pdf $f(x;\theta)$ satisfies the following conditions for θ in an interval containing the true value:*

(i) $\log f(x;\theta)$ *is three times differentiable with respect to θ.*

(ii) $\int \frac{\partial}{\partial\theta} f(x;\theta)dx = 0$. *This implies that the derivative may be taken outside the integral and so we are just differentiating the constant 1.*

(iii)$\int \frac{\partial^2}{\partial\theta^2} f(x;\theta)dx = 0$. *This is the same concept as for the second derivative.*

(iv) $-\infty < \int f(x;\theta)\frac{\partial^2}{\partial\theta^2}\log f(x;\theta)dx < 0$. *This establishes that the indicated integral exists and that the location where the derivative is zero is a maximum.*

(v) There exists a function $H(x)$ such that $\int H(x)f(x;\theta)dx < \infty$ with $|\frac{\partial^3}{\partial\theta^3}\log f(x;\theta)| < H(x)$. This makes sure that the population is not overrepresented with regard to extreme values.

Then the following results holds:

(a) As $n \to \infty$, the probability that the likelihood equation $[L'(\theta) = 0]$ has a solution goes to 1.

(b) As $n \to \infty$, the distribution of the maximum likelihood estimator $\hat{\theta}_n$ converges to a normal distribution with mean θ and variance satisfying $\mathcal{I}(\theta)Var(\hat{\theta}_n) \to 1$, where

$$\begin{aligned}\mathcal{I}(\theta) &= -nE\left[\frac{\partial^2}{\partial\theta^2}\log f(x;\theta)\right] = -n\int f(x;\theta)\frac{\partial^2}{\partial\theta^2}\log f(x;\theta)dx \\ &= nE\left[\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right)^2\right] = n\int f(x;\theta)\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right)^2 dx.\end{aligned}$$

Therefore $[\mathcal{I}(\theta)]^{-1}$ is a useful approximation for $Var(\hat{\theta}_n)$. The quantity $\mathcal{I}(\theta)$ is called Fisher's information. Note that the results stated above assume that the sample consists of independent and identically distributed random observations. A more general version of the result uses the logarithm of the likelihood function,

$$\mathcal{I}(\theta) = -nE\left[\frac{\partial^2}{\partial\theta^2}l(\theta)\right] = E\left[\left(\frac{\partial}{\partial\theta}l(\theta)\right)^2\right].$$

If there is more than one parameter, the only change is that the vector of maximum likelihood estimates now has an asymptotic multivariate normal distribution. The covariance matrix of this distribution is obtained from the inverse of the matrix with (r, s)th element,

$$\begin{aligned}\mathcal{I}(\theta)_{rs} &= -E\left[\frac{\partial^2}{\partial\theta_s\partial\theta_r}l(\theta)\right] = -nE\left[\frac{\partial^2}{\partial\theta_s\partial\theta_r}\log f(X;\theta)\right] \\ &= E\left[\frac{\partial}{\partial\theta_r}l(\theta)\frac{\partial}{\partial\theta_s}l(\theta)\right] = nE\left[\frac{\partial}{\partial\theta_r}\log f(X;\theta)\frac{\partial}{\partial\theta_s}\log f(X;\theta)\right].\end{aligned}$$

The first expression on each line is always correct. The second expression assumes that the likelihood is the product of n identical densities. This matrix is called the information matrix. To obtain this matrix, it is necessary to take both derivatives and expected values. This is not always easy to do (for example, in the case of the dPlN distribution). A way to avoid this problem is to replace the pdf by its empirical version. A sample-based version of the Fisher information would be to plug in the observed values rather than calculating the expected value. So the information matrix can be approximated by

$$I(\theta)_{rs} = -\left[\frac{\partial^2}{\partial\theta_s\partial\theta_r}l(\theta)\right],$$

which is called observed information. In practice, since the true value of θ is not known, this matrix is evaluated at the maximum likelihood estimate $\hat{\theta}_n$ to give $I(\hat{\theta}_n)$.

Chapter 3

DOUBLE PARETO-LOGNORMAL DISTRIBUTION WITH COVARIATES

In extreme value analysis for environmental variables, the statistics of extremes, especially those distributions with heavy upper tails, may be very useful. Additionally, the incorporation of covariates into the analysis makes the resultant models both more accurate and physically more realistic.

In the property and casualty insurance industry, the distribution of loss claims related to extreme environmental events, e.g. a cyclone or a flood, tend to be heavy-tailed. However, the distribution of the underlying geophysical phenomenon may not necessarily be heavy-tailed (Reiss and Thomas 2007).

In this chapter, we will fit the distribution of a loss random variable associated with extreme events by the double Pareto-lognormal distribution with an underlying covariate, which may require a transformation to use. We assume that the covariate does not necessarily exhibit a heavy tail behaviour or a linear relation with the loss random variable. The technique of maximum likelihood will also be adopted to estimate parameters.

3.1. The model

Consider a random variable Y with observed data $y_1, ..., y_n$ (e.g. loss caused by fires, hurricanes or floods) assumed to follow the double Pareto-lognormal distribution with density function $f(y)$, such that

$$f(y) = \frac{\alpha\beta}{\alpha+\beta}\left[y^{-\alpha-1}A(\alpha,\mu,\sigma)\Phi\left(\frac{\log y - \mu - \alpha\sigma^2}{\sigma}\right) + y^{\beta-1}A(-\beta,\mu,\sigma)\Phi^c\left(\frac{\log y - \mu + \beta\sigma^2}{\sigma}\right)\right]$$

where

$$A(\theta,\mu,\sigma) = \exp(\theta\mu + \theta^2\sigma^2/2).$$

Suppose that a covariate X is also available, say with observed data $x_1, ..., x_n$ (e.g. floor space, wind speed or precipitation). Since we know that the distribution of the underlying geophysical phenomenon may be heavy-tailed, we suggest using the Box-Cox transformation to transform the covariate in order to simplify the model. Given a value of covariates, say x, the transformed data can be written as

$$x^{(\lambda)} = \begin{cases} \frac{x^\lambda - 1}{\lambda}, & \text{if } \lambda \neq 0 \\ \log x, & \text{if } \lambda = 0. \end{cases}$$

Then the conditional distribution $f(y|x)$ (or $f(y|x^{(\lambda)})$) remains a dPlN distribution, but now with parameters that possibly depend on x (or $x^{(\lambda)}$). We also assume that there is no linear relation between Y and X.

To reduce the number of parameters, we assume that only μ is a function of x. The conditional distribution of Y given x can be written as

$$f(y|x) = \frac{\alpha\beta}{\alpha+\beta}\left[y^{-\alpha-1}A(\alpha,\mu(x),\sigma)\Phi\left(\frac{\log y - \mu(x) - \alpha\sigma^2}{\sigma}\right) + y^{\beta-1}A(-\beta,\mu(x),\sigma)\Phi^c\left(\frac{\log y - \mu(x) + \beta\sigma^2}{\sigma}\right)\right], \tag{3.1.1}$$

where

$$\mu(x) = a + bx^{(\lambda)} \text{ and } A(\theta,\mu(x),\sigma) = e^{\theta\mu(x)+\theta^2\sigma^2/2}.$$

Therefore, we have to deal with the unknown parameters $\lambda, \alpha, \beta, \sigma, a$ and b. The transformation parameter λ could be predetermined by using the method introduced in Appendix A. The estimation of other parameters may be carried out by means of maximum likelihood estimates.

As mentioned in chapter 1, Reed (2004) defined the normal-Laplace distribution (NL) as the convolution of a normal distribution $N(\mu, \sigma^2)$ and an asymmetric Laplace distribution $L(\alpha, \beta)$. Therefore, the normal-Laplace variable

$W \sim NL(\mu, \sigma^2, \alpha, \beta)$ can be represented as

$$W \stackrel{d}{=} N(\mu, \sigma^2) + L(\alpha, \beta).$$

With a covariate x and one parameter μ, depending on x, W can be redefined as

$$(W|X = x) \stackrel{d}{=} N(\mu(x), \sigma^2) + L(\alpha, \beta).$$

Since the double Pareto-lognormal distribution can be considered as an "exponentiated normal-Laplace distribution", i.e. $Y = e^W \sim dPlN(\mu, \sigma^2, \alpha, \beta)$, the estimates of the parameters could be obtained by using the normal-Laplace distribution with log-transformed data $\log Y$. Hence the conditional density function of W given x will be a normal-Laplace density function with μ as a linear function of a and b,

$$g(w|x) = \frac{\alpha\beta}{\alpha+\beta}\phi\left(\frac{y-\mu(x)}{\sigma}\right)[R(\alpha\sigma - (w-\mu(x))/\sigma) + R(\beta\sigma + (w-\mu(x))/\sigma)], \tag{3.1.2}$$

where

$$\mu(x) = a + bx^{(\lambda)}.$$

3.2. Estimation

Consider a data set $y_1, ..., y_n$ (e.g. loss caused by fires, hurricanes or floods) and covariates $x_1, ..., x_n$ (e.g. floor space, wind speed or precipitation); we try to fit the above model to our data. This involves a maximization of six unknown parameters, $\lambda, \alpha, \beta, \sigma, a$ and b. We propose the following steps to estimate these parameters:

•*Step 1.* Fit the double Pareto-lognormal distribution to the data $y_1, ..., y_n$, using first the method of moments estimates of α, β, σ and μ. If there is no real solution to the equations established by (1.3.17), use the log-transformed data $\log y_1, ..., \log y_n$ in order to estimate the parameters with higher order cumulants of the normal-Laplace distribution.

•*Step 2.* Use the method of moments estimates as starting values for the maximum likelihood estimation procedure based on the log-likelihood function (2.2.2). A chi-square test may be applied to check the goodness of fit of the distribution. Note that we do not take account into covariates at this point.

•*Step 3.* Estimate the parameter λ by using the covariate data $x = x_1, ..., x_n$. The estimation of λ is detailed in Appendix A.

•*Step 4.* Assuming that only the parameter μ depends on the covariate x, the log-likelihood function of the dPlN distribution is

$$l = n \log \alpha\beta - n \log(\alpha + \beta) + \sum_{i=1}^{n} \log \left[y_i^{-\alpha-1} A(\alpha, \mu(x), \sigma) \Phi \left(\frac{\log y_i - \mu(x) - \alpha\sigma^2}{\sigma} \right) + y_i^{\beta-1} A(-\beta, \mu(x), \sigma) \Phi^c \left(\frac{\log y_i - \mu(x) + \beta\sigma^2}{\sigma} \right) \right] \tag{3.2.1}$$

where

$$\mu(x) = a + bx^{(\lambda_0)} \text{ and } A(\theta, \mu(x), \sigma) = e^{\theta\mu(x) + \theta^2\sigma^2/2}.$$

If one uses log-transformed data $w_1 = \log y_1, ..., w_n = \log y_n$, the log-likelihood function is

$$l = n \log \alpha + n \log \beta - n \log(\alpha + \beta) + \sum_{i=1}^{n} \log \phi \left(\frac{w_i - \mu(x)}{\sigma} \right) + \sum_{i=1}^{n} \log[R(\alpha\sigma - (w_i - \mu(x))/\sigma) + R(\beta\sigma + (w_i - \mu(x))/\sigma)], \tag{3.2.2}$$

where $\mu(x) = a + bx^{(\lambda_0)}$.

Both log-likelihood equations have to be maximized numerically. We may use maximum likelihood estimates of α, β and σ obtained in step 2 as starting values. The selection of starting values of a and b could be arbitrary, for example, we could try $a = -1$ and $b = 1$ at first. Note that a likelihood ratio test may be used to determine whether adding a covariate to our model is necessary.

In the next chapter, we will illustrate some potential applications of our model in finance and property and casualty insurance industry.

Chapter 4

NUMERICAL ILLUSTRATIONS

In this chapter, the potential application of the double Pareto-lognormal distribution in finance and in property and casualty insurance will be discussed. First, stock price returns will be fitted to the dPlN model. Then, we fit the Danish fire loss data to the dPlN distribution with the floor space as a covariate. Some useful statistical tests will also be conducted in order to analyse how well the model fits the data.

4.1. Application of the model in finance

The logarithmic returns of a stock price $r(t)$ are defined as

$$r(t) = \log(P_{t+1}) - \log(P_t) = \log(P_{t+1}/P_t), \tag{4.1.1}$$

where P_t is the price of a stock at time t. Empirical evidence (see Rydberg, 2000) shows that logarithmic returns tend to follow a distribution with a fatter tail than that of a normal distribution. Therefore the normal-Laplace distribution may be a good alternative to analyse logarithmic returns. Alternatively, the stock price returns, P_{t+1}/P_t could be fitted to the double Pareto-lognormal model.

4.1.1. Description of the data set

Our data set corresponds to daily adjusted closing prices of Bank of Montreal (BMO) ordinary stock from 4 January 2010 to 1 June 2012 (Finance.yahoo.com, 2013). The adjusted closing price is employed to examine historical returns because it gives an accurate representation of the firm's equity value beyond the simple market price. It accounts for all corporate actions such as stock splits, dividends or distributions and rights offerings. Let Y denote the random variable of daily stock price returns of BMO, P_{t+1}/P_t, and $\log Y$ denote the daily logarithmic returns for the BMO stock price. In order to insure that the stock price returns are computed at equally-spaced moments in time, the Monday returns are not taken into account.

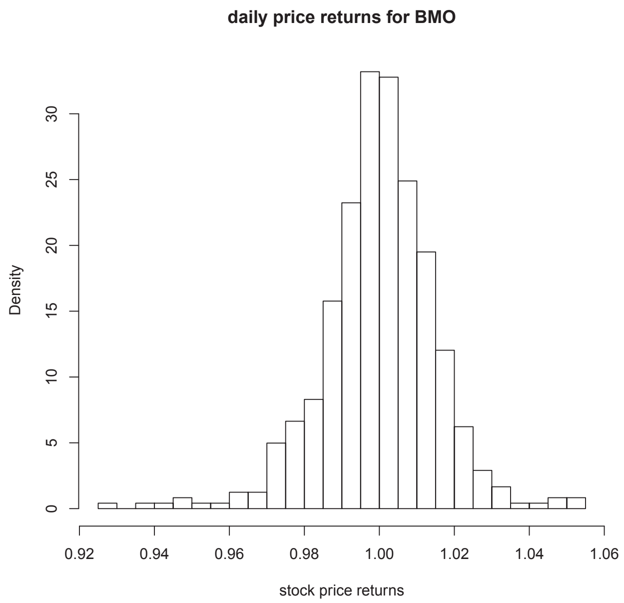

FIGURE 4.1. Histogram of daily stock price returns.

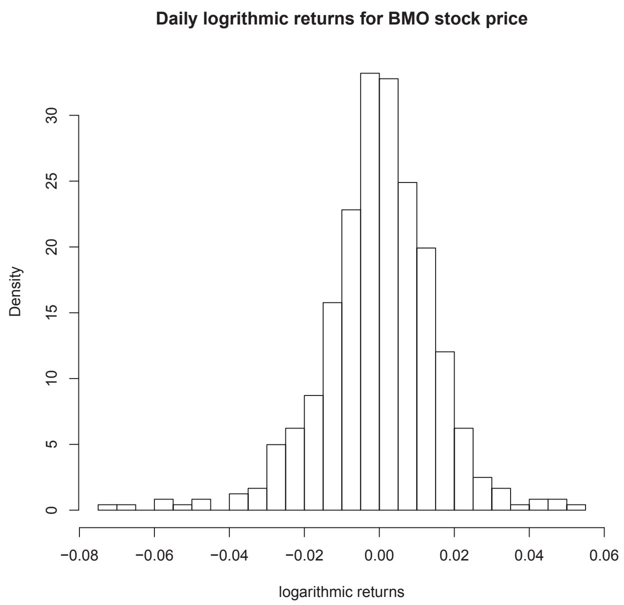

FIGURE 4.2. Histogram of daily logarithmic returns.

First a histogram is sketched to visualize Y, the stock price returns (482 observations). Figure 4.1 shows that stock price returns may follow an asymmetric distribution and this distribution does not seem to be too heavy-tailed. Figure 4.2 gives an histogram of daily logarithmic returns; one can notice that its form is similar to that of daily price returns.

Table 4.1 provides some descriptive statistics produced with **R**. The coefficient of skewness indicates that the distribution is left-skewed. We also observe the kurtosis is smaller than that of the normal distribution. The 5 lowest observations are 0.9288, 0.9354, 0.9422, 0.9456 and 0.9491.

TABLE 4.1. Statistical summary for daily price returns for BMO.

Mean	Std Dev	Skewness	Kurtosis	$25th$	$50th$	$75th$	$95th$
0.99989	0.03097	-0.48954	2.66232	0.99235	1.00028	1.00878	1.02186

4.1.2. Fit of stock price returns

One may first fit the daily price returns to the dPlN distribution directly by using the method of moments to estimate the four parameters. Setting the first four raw moments equal to their theoretical counterparts gives the following equations:

$$\frac{\alpha\beta\exp(\mu+\sigma^2/2)}{(\alpha-1)(\beta+1)} = 0.99989$$
$$\frac{\alpha\beta\exp(2\mu+2\sigma^2)}{(\alpha-2)(\beta+2)} = 1.00000$$
$$\frac{\alpha\beta\exp(3\mu+9\sigma^2/2)}{(\alpha-3)(\beta+3)} = 1.00035$$
$$\frac{\alpha\beta\exp(4\mu+8\sigma^2)}{(\alpha-4)(\beta+4)} = 1.00092$$

As no real solutions could be found with the *NSolve* function in **MATHEMATICA**, we fitted the normal-Laplace distribution to the log-transformed data, i.e. the daily logarithmic returns. We set the four sample cumulants $\kappa_1, \kappa_2, \kappa_4$ and κ_5 of the logarithmic returns (2.1.4, 2.1.5, and 1.2.26) equal to their theoretical counterparts in order to find MMEs of the four parameters, that is,

$$\mu + \alpha^{-1} - \beta^{-1} = -0.00023$$
$$\sigma^2 + \alpha^{-2} + \beta^{-2} = 0.00023$$
$$6\alpha^4 + 6\beta^4 = 1.56106 \times 10^{-7}$$

$$24\alpha^{-5} - 24\beta^{-5} = -4.53149 \times 10^{-9}$$

Using the *NSolve* function of **MATHEMATICA**, we found the method of moments estimates presented in Table 4.2. Note that no real solutions could be obtained with κ_3 and κ_4. These estimates were used as starting values for finding

TABLE 4.2. NL (using MME) fitted to the daily logarithmic returns.

MME	α	β	μ	σ
Estimators	114.2750	83.9290	0.0029	0.0037

the maximum likelihood estimates (MLEs). After maximizing the log-likelihood function (2.2.2) in **R** (Geyer, 2003), we obtain the MLEs listed in Table 4.3. We

TABLE 4.3. NL (using MLE) fitted to the daily logarithmic returns.

MLE	Log-likelihood	α	β	μ	σ
Estimators	-1360.555	117.2474	90.8765	0.0023	0.0056
Standard Errors	N/A	13.65076	7.69729	0.00114	0.00137

also computed an approximation to the asymptotic variance-covariance matrix of the parameter estimates numerically by the observed Fisher information. This matrix is defined as

$$\begin{pmatrix} Var(\hat{\alpha}) & Cov(\hat{\alpha},\hat{\beta}) & Cov(\hat{\alpha},\hat{\mu}) & Cov(\hat{\alpha},\hat{\sigma}) \\ Cov(\hat{\beta},\hat{\alpha}) & Var(\hat{\beta}) & Cov(\hat{\beta},\hat{\mu}) & Cov(\hat{\beta},\hat{\sigma}) \\ Cov(\hat{\mu},\hat{\alpha}) & Cov(\hat{\mu},\hat{\beta}) & Var(\hat{\mu}) & Cov(\hat{\mu},\hat{\sigma}) \\ Cov(\hat{\sigma},\hat{\alpha}) & Cov(\hat{\sigma},\hat{\beta}) & Cov(\hat{\sigma},\hat{\mu}) & Var(\hat{\sigma}) \end{pmatrix},$$

where $\hat{\alpha}, \hat{\beta}, \hat{\mu}$ and $\hat{\sigma}$ are the maximum likelihood estimates.

We obtain the following matrix using **R**

$$\begin{pmatrix} 186.3432347 & 12.31974781 & 0.00999598 & 0.01154041 \\ 12.31974781 & 59.248246677 & -0.004202135 & 0.004770564 \\ 0.00999598 & -0.004202135 & 0.0000013025 & 0.0000002627 \\ 0.01154041 & 0.004770564 & 0.0000002627 & 0.0000018859 \end{pmatrix}.$$

Figure 4.3 suggests that the normal-Laplace distribution fits the logarithmic returns data well. Visually, the fitted pdf (the red line) is a left-skewed distribution which is similar to the form of the empirical pdf (histogram). To test the goodness-of-fit of our model, a chi-square test will be conducted.

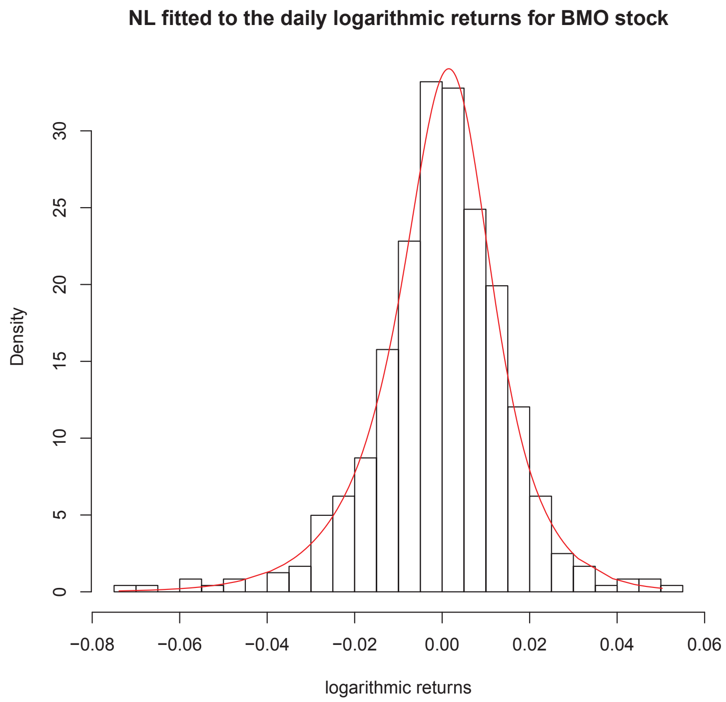

FIGURE 4.3. Normal-Laplace distribution fitted to daily logarithmic returns.

The chi-square test provides us a measure of how close the model distribution function is to the empirical distribution function; the following hypotheses will be tested:

H_0: The daily logarithmic returns come from the normal-Laplace model.

H_1: The data do not come from such a model.

The chi-square test begins with the selection of $k-1$ arbitrary values, $-\infty = c_0 < c_1 < ... < c_k = \infty$. Let $E_i = n(G(c_i) - G(c_{i-1}))$ be the number of expected observations in the interval, where n is the sample size and G is the cdf of the fitted NL distribution (1.2.22). Let $O_i = n(F_n(c_i) - F_n(c_{i-1}))$ be the number of observations in the interval, where F_n is the empirical cdf. The test statistic is then

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

The critical value for this test comes from the chi-square distribution with degrees of freedom equal to the number of classes k minus 1 minus the number of estimated parameters. The null hypothesis H_0 is not rejected if the test statistic is smaller than the critical value.
We set boundaries at -0.05, -0.03, -0.02, -0.01, -0.005, 0, 0.005, 0.01, 0.02, 0.03, 0.05 and infinity. The results are presented in Table 4.4. The test statistic

TABLE 4.4. Chi-square test.

i	Interval from c_{i-1} to c_i	Observed O_i	Expected E_i
1	$(-\infty, -0.05]$	5	2.68354
2	(-0.05, -0.03]	9	13.83799
3	(-0.03, -0.02]	27	24.47228
4	(-0.02, -0.01]	59	60.44447
5	(-0.01, -0.005]	55	55.62738
6	(-0.005, 0]	80	74.66608
7	(0, 0.005]	79	80.64882
8	(0.005, 0.01]	60	66.81625
9	(0.01, 0.02]	77	70.10146
10	(0.02, 0.03]	21	22.57411
11	(0.03, 0.05]	9	9.15687
12	$(0.05, \infty)$	1	0.97075

χ^2 is then 5.8960. With seven degrees of freedom (12 rows minus 1 minus 4 estimated parameters), the critical value for the test at the 0.05 significance level is 14.0671. Since the test statistic is smaller than the critical value, the normal-Laplace distribution is a good fit to the logarithmic returns data, thus the original stock price returns for BMO can be fitted with the double-Pareto lognormal model. Note that in assessing whether a given distribution is suited to a data set, other statistical hypothesis tests such as Kolmogorov-Smirnov test and Anderson-Darling test, can also be used.

4.2. Application of the model in property and casualty insurance

In the property and casualty insurance industry, an insurer must compensate the loss after the payment of an appropriate premium. Actuaries are first of all interested in estimating this net premium which is the mean of the total claim amount for an individual or a portfolio of risks. The double Pareto-lognormal distribution should be useful in modelling the distribution of loss claims of various phenomena, which exhibit a large potential risk, such as floods, fires and hurricanes.

This chapter will discuss the application of the dPlN distribution with an underlying covariate. First we will describe our data set, and then fit the dPlN model to the data; the covariate will be examined and employed to explain the parameter μ of the model. The method of moments and maximum likelihood method can be used to estimate the parameters; some statistical tests will be applied to detect the goodness-of-fit of the model.

4.2.1. Description of the data set

Our data set (Ramlau-Hansen, 1988) consists of 793 fire insurance claims of a Danish insurance company in 1981 (mesured in Danish krone) and the floor space (measured in square meters) associated to the corresponding claim. Let Y denote the random variable of the loss claim amount, and X denote the floor space covariate. First we will use some graphic tools, such as a histogram, to visualize our data set. Note that the observations of Y are already sorted in increasing order.

Figure 4.4 illustrates the histograms by regrouping the loss claim amount Y greater than 100000 in several classes. One can observe a very long tail on the right side of the graphic. If we log-transform the data (say $W = \log Y$), the magnitude of the data values is reduced significantly. The resulting histogram (Figure 4.5) does not resemble a normal distribution, and shows an asymmetric shape.

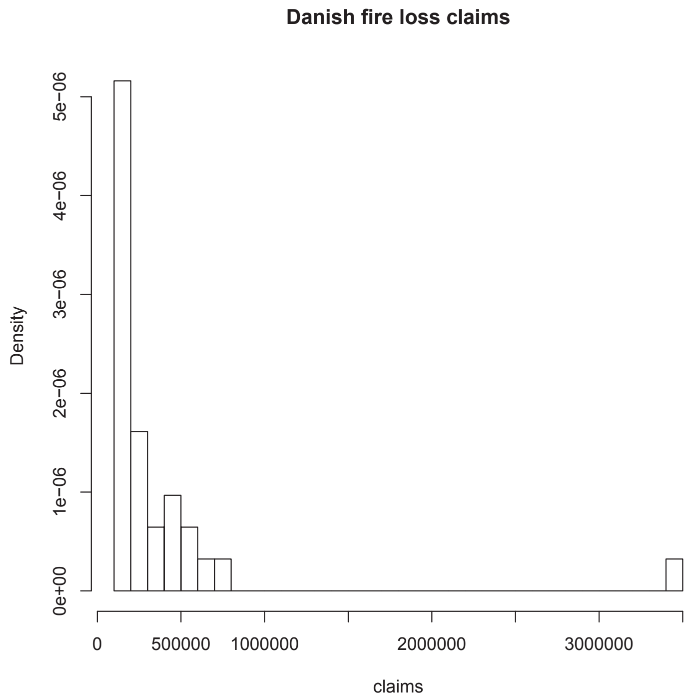

FIGURE 4.4. Histogram of Danish fire loss claims greater than 100000.

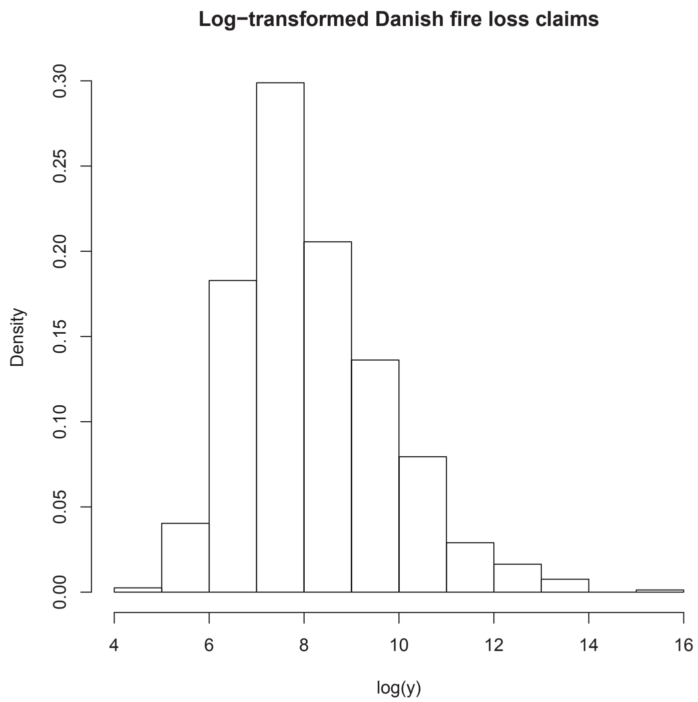

FIGURE 4.5. Histogram of log-transformed Danish fire loss claims.

Table 4.5 provides descriptive statistics produced with **R**; one may observe extreme values from the $99th$ percentile, and the 5 highest observations are 5228877.54, 598389.35, 688498.22, 777477.01 and 3408712.49. Figure 4.6 shows

TABLE 4.5. Statistical summary for Danish fire loss claims.

Mean	Std Dev	$25th$	$50th$	$75th$	$95th$	$99th$
22232.31	135568.13	1143.14	2676.09	9200.43	69553.66	326847.09

the empirical density of the floor space.

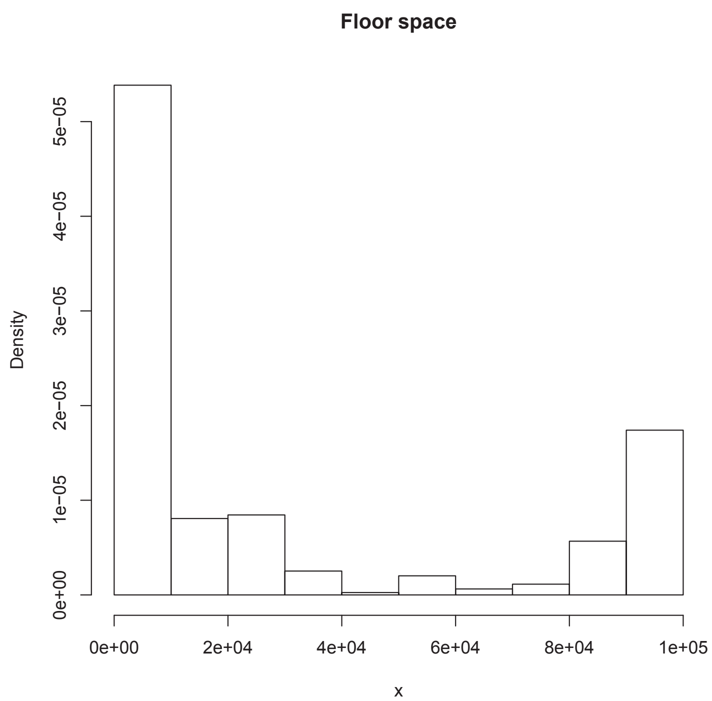

FIGURE 4.6. Histogram of floor space.

Consider a simple linear regression model, $\log y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, where $\log y_i$ and x_i are observations of the log-transformed random variable $\log Y$ and covariate x respectively, β_0 and β_1 are unknown constants, and the residual error is ϵ_i. We may examine whether there is a linear relation between the dependent variable Y and the explanatory variable x.

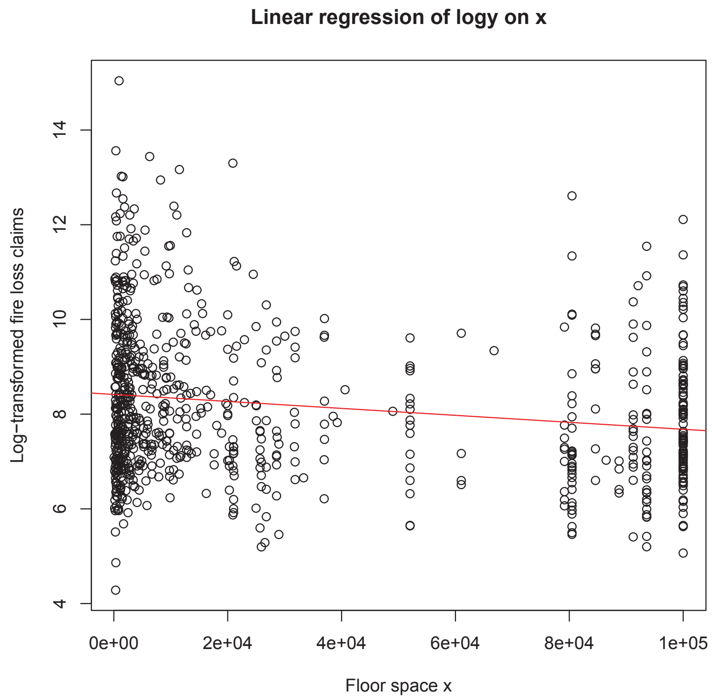

FIGURE 4.7. Scatter plot: log-transformed fire losses vs floor space.

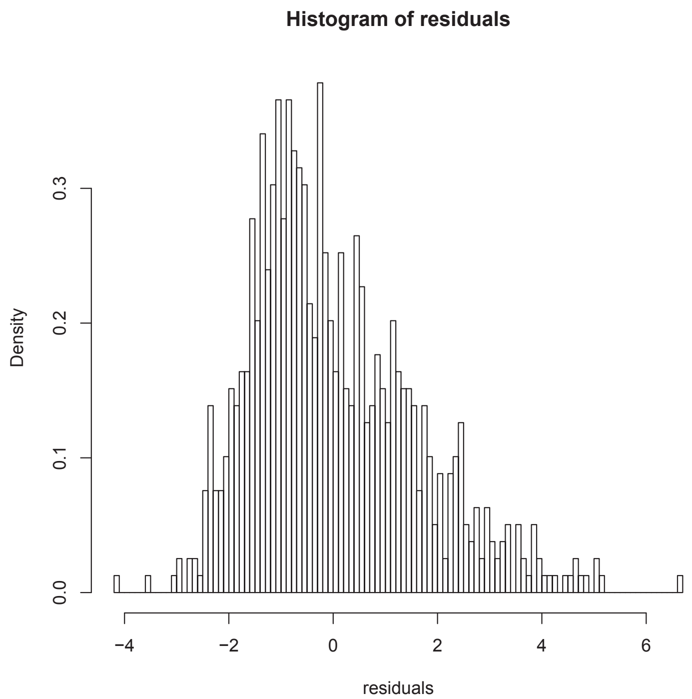

FIGURE 4.8. Histogram of residuals.

Figure 4.7 gives a scatter plot of log-transformed fire losses against floor space, and a regression line is added in order to visualize the fitted simple linear regression model. One can observe that a few points are situated on the regression line.

Before accepting a linear regression model it is important to evaluate its suitability at explaining the data. One of the many ways to do this is to visually examine the residuals. If the model is appropriate, then the residual errors should be random and normally distributed.

According to the histogram of residuals (Figure 4.8), we observe a right-tailed distribution, which suggests that the residuals are not normally distributed. We can also assess normality of residuals with a Q-Q plot. Recall that if the residuals are normally distributed, the points in the normal Q-Q (quantile-quantile) plot will approximately lie on a straight line. From a standard Q-Q plot (Figure 4.9), we notice that there is a strong deviation from a straight line in the upper and lower tails, which implies that the residual errors are not normally distributed.

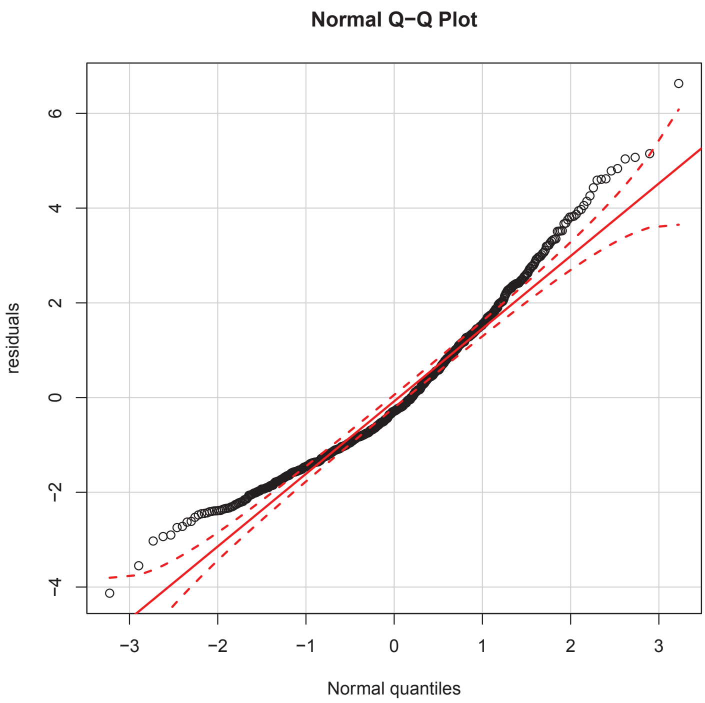

FIGURE 4.9. Normal Q-Q plot of residuals.

Therefore, we can conclude that the simple linear regression model with normal errors is not appropriate for the log-transformed data $\log Y$ with a covariate x.

4.2.2. Application of the model with a covariate

4.2.2.1. *Fit of the fire losses*

We will try to fit the double Pareto-lognormal distribution to the fire loss claims and consider the floor space as an explanatory variable of one parameter of the fitted distribution. Note that we also fit the fire loss claim variable Y to the lognormal distribution and the inverse Gaussian distribution, but both models are not appropriate (see Appendix B). We will now fit Y to the dPlN distribution directly; the method of moment estimate (MME) will be applied to find the four parameters by setting the first four raw moments equal to their theoretical counterparts. We may try to solve the following equations with the *NSolve* function of **MATHEMATICA**:

$$\mu'_1 = 22232.31$$
$$\mu'_2 = 18849817544$$
$$\mu'_3 = 5.197414 \times 10^{16}$$
$$\mu'_4 = 1.714522 \times 10^{23}$$

Unfortunately, we cannot find numerical solutions with these equations. Recall that the kth raw moment exists only for $k < \alpha$. It is possible that the estimate of α is smaller than 4, so that no real solution can be found. In this case, we have to get the random variable Y log-transformed, then fit $W = \log Y$ with the normal-Laplace distribution. We could set four sample cumulants $\kappa_1, \kappa_2, \kappa_4$ and κ_5 of W (2.1.4, 2.1.5, 2.1.7 and 1.2.26) equal to their theoretical counterparts in order to find MMEs of the four parameters, that is,

$$\mu + \alpha^{-1} - \beta^{-1} = 8.19761$$
$$\sigma^2 + \alpha^{-2} + \beta^{-2} = 2.52576$$
$$6\alpha^{-4} + 6\beta^{-4} = 3.99545$$
$$24\alpha^{-5} - 24\beta^{-5} = -5.72187$$

Using **MATHEMATICA**, after eliminating all negative values, we could find the following method of moments estimates of the four parameters as presented in Table 4.6. Note that we replace κ_3 by κ_5 because no real solution is obtained with the third cumulant.

Table 4.6. NL (using MME) fitted to the log-transformed data.

MME	α	β	μ	σ
Estimators	1.48286	1.21486	8.34638	1.18043

We may use the MMEs as starting values for finding the maximum likelihood estimates (MLEs); after maximizing the log-likelihood function (2.2.2) in **R**, we obtain the MLEs listed in Table 4.7. Figure 4.10 allows us to visualize whether the normal-Laplace distribution fits the log-transformed data well or not. Graphically, the fitted pdf (the red line) is a right-skewed distribution without a heavy tail, which resembles the form of the empirical pdf (histogram).

Table 4.7. NL (using MLE) fitted to the log-transformed data.

MLE	Log-likelihood	α	β	μ	σ
Estimators	-1448.182	0.68739	10.86952	6.83502	0.81916

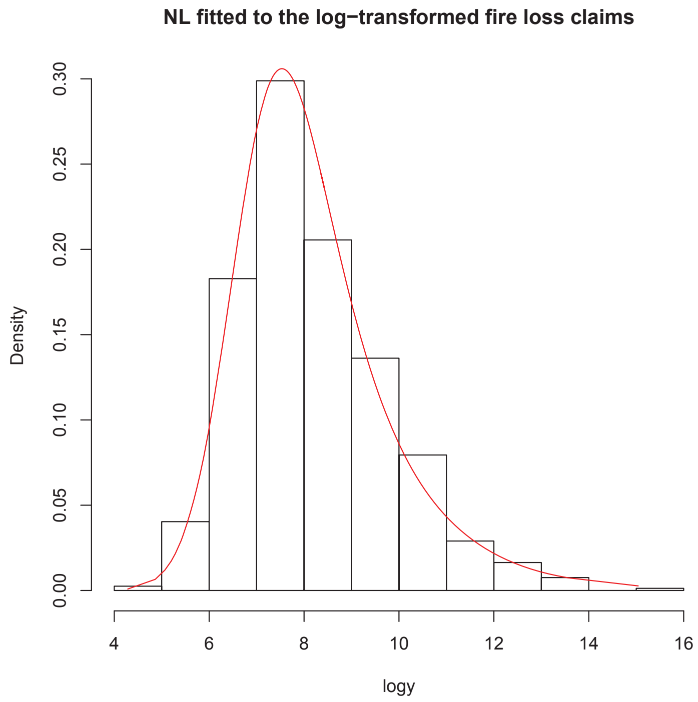

FIGURE 4.10. Normal-Laplace fitted to log-transformed Danish fire loss claims.

We will employ a chi-square test to check the goodness-of-fit of our model. The following hypotheses will be tested:
H_0: The data $\log Y$ come from the normal-Laplace distribution.
H_1: The data $\log Y$ do not come from such model.

We establish boundaries at 5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 10, 11, 12, 13, 15 and infinity. The results appear in Table 4.8.

The χ^2 test statistic is equal to 15.6690. With nine degrees of freedom (14 rows minus 1 minus 4 estimated parameters) the critical value for the test at the 0.05 significance level is 16.919. We conclude that the normal-Laplace distribution provides an acceptable fit to the log-transformed data, thus the original fire loss data can be fitted with the double-Pareto lognormal model.

4.2.2.2. *Inclusion of an explanatory variable*

The parameter μ will now be assumed to be a linear function of a and b, $\mu = a + b \log x$ (see Appendix A); if we use log-transformed data $\log Y$, the log-likelihood function with the form (3.2.2) must be maximized numerically. We

TABLE 4.8. Chi-square test.

i	Interval from c_{i-1} to c_i	Observed O_i	Expected E_i
1	(-∞, 5]	2	2.3648
2	(5, 6]	32	33.0786
3	(6, 6.5]	54	55.0917
4	(6.5, 7]	91	91.2996
5	(7, 7.5]	134	116.2391
6	(7.5, 8]	103	118.3228
7	(8, 8.5]	84	101.5856
8	(8.5, 9]	79	78.1865
9	(9, 10]	108	97.6030
10	(10, 11]	63	49.3253
11	(11, 12]	23	24.8073
12	(12, 13]	13	12.4753
13	(13, 15]	6	9.4287
14	(15, ∞)	1	3.1916

may use maximum likelihood estimates of α, β and σ obtained in Table 4.7 as starting values for the numerical procedure. We use -1 and 1 for the starting values of a and b respectively. We obtain MLEs for the five parameters presented in Table 4.9 and the log-likelihood function is maximized at -1442.099.

One may ask an interesting question: Is it necessary to include a covariate to our double Pareto-lognormal model? As a matter of fact, we only need to test whether the parameter b can be set equal to zero. A likelihood ratio test could help us make this decision. Such a test will be conducted as follows; the

TABLE 4.9. NL (using MLE) fitted to the log-transformed data with a covariate.

MLE	Log-likelihood	α	β	a	b	σ
Estimators	-1442.099	0.69706	9.20529	7.60455	-0.08160	0.81785

hypotheses are:
H_0: The data Y come from the dPlN model (without a covariate) with $b = 0$.
H_1: The data Y came from the dPlN model (with a covariate) with $b \neq 0$.
The test statistic is

$$T = 2(loglikelihood\ for\ alternative\ model - loglikelihood\ for\ null\ model).$$

The null hypothesis is rejected if T is greater than the critical value, which comes from a chi-square distribution with degrees of freedom equal to the number of free parameters in the model under the alternative hypothesis less the number of free parameters in the model under the null hypothesis.

From Table 4.9, the maximized log-likelihood for the alternative model with $\hat{b} = -0.08160$ is -1442.099, while the maximized log-likelihood for the null model is -1448.182 (see Table 4.7). The test statistic is $T = 2(-1442.099 + 1448.182) = 12.166$. The difference between the number of free parameters of the model under H_1 and the model under H_0 is one. For a chi-square distribution with one degree of freedom, the critical value is 3.8415. Because $12.166 > 3.8415$, the null hypothesis is rejected. The probability that a chi-square random variable with one degree of freedom exceeds 12.166 is nearly 0. This indicates strong support to reject the null hypothesis in favor of the dPlN model with a covariate. Note that the dPlN model already provides a good fit to the Danish fire loss data, however, our model can be improved by including the floor space as a covariate.

Chapter 5

CONCLUSION

In this thesis, we review two new related distributions introduced by Reed (2004): the normal-Laplace (NL) distribution and the double Pareto-lognormal (dPlN) distribution and explore the possibility to extend the model with covariates.

We show how to derive the NL distribution from the convolution of a normal distribution and a Laplace distribution, and review its properties studied in the paper of Reed (2004). Based on the fact that the double Pareto-lognormal distribution is related to the normal-Laplace distribution in the same way as the lognormal is related to the normal distribution, we illustrate how the density function of the dPlN distribution can be found by transforming a NL distributed random variable. Several properties proposed by Reed and Jorgensen (2004) are also reviewed and some of them are demonstrated in this thesis.

Besides, we use the methods of moments and maximum likelihood to estimate the parameters of the double Pareto-lognormal distribution. To make the model physically more realistic, we try to include an explanatory variable into our model to create a better model. The Box-Cox power transformation can be employed to utilize the explanatory variable data set. We also show how to estimate the parameters of the new model, relying on the original dPlN by maximum likelihood estimation.

We give two examples to show the large potential of applications of the double Pareto-lognormal distribution and our extended model to real financial and insurance data sets. In the first example, we fit the dPlN distribution to daily stock price returns for Bank of Montreal (BMO) on the New York Stock Exchange (NYSE), and the goodness-of-fit of our model is confirmed by a chi-square test.

In the second example, we model Danish fire losses with the dPlN distribution and consider the floor space as an explanatory variable of the parameter μ in the dPlN distribution. A likelihood ratio test is conducted to justify the fact that the model with a covariate gives a better fit than without it.

As mentioned in the work of Reed and Jorgensen (2004), the usefulness of the double Pareto-lognormal distribution is shown for modeling incomes, particle sizes, settlement sizes, oil-fields and stock price returns, etc. By incorporating explanatory variables into the analysis, the dPlN distribution could be employed to satisfactorily model rare events with potential underlying covariates, such as flood with precipitation, hurricane with wind speed, fire loss claims with floor space or even stock price returns with volume of trade. These other applications of the model would be interesting to explore in further research.

BIBLIOGRAPHY

[1] Box, G. E. P. and Cox, D. R., *An analysis of transformations*, Journal of the Royal Statistical Society, Series B, 26, 211-252, 1964.

[2] Balakrishnan, N., Kotz, S. and Johnson, N. L., *Continuous univariate distributions*, 2nd ed., Wiley, New York, 1995.

[3] Colombi, R., *A new model of income distribution: The Pareto-lognormal distribution*, Income and wealth distribution, inequality and poverty, C. Dagum and M. Zenga, (eds.), Springer, Berlin, 1990.

[4] Finace.yahoo.com, *Historical Prices of Bank of Montreal (BMO) stock*, viewed 1 November 2013. http://finance.yahoo.com/q/hp?s=BMO+Historical+Prices

[5] Geyer, C. J., *Maximum Likelihood in R*, University of Minnesota, 2003. http://www.stat.umn.edu/geyer/5931/mle/mle.pdf

[6] Groparu-Cojocaru, I., *Quadratic distance methods applied to generalized normal Laplace distribution*, Master's thesis, Université de Montréal, 2007.

[7] Groparu-Cojocaru, I. and Doray, L.G., *Inference for the Generalized Normal Laplace Distribution*, Communications in Statistics: Simulation and Computation, 42, 1989-1997, 2013.

[8] Kendall, M.G., Stuart, A., *Kendall's advanced theory of statistics, Volume 1*, 5th edition, Griffin, London. 1987.

[9] Kotz, S., Kozubowski, T. J. and Podgórski, K.,*The Laplace Distribution and Generalizations*, Birkhäuser, Boston, 2001.

[10] Klugman, S. A., Panjar, H. H. and Willmot, G. E., *Loss Models: From Data to Decisions*, 3rd ed., Wiley, New York, 2008.

[11] Osborne, J. W., *Improving your data transformations: Applying the Box-Cox transformation*, Practical Assessment, Research & Evaluation, Vol. 15, No. 12, North Carolina State University, 2010.

[12] Ramlau-Hansen, H., *A solvency study in non-life insurance. Part 1: Analysis of fire, windstorm, and glass claims.*, Scand. Actuar. J., 3-34. [34, 54] 523, 1988.

[13] Reed, W. J., *The normal-Laplace distribution and its relatives. In* Order Statistics and Inference, N. Balakrishna et al. (eds), Birkhäuser, 2004.

[14] Reed, W. J., *Brownian-Laplace motion and its use in financial modelling*, Comm. Stat. A, 36 473-484, 2007.

[15] Reed, W. J. and Jorgensen, M., *The double Pareto-lognormal distribution: A new parametric model for size distribution.*, Com. Stats - Theory & Methods, Vol. 33, No. 8., 1733-1753, 2004.

[16] Reiss, R. D. and Thomas, M., *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields*, 3rd ed., Birkhäuser, Basel, 2007.

[17] Rydberg, T. H., *Realistic Statistical Modelling of Financial Data*, International Statistical Review, Vol. 68, No. 3., pp. 233-258, 2000

[18] Sakia, R. M., *The Box-Cox transformation technique: a review*, The statistician, Vol. 41, pp. 169-178, 1992.

[19] Tukey, G. W., *The comparative anatomy of transformation*, Annals of Mathematical Statistics, 28, 602-632, 1957.

Appendix A

THE BOX-COX TRANSFORMATION

A.1. Definition

Many standard statistical procedures make the assumptions that the variables (or their error terms, more technically) are normally distributed, and the variance of the variables remains constant over the observed range of some other variables, i.e. homoscedasticity or homogeneity of variance. In situations where these assumptions are seriously violated, most researchers may try to design a new model that has important aspects of the original model and satisfies all the assumptions, for example, by applying a proper transformation to the data or filtering out some suspect data points which may be considered outlying.

In our model, we consider data transformations as appropriate tools that can serve many functions in the quantitative analysis of data, including improving normality of a distribution and equalizing variances to meet assumptions. There are as many potential types of data transformations as there are mathematical functions. Some of the more commonly-used traditional transformations include: adding constants, square root, converting to logarithmic scales, inverting and reflecting, etc. Note that all these potential transformations are members of a class of transformation called power transformation. Tukey (1957) is often credited with presenting the initial idea that transformations can be thought of as a class or family of similar mathematical functions. He introduced a family of power transformations from y to $y^{(\lambda)}$, for $y > 0$, such that the transformed values are a monotonic function of the original values over some admissible range and indexed by λ:

$$y^{(\lambda)} = \begin{cases} y^{\lambda}, & \text{if } \lambda \neq 0 \\ \log y, & \text{if } \lambda = 0. \end{cases} \tag{A.1.1}$$

This idea was modified by Box and Cox (1964) to take account of the discontinuity at $\lambda = 0$,

$$y^{(\lambda)} = \frac{y^{\lambda} - 1}{\lambda}, \text{ if } \lambda \neq 0,$$

and at the point $\lambda = 0$, we can write

$$y^{(\lambda)} = \frac{e^{\lambda \log y} - 1}{\lambda} = \frac{(1 + \lambda \log y + \frac{1}{2}\lambda^2 \log(y)^2 + \frac{1}{6}\lambda^3 \log(y)^3 + ...) - 1}{\lambda}$$
$$= \log y + \frac{1}{2}\lambda \log(y)^2 + \frac{1}{6}\lambda^2 \log(y)^3 + ... = \log y, \text{ if } \lambda = 0.$$

Definition A.1.1. *The Box-Cox transformation (Box and Cox 1964) can be defined as :*

$$y^{(\lambda)} = \begin{cases} \frac{y^{\lambda}-1}{\lambda}, & \text{if } \lambda \neq 0 \\ \log y, & \text{if } \lambda = 0. \end{cases} \tag{A.1.2}$$

where $y > 0$ and the transformation parameter λ can take any real values.

The function $y^{(\lambda)}$ defined in (A.1.2) is continuous at $\lambda = 0$.

TABLE A.1. Some traditional transformations in Box-Cox transformation

λ chosen	Name of the transformation
$\lambda = 2$	Square transformation
$\lambda = 1$	No transformation needed
$\lambda = 0.5$	Square root transformation
$\lambda = 0.33$	Cube root transformation
$\lambda = 0$	Natural log transformation
$\lambda = -0.5$	Inverse square root transformation
$\lambda = -1$	Reciprocal (inverse) transformation
$\lambda = -2$	Inverse square transformation

This transformation represents a family of power transformations that incorporates and extends the traditional options (Osborne, 2010) to help researchers easily find the optimal normalizing transformation for a particular variable, eliminating the need to randomly try different transformations to determine the best option (see Table A.1). Precisely, we may enumerate some of the transformations,

$\lambda = 0.5,\ y^{(0.5)} = 2(\sqrt{y} - 1);$
$\lambda = 1,\ y^{(1)} = y - 1;$
$\lambda = 2,\ y^{(2)} = \frac{y^2-1}{2};$
$\lambda = -0.5,\ y^{(-0.5)} = 2(1 - \frac{1}{\sqrt{y}});$
$\lambda = -1,\ y^{(-1)} = 1 - \frac{1}{y};$
$\lambda = -2,\ y^{(-2)} = \frac{1}{2} - \frac{1}{2y^2}.$

Box and Cox (1964) originally envisioned this transformation as a solution for simultaneously correcting normality, linearity and homoscedasticity. While this transformation often improves all of these aspects of a distribution or analysis, Sakia (1992) argued that it does not always accomplish these challenging goals.

A.2. Estimation of λ

Suppose that we observe an $n \times 1$ vector of observations $y = y_1, ..., y_n$, and that the appropriate linear model for the problem is specified by

$$E[y^{(\lambda)}] = X\delta, \tag{A.2.1}$$

where $y^{(\lambda)} = (y_1^{(\lambda)}, ..., y_n^{(\lambda)})$ is the vector of transformed data, X is an observed design matrix, and δ a vector of unknown parameters associated with the transformed observations.

We assume that for some unknown λ, the transformed observations $y_i^{(\lambda)}(i = 1, ..., n)$ are independently and normally distributed with constant variance υ^2, and with expectations (A.2.1). The probability density function for the vector $y^{(\lambda)}$ can be written as

$$f(y^{(\lambda)}) = \frac{\exp(-\frac{1}{2\upsilon^2}(y^{(\lambda)} - X\delta)'(y^{(\lambda)} - X\delta))}{(2\pi\upsilon^2)^{\frac{n}{2}}}.$$

The pdf for the untransformed observations y is obtained by multiplying the normal density by the Jacobian of the transformation $J(\lambda, y)$, which is also the likelihood in relation to the original observations, that is

$$L(\delta, \upsilon^2, \lambda) = f(y|\delta, \upsilon^2, \lambda) = \frac{\exp(-\frac{1}{2\upsilon^2}(y^{(\lambda)} - X\delta)'(y^{(\lambda)} - X\delta))}{(2\pi\upsilon^2)^{\frac{n}{2}}} J(\lambda, y), \tag{A.2.2}$$

where

$$J(\lambda, y) = \prod_{i=1}^{n} \left| \frac{dy_i^{(\lambda)}}{dy_i} \right| = \prod_{i=1}^{n} y_i^{\lambda-1}.$$

From (A.2.2), the log-likelihood can be derived as

$$l(\delta, \upsilon^2, \lambda) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log \upsilon^2 - \frac{1}{2\upsilon^2}(y^{(\lambda)} - X\delta)'(y^{(\lambda)} - X\delta) + \log J(\lambda, y). \tag{A.2.3}$$

By partial derivation, the maximum likelihood estimate of υ^2 can be easily found

$$\frac{dl}{d\upsilon} = (y^{(\lambda)} - X\delta)'(y^{(\lambda)} - X\delta)/\upsilon^3 - n/\upsilon = 0$$

Thus, for given λ, the estimate of υ^2 is denoted

$$\hat{\upsilon}^2(\lambda) = (y^{(\lambda)} - X\delta)'(y^{(\lambda)} - X\delta)/n = S(\lambda)/n \tag{A.2.4}$$

where $S(\lambda)$ is the residual sum of squares in the analysis of variance of $y^{(\lambda)}$. Substituting $\hat{\upsilon}^2(\lambda)$ into the likelihood equation, we only need to maximize the log-likelihood, except for a constant,

$$l(\lambda) = -\frac{1}{2}n\log \hat{\upsilon}^2(\lambda) + (\lambda - 1)\sum_{i=1}^{n}\log y_i. \tag{A.2.5}$$

Then we can plot the maximized log-likelihood $l_{max}(\lambda)$ against λ for a trial series of values of λ. From this plot the maximizing value of $\hat{\lambda}$ may be read off and we can obtain an approximate $100(1-\alpha)$ per cent confident region as well. Note that all these could be done by the *MASS* package in **R**.

A.3. Numerical example

In this section, we will use the Box-Cox transformation just introduced to try to normalize the floor space covariate in section 4.2.1.

The transformation parameter λ will be obtained by maximum likelihood estimation. Using the function *boxcox* in the package *MASS* of **R**, Figure A.1 illustrates a plot of maximized log-likelihood (with likelihood function (A.2.3)) against a series of values of λ ranging from -1 to 1. From this graphic, one can observe that the maximizing value $\hat{\lambda}$ may be close to zero. Precisely, with $\hat{\lambda} = 0.0505$, the log-likelihood (i. e. $\log L(x|\delta, \upsilon^2, \lambda)$) is maximized at -3155.917. But the value 0 is also within the 95 per cent confidence region, thus we suggest taking $\hat{\lambda} = 0$. According to definition A.1.1, the value 0 is equivalent to getting the data log-transformed, i.e. $x^{(\lambda)} = \log x$. This is the transformation we will use in Chapter 4 for the floor space, i.e. we will assume that $\mu(x) = a + b\log x$.

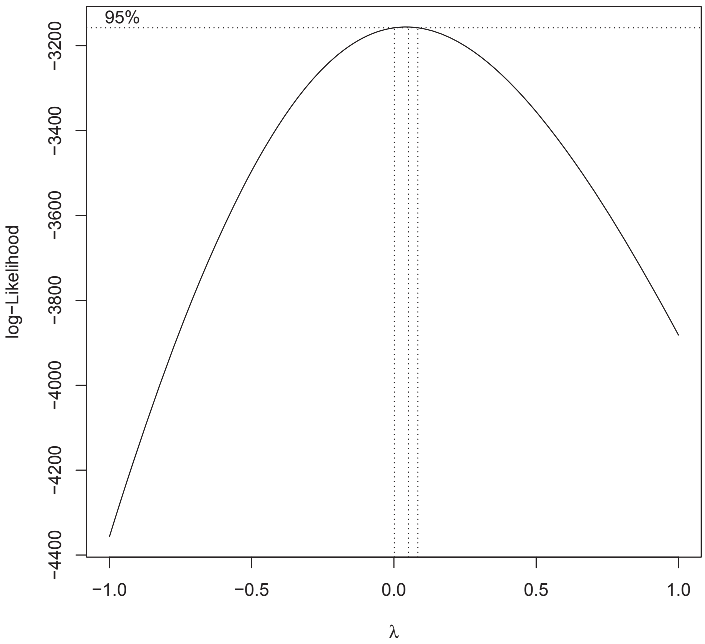

Figure A.1. Estimation of λ.

Appendix B

FIT LOGNORMAL AND INVERSE GAUSSIAN TO THE FIRE LOSS CLAIMS

B.1. Fit lognormal distribution

We will now fit lognormal distribution to the fire loss claim variable Y. Note that the density function of a lognormal distribution with parameters $\mu \in \mathbb{R}$ and $\sigma > 0$ is

$$f(y) = \frac{1}{y\sigma\sqrt{2\pi}} \exp\left[\frac{-(\log y - \mu)^2}{2\sigma^2}\right], \quad y > 0.$$

If $y_1, ...y_n$ are independent and identically distributed and assumed to follow the lognormal distribution with the density function $f(y)$, the likelihood function is

$$L = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \prod_{i=1}^{n} (y_i)^{-1} \exp\left[\frac{-\sum_{i=1}^{n}(\log y_i - \mu)^2}{2\sigma^2}\right],$$

the log-likelihood function is thus

$$l = -n\log(\sigma\sqrt{2\pi}) - \sum_{i=1}^{n} y_i - \frac{\sum_{i=1}^{n}(\log y_i - \mu)^2}{2\sigma^2}. \tag{B.1.1}$$

First, we apply the method of moment estimate in order to find the two parameters by setting the first two raw moments equal to their theoretical counterparts. We may solve the following equations

$$\mu_1' = \exp(\mu + \sigma^2/2) = 22232.31,$$

$$\mu_2' = \exp(2\mu + 2\sigma^2) = 18872993734.$$

We find the following method of moments estimates of the two parameters as presented in Table B.1.

TABLE B.1. Log-normal (using MME) fitted to the Danish fire loss data.

MME	μ	σ
Estimators	8.18811	1.90851

We may use the MMEs as starting values for finding the maximum likelihood estimates; after maximizing the log-likelihood function (B.1.1) in R, we obtain the MLEs listed in Table B.2.

TABLE B.2. Log-normal (using MLE) fitted to the Danish fire loss data.

MLE	Log-likelihood	μ	σ
Estimators	7993.298	8.19761	1.58927

Then, we will employ a chi-square test to check the goodness-of-fit of the log-normal model. The following hypotheses will be tested:
H_0: The data Y come from the lognormal distribution.
H_1: The data Y do not come from such model.

To make the test comparable with the log-transformed data, we establish boundaries at e^5, e^6, $e^{6.5}$, e^7, $e^{7.5}$, e^8, $e^{8.5}$, e^9, e^{10}, e^{11}, e^{12}, e^{13}, e^{15} and infinity. The results appear in Table B.3.

The χ^2 test statistic is equal to 243.4494. With eleven degrees of freedom (14 rows minus 1 minus 2 estimated parameters) the critical value for the test at the 0.05 significance level is 19.675. We conclude that the null hypothesis must be rejected, thus the original fire loss data cannot be fitted with the lognormal model.

TABLE B.3. Chi-square test.

i	Interval from c_{i-1} to c_i	Observed O_i	Expected E_i
1	$(0,\ e^5]$	2	17.5333
2	$(e^5,\ e^6]$	32	48.5762
3	$(e^6,\ e^{6.5}]$	54	47.0691
4	$(e^{6.5},e^7]$	91	65.6880
5	$(e^7,\ e^{7.5}]$	134	83.1001
6	$(e^{7.5},\ e^8]$	103	95.2980
7	$(e^8,\ e^{8.5}]$	84	99.0679
8	$(e^{8.5},\ e^9]$	79	93.3576
9	$(e^9,\ e^{10}]$	108	141.5071
10	$(e^{10},\ e^{11}]$	63	70.9364
11	$(e^{11},e^{12}]$	23	24.2320
12	$(e^{12},\ e^{13}]$	13	5.6379
13	$(e^{13},\ e^{15}]$	6	0.9891
14	$(e^{15},\ \infty)$	1	0.0074

B.2. FIT INVERSE GAUSSIAN DISTRIBUTION

We will now fit the fire loss claim variable Y to the inverse Gaussian distribution. The density function of a log-normal distribution with positive parameters μ and θ is

$$f(y) = \left(\frac{\theta}{2\pi y^3}\right)^{1/2} \exp\left[-\frac{\theta(y-\mu)^2}{2\mu^2 y}\right], \quad 0 < y < \infty.$$

If $y_1, ...y_n$ are independent and identically distributed and assumed to follow the inverse Gaussian distribution, the likelihood function is

$$L = \left(\frac{\theta}{2\pi}\right)^{n/2} \prod_{i=1}^{n} \frac{1}{y_i^{3/2}} \exp\left(-\frac{\theta}{2\mu^2}\sum_{i=1}^{n} y_i + \frac{n\theta}{\mu} - \frac{\theta}{2}\sum_{i=1}^{n}\frac{1}{y_i}\right),$$

and the log-likelihood function is

$$l = \frac{n}{2}\log\left(\frac{\theta}{2\pi}\right) - \frac{3}{2}\sum_{i=1}^{n}\log y_i - \frac{\theta}{2\mu^2}\sum_{i=1}^{n} y_i + \frac{n\theta}{\mu} - \frac{\theta}{2}\sum_{i=1}^{n}\frac{1}{y_i}. \tag{B.2.1}$$

The method of moment estimate will be applied to find the two parameters by setting the mean and the variance equal to their theoretical counterparts. We may solve the following equations

$$\mu_1' = \mu = 22232.31,$$

$$Var(Y) = \mu^3/\theta = 18378718062.$$

We find the following method of moments estimates of the two parameters as presented in Table B.4.

TABLE B.4. Inverse Gaussian (using MME) fitted to the Danish fire loss data.

MME	μ	θ
Estimators	22232.31	597.9139

We may use the MMEs as starting values for finding the maximum likelihood estimates; after maximizing the log-likelihood function (B.2.1) in R, we obtain the MLEs listed in Table B.5.

TABLE B.5. Inverse Gaussian (using MLE) fitted to the Danish fire loss data.

MLE	Log-likelihood	μ	θ
Estimators	7951.285	22232.31	1561.589

We also employ a chi-square test to assess the goodness-of-fit of our model. The following hypotheses will be tested:
H_0: The data Y come from the inverse Gaussian distribution.
H_1: The data Y do not come from such model.

As the previous section, we establish boundaries at e^5, e^6, $e^{6.5}$, e^7, $e^{7.5}$, e^8, $e^{8.5}$, e^9, e^{10}, e^{11}, e^{12}, e^{13}, e^{15} and infinity. The results appear in Table B.6.

The χ^2 test statistic is equal to 172.5658. With eleven degrees of freedom (14 rows minus 1 minus 2 estimated parameters) the critical value for the test at the 0.05 significance level is 19.675. The chi-square test suggests that the null hypothesis should be rejected, therefore the fire loss data cannot be fitted with the inverse Gaussian distribution.

TABLE B.6. Chi-square test.

i	Interval from c_{i-1} to c_i	Observed O_i	Expected E_i
1	$(0, e^5]$	2	1.0034
2	$(e^5, e^6]$	32	40.7748
3	$(e^6, e^{6.5}]$	54	64.8796
4	$(e^{6.5}, e^7]$	91	91.1422
5	$(e^7, e^{7.5}]$	134	101.8204
6	$(e^{7.5}, e^8]$	103	98.7364
7	$(e^8, e^{8.5}]$	84	87.7436
8	$(e^{8.5}, e^9]$	79	73.8593
9	$(e^9, e^{10}]$	108	107.7482
10	$(e^{10}, e^{11}]$	63	65.6022
11	$(e^{11}, e^{12}]$	23	36.6221
12	$(e^{12}, e^{13}]$	13	17.1942
13	$(e^{13}, e^{15}]$	6	5.8670
14	(e^{15}, ∞)	1	0.0065

Appendix C

CODE R AND MATHEMATICA

C.1. Code R

C.1.1. Application to daily logarithmic returns for BMO stock

```
library(e1071)
library(Matrix)
library(car)
library(stats)
library(MASS)
library(graphics)
pr<-read.table("C:/***/BMO.txt",header=T)
y<-pr$PriceReturn
logy<-log(y);
logy<-logy[order(logy, decreasing = FALSE)];

#Graphic#
hist(y,probability=TRUE,xlab="stock price returns",main="daily price
returns for BMO",breaks=40)

hist(logy,probability=TRUE,xlab="logarithmic returns",main="Daily
logrithmic returns for BMO stock price",breaks=40)

#Statistical summary#
library(pastecs)
stat.desc(y)
quantile(y, c(0.0005,.25,.5, .75, .95,.99))
skewness(y)
kurtosis(y)
```

```
#MLE normal-Laplace#
NL.lik1<-function(theta,y){
alpha<-theta[1]
beta<-theta[2]
mu<-theta[3]
sigma<-theta[4]
R1<-pnorm(-y, mean = -(mu+alpha*sigma^2), sd = sigma, lower.tail =
FALSE, log.p = FALSE)/dnorm(alpha*sigma-(y-mu)/sigma, mean = 0,
sd = 1, log = FALSE)
R2<-pnorm(y, mean = mu-beta*sigma^2, sd = sigma, lower.tail = FALSE,
log.p = FALSE)/dnorm(beta*sigma+(y-mu)/sigma, mean = 0, sd = 1,
log = FALSE)
n<-NROW(y)
logl<-n*log(alpha)+n*log(beta)-n*log(alpha+beta)+sum(log(dnorm
((y-mu)/sigma, mean = 0, sd = 1, log = FALSE)))+sum(log(R1+R2))
return(-logl)
}
optim(c(114.275,83.929,0.0029339,0.00372454),NL.lik1,y=logre,
method="BFGS")

#Observed asymptotic variance-covariance matrix#
p<-optim(c(114.275,83.929,0.0029339,0.00372454),NL.lik1,y=logre,
hessian=TRUE,method="BFGS")
VCV<-solve(p$hessian)
VCV

#Graphic#
hist(logy,probability="TRUE",xlab="logarithmic returns",main="NL
fitted to the daily logarithmic returns for BMO stock",breaks=40)
lines(logy, 117.2474*90.87647/(117.2474+90.87647)*dnorm((logy-
0.002253559)/ 0.005646578, mean = 0, sd = 1, log = FALSE)*
(pnorm(-logy, mean = -(0.002253559+117.2474 *0.005646578^2),
sd = 0.005646578, lower.tail = FALSE, log.p = FALSE)/dnorm
(117.2474*0.005646578 -(logy-0.002253559)/ 0.005646578, mean = 0,
sd = 1, log = FALSE)+ pnorm(logy, mean = 0.002253559-
90.87647*0.005646578^2, sd = 0.005646578, lower.tail = FALSE,
log.p = FALSE)/dnorm(90.87647*0.005646578+(logy-0.002253559)
```

```
/0.005646578, mean = 0, sd = 1, log = FALSE)),ylab="probability",
col="red")

#Chi-square test NL#
F<-function(y){
a<-117.2474 #alpha
b<-90.87647 #beta
s<-5.646578e-03 #sigma
m<-2.253559e-03 #mu
pnorm(y,mean=m,sd=s,log=FALSE)-1/(a+b)*dnorm((y-m)/s, mean = 0,
sd = 1, log = FALSE)*(b*pnorm(-y, mean = -(m+a*s^2), sd = s,
lower.tail = FALSE, log.p = FALSE)/dnorm(a*s-(y-m)/s, mean = 0,
sd = 1, log = FALSE)-a*pnorm(y, mean = m-b*s^2, sd = s, lower.tail
 = FALSE, log.p = FALSE)/dnorm(b*s+(y-m)/s, mean = 0, sd = 1,
log = FALSE))}
nl.cut<-cut(logy,breaks=c(-1,-0.05,-0.03,-0.02,-0.01,-0.005,
0,0.005,0.01,0.02,0.03,0.05,1))
table(nl.cut)
(F(-0.05))*482
(F(-0.03)-F(-0.05))*482
(F(-0.02)-F(-0.03))*482
(F(-0.01)-F(-0.02))*482
(F(-0.005)-F(-0.01))*482
(F(0)-F(-0.005))*482
(F(0.005)-F(0))*482
(F(0.01)-F(0.005))*482
(F(0.02)-F(.01))*482
(F(0.03)-F(.02))*482
(F(0.05)-F(0.03))*482
(1-F(0.05))*482

f.ex<-c(2.683535,13.83799, 24.47228, 60.44447, 55.62738, 74.66608,
80.64882, 66.81625, 70.10146, 22.57411, 9.156865, 0.9707547)
f.os<-vector()
for(i in 1:12) f.os[i]<- table(nl.cut)[[i]] #empirical frequencies
X2<-sum(((f.os-f.ex)^2)/f.ex)
print(X2)
```

C.1.2. Application to Danish fire insurance data

```
library(Matrix)
library(car)
library(stats)
library(MASS)
library(graphics)
clsp<-read.table("C:/***/claimdk.txt",header=T)

#X denotes the floor space, Y denotes the fire loss claims#
x<-clsp$space
y<-clsp$claims
logy<-log(y)
logx<-log(x)

#Statistical summary for Danish fire loss claims#
library(pastecs)
stat.desc(y)
quantile(y, c(.25,.5, .75, .95,.99))

#Graphics#
claims<-subset(y,y>100000)
hist(claims,probability=TRUE,main="Danish fire loss claims",
breaks=30)

hist(log(y),probability=TRUE,main="Log-transformed Danish fire
loss claims")

hist(x,probability=TRUE,main="Floor space")

plot(x, logy, xlab="Floor space x",ylab="Log-transformed fire loss
claims",main="Linear regression of logy on x")
abline(lm(logy~x),col="red")

fit<-lm(logy~x)
hist(residuals(fit),main="Histogram of residuals",xlab="residuals",
probability=TRUE,breaks=100)
```

```
qqPlot(residuals(fit), ylab="residuals",xlab="Normal quantiles",
main="Normal Q-Q Plot")


#MLE normal-Laplace#
NL.lik1<-function(theta,y){
alpha<-theta[1]
beta<-theta[2]
mu<-theta[3]
sigma<-theta[4]
R1<-pnorm(-y, mean = -(mu+alpha*sigma^2), sd = sigma, lower.tail =
FALSE, log.p = FALSE)/dnorm(alpha*sigma-(y-mu)/sigma, mean = 0,
sd = 1, log = FALSE)
R2<-pnorm(y, mean = mu-beta*sigma^2, sd = sigma, lower.tail = FALSE,
log.p = FALSE)/dnorm(beta*sigma+(y-mu)/sigma, mean = 0, sd = 1,
log = FALSE)
n<-NROW(y)
logl<-n*log(alpha)+n*log(beta)-n*log(alpha+beta)+sum(log(dnorm((y-mu)
/sigma, mean = 0, sd = 1, log = FALSE)))+sum(log(R1+R2))
return(-logl)
}
optim(c(1.48286,1.21486,8.34638,1.18043),NL.lik1,y=log(y),
method="BFGS")

#Graphic#
hist(logy,probability="TRUE",main="NL fitted to the log-transformed
fire loss claims")
lines(logy,0.6873861*10.8695179/(0.6873861 +10.8695179)*dnorm
((logy-6.8350194)/0.8191595, mean = 0, sd = 1, log = FALSE)*
(pnorm(-logy, mean = -(6.8350194+0.6873861 *0.8191595^2), sd =
 0.8191595, lower.tail = FALSE, log.p = FALSE)/dnorm(0.6873861*
 0.8191595-(logy-6.8350194)/0.8191595, mean = 0, sd = 1,
log = FALSE)+ pnorm(logy, mean = 6.8350194-10.8695179*0.8191595^2,
sd = 0.8191595, lower.tail = FALSE,log.p = FALSE)/dnorm(10.8695179*
0.8191595+(logy-6.8350194)/0.8191595,mean = 0, sd = 1, log = FALSE)),
main="Fire Loss Insurance Claims",ylab="probability",col="red")

#Chi-square test NL#
```

```
clsp<-read.table("C:/***/claimdk.txt",header=T)
y<-clsp$claims
logy<-log(y)
F<-function(y) {pnorm(y,mean=6.8350194,
sd=0.8191595,log=FALSE)-1/(0.6873861+10.8695179)*dnorm
((y-6.8350194)/0.8191595, mean = 0, sd = 1,log = FALSE)*(10.8695179*
pnorm(-y, mean = -(6.8350194+0.6873861*0.8191595^2),sd = 0.8191595,
lower.tail = FALSE, log.p = FALSE)/dnorm(0.6873861*0.8191595-
(y-6.8350194)/0.8191595, mean = 0, sd = 1,
log = FALSE)-0.6873861*pnorm(y, mean = 6.8350194-10.8695179*
0.8191595^2, sd = 0.8191595, lower.tail = FALSE, log.p = FALSE)/
dnorm(10.8695179*0.8191595+(y-6.8350194)/0.8191595, mean = 0,
sd = 1, log = FALSE))}

nl.cut<-cut(logy,breaks=c(0,5,6,6.5,7,7.5,8,8.5,9,10,11,12,13,15,16))
table(nl.cut)
(F(5)-F(0))*793
(F(6)-F(5))*793
(F(6.5)-F(6))*793
(F(7)-F(6.5))*793
(F(7.5)-F(7))*793
(F(8)-F(7.5))*793
(F(8.5)-F(8))*793
(F(9)-F(8.5))*793
(F(10)-F(9))*793
(F(11)-F(10))*793
(F(12)-F(11))*793
(F(13)-F(12))*793
(F(15)-F(13))*793
(1-F(15))*793

f.ex<-c(2.364764,33.0786,55.09176,91.29956,116.2391,118.3228,
101.5856,78.1865,97.60304,49.32528,24.80734,12.47534,9.428688,3.191648)
f.os<-vector()
for(i in 1:14) f.os[i]<- table(nl.cut)[[i]] ## empirical frequencies
X2<-sum(((f.os-f.ex)^2)/f.ex)
print(X2)
ddl<-14-4-1
```

```
#Estimation of lambda#
library(MASS)
bc<-boxcox(x~1, lambda = seq(-1, 1, 0.1))
which.max(bc$y)
lambda <- bc$x[which.max(bc$y)]
lambda

#MLE normal-Laplace WITH A COVARIATE#
NL.lik2<-function(theta,y){
alpha<-theta[1]
beta<-theta[2]
a<-theta[3]
b<-theta[4]
sigma<-theta[5]
R1<-pnorm(-y, mean = -(a+b*logx+alpha*sigma^2), sd = sigma,
lower.tail = FALSE, log.p = FALSE)/dnorm(alpha*sigma-(y-(a+b*logx))/
sigma, mean = 0, sd = 1, log = FALSE)
R2<-pnorm(y, mean = (a+b*logx)-beta*sigma^2, sd = sigma,
lower.tail = FALSE, log.p = FALSE)/dnorm(beta*sigma+(y-(a+b*logx))/
sigma, mean = 0, sd = 1, log = FALSE)
n<-NROW(y)
logl<-n*log(alpha)+n*log(beta)-n*log(alpha+beta)+sum(log
(dnorm((y-(a+b*logx))/sigma, mean = 0, sd = 1, log = FALSE)))+
sum(log(R1+R2))
return(-logl)
}
optim(c(0.6873861,10.8695179 ,-1,1,0.8191595),NL.lik2,y=log(y),
method="BFGS")

#MME and MLE log-normal#
y<-clsp$claims
cl<-y
mean(cl)
var(cl)+(mean(cl))^2
sqrt(log((var(cl)+mean(cl)^2)/(mean(cl))^2))
log(mean(cl))-log((var(cl)+(mean(cl))^2)/(mean(cl))^2)/2
```

```
LN.lik<-function(theta,y){
sig<-theta[1]
mu<-theta[2]
n<-NROW(y)
logl<-(-n)*log(sig*sqrt(2*pi))-sum(log(y))-(1/(2*(sig)^2))*
sum((log(y)-mu)^2)
return(-logl)
}
optim(c(1.908506,8.188105),LN.lik,y=cl,method="BFGS")

#Chi-square test log-normal#
cl<-cl[order(re, decreasing = FALSE)];
cl
F<-function(y){
sig<-1.589271
mu<-8.197611
pnorm((log(y)-mu)/sig, mean = 0, sd = 1, log=FALSE)}
lines(cl,F(cl),col="red")
plot(ecdf(cl))
LN.cut<-cut(cl,breaks=c(0,exp(5),exp(6),exp(6.5),exp(7),exp(7.5),
exp(8),exp(8.5),exp(9),exp(10),exp(11),exp(12),exp(13),exp(15),
exp(16)))
table(LN.cut)
(F(exp(5))-F(0))*793
(F(exp(6))-F(exp(5)))*793
(F(exp(6.5))-F(exp(6)))*793
(F(exp(7))-F(exp(6.5)))*793
(F(exp(7.5))-F(exp(7)))*793
(F(exp(8))-F(exp(7.5)))*793
(F(exp(8.5))-F(exp(8)))*793
(F(exp(9))-F(exp(8.5)))*793
(F(exp(10))-F(exp(9)))*793
(F(exp(11))-F(exp(10)))*793
(F(exp(12))-F(exp(11)))*793
(F(exp(13))-F(exp(12)))*793
(F(exp(15))-F(exp(13)))*793
(1-F(exp(15)))*793
```

```
f.ex<-c( 17.53326,48.57619,47.06909,65.68795,83.10008,95.29799,
99.06794,93.35762,141.5071,70.93641,24.23195,5.637924,0.9890527,
0.007403846)
f.os<-vector()
for(i in 1:14) f.os[i]<- table(LN.cut)[[i]] ## empirical frequencies
X2<-sum(((f.os-f.ex)^2)/f.ex)
print(X2)

#MME and MLE inverse Gaussian#
mean(cl)
var(cl)
(mean(cl))^3/var(cl)
IG.lik<-function(theta,y){
th<-theta[1]
mu<-theta[2]
n<-NROW(y)
logl<-n/2*log(th/2/pi)-3/2*sum(log(y))-th/2/mu^2*sum(y)+n*th/mu
-th/2*sum(1/y)
return(-logl)
}
optim(c(597.9139,22232.31),IG.lik,y=cl,method="BFGS")

#Chi-square test inverse Gaussian#
cl<-cl[order(re, decreasing = FALSE)];
cl
G<-function(y){
th<-1561.589
mu<-22232.310
pnorm(sqrt(th/y)*(y/mu-1), mean = 0, sd = 1, log= FALSE)+
exp(2*th/mu)*pnorm(-sqrt(th/y)*(y/mu+1), mean = 0, sd = 1,
log = FALSE)}
lines(cl,G(cl),col="red")
plot(ecdf(cl))
IG.cut<-cut(cl,breaks=c(0,exp(5),exp(6),exp(6.5),exp(7),exp(7.5),
exp(8),exp(8.5),exp(9),exp(10),exp(11),exp(12),exp(13),exp(15),
exp(16)))
table(IG.cut)
(G(exp(5))-G(0))*793
```

```
(G(exp(6))-G(exp(5)))*793
(G(exp(6.5))-G(exp(6)))*793
(G(exp(7))-G(exp(6.5)))*793
(G(exp(7.5))-G(exp(7)))*793
(G(exp(8))-G(exp(7.5)))*793
(G(exp(8.5))-G(exp(8)))*793
(G(exp(9))-G(exp(8.5)))*793
(G(exp(10))-G(exp(9)))*793
(G(exp(11))-G(exp(10)))*793
(G(exp(12))-G(exp(11)))*793
(G(exp(13))-G(exp(12)))*793
(G(exp(15))-G(exp(13)))*793
(1-G(exp(15)))*793
g.ex<-c(1.003353,40.77476,64.87964,91.14224,101.8204,98.73644,
87.74361, 73.8593,107.7482, 65.60215,36.62212, 17.19424,5.866949,
0.00654503)
g.os<-vector()
for(i in 1:14) g.os[i]<- table(IG.cut)[[i]] ## empirical frequencies
X2<-sum(((g.os-g.ex)^2)/g.ex)
print(X2)
```

C.2. Code MATHEMATICA for obtaining MME

```
(*MME for daily logarithmic returns for BMO*)
(*logx denotes the daily logarithmic returns*)
(*a<-alpha, b<-beta, m<-mu and s<-sigma*)
NSolve[{m + 1/a - 1/b == Mean[logx],
  s^2 + 1/a^2 + 1/b^2 == Moment[logx, 2] - (Mean[logx])^2,
  6/a^4 + 6/b^4 == Cumulant[logx, 4],
  24/a^5 - 24/b^5 == Cumulant[logx, 5]}, {a, b, s, m}, Reals]

(*MME for log-transformed Danish fire loss claims data*)
(*logx denotes the log-transformed data*)
NSolve[{m + 1/a - 1/b == Mean[logx],
  s^2 + 1/a^2 + 1/b^2 == Moment[logx, 2] - (Mean[logx])^2,
  6/a^4 + 6/b^4 == Cumulant[logx, 4],
  24/a^5 - 24/b^5 == Cumulant[logx, 5]}, {a, b, s, m}, Reals]
```