CHAPTER I

INTRODUCTION

1.1 Introduction and Motivation

Many problems in actuarial science involve the building of models that can be used to forecast or predict insurance costs. Modeling is an important procedure for actuaries so that they can estimate the degree of uncertainty as to when a claim will be made and how much will be paid. In particular, the modeling of claims and outstanding claims lead to the pricing of insurance premiums and an estimation of claim reserve, respectively. The most useful approach to uncertainty representation is through probability, so we will concentrate on probability models.

Losses depend on two random variables, i.e., the number of losses and the amount of loss which occur in a specified period. The number of losses (claim number) is referred to as the frequency of loss (claim frequency) and its probability distribution is called *the frequency distribution*. The amount of loss (claim size) is referred to as the severity of loss (claim severity) and its probability distribution is called *the severity distribution*. Loss distribution and its modeling are described in detail in the book of Klugman (2008) and in the papers of Burnecki, Janczura, and Weron (2010). A building of a credible model for claim severity is usually more difficult than for claim frequency. Thus we are interested in claim severity, that is, the severity distribution will be considered in this study.

The mixture of distributions is sometimes called *compounding*, which is extremely important as it can provide a superior fit. A successful use of this technique is illustrated in Hewitt and Lefkowitz (1979). In the 1960s and 1970s, finite mixture models appeared in the statistical literature and they proved to be useful for modeling discrete unobserved heterogeneity in the population. Since there are many different modes of claim possibilities, a finite mixture model should work well.

An Expectations Maximization (EM) algorithm is provided to fit the model that introduces unobserved indicators with the goal of maximizing the complete likelihood functions. The EM algorithm is also applicable for parameter estimation of mixture models. For more details, see Dempster, Laird and Rubin (1977), McLachlan and Peel (2000), Aitkin and Rubin (1985) and Hogg *et al*. (2005).

The bootstrap process is a tool for fitting and it is not complicated to implement. Usually, the bootstrap process involves resampling with replacements from the residual or the data themselves. We apply the bootstrap technique to recalculate the estimated parameters for model fitting. For more details, see Efron and Tibshirani (1993).

An insurance contract is a risk exchange between two parties, i.e., the insurer and the policyholder (insured). The insurer promises to pay for the financial consequences of the claims as the policyholder pays a fixed premium. In this study, the term of risk, in insurance, refers to a loss (claim) variable that quantifies the potential loss (claim) amount associated with an insurance contract. The insurer has understanding to price the premium to cover the uncertainty losses that will occur in the future. So the insurance pricing is therefore important to construct the model for premium calculation.

Risk is often used to mean uncertainty which creates both problems and opportunities for business and individuals. Pure risk exists when there is uncertainty as to whether loss will occur. Speculative risk exists when there is uncertainty about an event that could produce either a profit or a loss. In insurance risk is pure risk that can be insurable, while most of financial risks tend to have the characteristics of speculative risks that are uninsurable. The definitions and properties of risks are explained in the book of James, Robert and David (2005). The risk measures and its classification are described in the book of McNeil, Frey and Embrechts (2004) and the paper of Dhaene *et al*. (2006), in detail. The summarization of risk measure families is shown in Table C.1 of Appendix C. The premium calculation principle is the one of risk measures families that we consider for insurance pricing in this study.

As for insurance premium, the insurer needs not only price it to cover the losses but also to make it competitive in the market. Traditionally, the expected value and the standard deviation are the most widely used to obtain the premium which tends to make it be higher than needed. To provide a competitve premium in the market, we work in the opposite direction. That is, we are interested in how much the premium should be discounted relative to the market price of risk. The premium which is calculated depending on both risk and market conditions, is called *the economic premium*. Then we study economic premium principles for insurance pricing.

1.2 Historical Review

Claim modeling: Many authors have proposed and compared the parameter estimation methods for fitting of claim severity. Some authors investigate some special distributions of the claim severity and apply them to calculate the insurance premium. Grzegorz and Richard (2005) proposed the modeling of hidden exposures in claim severity of normal distribution via the EM algorithm for 2, 3 and 4 components, using the R program. The actual auto bodily injury liability claims closed in Massachusetts in 2001 were applied for the model. Vytaras, Bruce and Ricardas (2009) suggested the method of trimmed moments (MTM) in the case of loss distribution of Lognormal and Pareto and they analyzed real data sets concerning hurricane damage in the United States. Recently, Mohamed, Ahmad and Noriszura (2010) investigated a model of claim severity which has compound Poisson-Pareto distribution, by simulation, and they used it to calculate insurance premiums under the retention limit.

Insurance pricing: In the actuarial literature, there have been many discussions on risk measures of financial and insurance risks in the context of premium calculation principles. Wang's premium principle has been discussed by many authors, e.g., Wang (1995; 1996), Wang, Young and Panjer (1997) and Young (1999). In Wang (2000), the author proposed a pricing method based on the following transform:

$$F^{*}(x) = \Phi\left[\Phi^{-1}(F(x)) + \theta\right]$$

where Φ is the standard normal cumulative distribution and $F(x)$ is the cumulative distribution function (CDF) of a risk interest. The key parameter θ is called *the market price of risk*. The transform is now better known as *the Wang transform* among financial engineers and risk managers. Recently, Kijima and Muromachi (2008) presented an extension of the Wang transform that is consistent with Bühlmann's pricing formula and proposed a new probability transform which is related to the Student's t distribution for pricing of financial and insurance risks.

The purpose of this study is to consider the claim modeling for finite mixture Lognormal distributions and the pricing of insurance premiums based on a new property transform related to finite mixture Lognormal distributions.

1.3 Objective and Overview of the Thesis

The purpose of this study is to find a statistical model for the claim modeling and insurance pricing. For claim modeling, we shall find a model that is fitted to the claim data. Two kinds of distributions are usually considered: one for the amounts of individual claims and the other for amounts of aggregate claims. We are interested in the amount of individual claims. In insurance companies, there are 2 types of claim data recording, i.e., individual and group data. We model the individual claim data in this study. A finite mixture of Lognormal distributions is fitted to the data and the estimated parameters for the model are calculated by the EM algorithm. We also use the bootstrap technique to fit the data and show that the bootstrap sample for observation and residual can be applied to the estimated parameters.

In insurance pricing; we study the premium calculation principle and propose a new transform, called *the Log-transform* that is related to the finite mixture of Lognormal distributions. The premium shall be calculated based on Log-transform and compared with premiums obtained by other methods.

Our work is organized as follows: In Chapter II, we present preliminaries which are useful for claim modeling and insurance pricing, some mathematical and statistical background are also shown in this section. In Chapter III, we present the claim modeling. That is, we present the statistical modeling for a finite mixture of Lognormal distributions, the EM algorithm is explained and the bootstrap technique is demonstrated. We have performed numerical experiments of empirical data for fitting by the finite mixture of Lognormal distributions. An application with actual claim data set is given in this chapter. In Chapter IV, we present the insurance premium calculation which is price based on the Log-transform related to the finite mixture Lognormal distributions. We show that the Log-transform can be derived from Bühlmann’s economic premium principle. The insurance pricing based on Log-transform is applied to the actual claim data set. The conclusions, discussion and further research are shown in Chapter V.

CHAPTER II

PRELIMINARIES

In this section, the concepts and theories of some mathematical and statistical material are presented that is useful for the claim modeling and insurance pricing. Some of the probabilistic tools are described in Appendix B.

2.1 Random Variables

Losses of insurance are losses caused by occurrences of unexpected events. Examples of insured events and their consequences are damage to property and casualties by fire, theft, flood, hail, accident, disability or death (loss of future income and support), illness (cost of medical treatment) and personal injury resulting from accidents or medical malpractice (cost of treatment and personal suffering).

Mostly, actuaries are interested in some consequences of random outcomes. For example, they are concerned with the amount which the insurance company will pay for claim possibilities. We can think of them as functions mapping insured events into the real line $\mathbb{R}$ (claim amount). Such functions are called *random variables* provided they satisfy certain desirable properties, precisely stated in the following definition:

Definition 2.1. If Ω is a given set, then a σ- algebra $\mathcal{F}$ on Ω is a family $\mathcal{F}$ of subsets of Ω with the following properties:

(i) $\varnothing \in \mathcal{F}$

(ii) $F \in \mathcal{F} \Rightarrow F^c \in \mathcal{F}$, where $F^c = \Omega \setminus F$ is the complement of F in Ω

(iii) $A_1, A_2, \ldots \in \mathcal{F} \Rightarrow A := \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

The pair $(\Omega, \mathcal{F})$ is called *a measurable space*. A probability measure P on a measurable space $(\Omega, \mathcal{F})$ is a function $P : \mathcal{F} \to [0,1]$ such that

(a) $P(\varnothing) = 0$, $P(\Omega) = 1$

(b) if $A_1, A_2, \ldots \in \mathcal{F}$ and $\{A_i\}_{i=1}^{\infty}$ is disjoint (i.e., $A_i \cap A_j = \varnothing$ if $i \neq j$) then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

The triple $(\Omega, \mathcal{F}, P)$ is called *a probability space*.

The subsets A of Ω which belong to $\mathcal{F}$ are called $\mathcal{F}$ - *measurable sets*. In a probability context these sets are called events and we use the interpretation

$P(A)$ = " the probability that the event A occurs"

If $(\Omega, \mathcal{F}, P)$ is a given probability space, then a function $Y : \Omega \to \mathbb{R}^n$ is called $\mathcal{F}$ - *measurable* if

$$Y^{-1}(U) := \{\omega \in \Omega ; Y(\omega) \in U\} \in \mathcal{F}$$

for all open sets $U \in \mathbb{R}^n$.

If $X : \Omega \to \mathbb{R}^n$ is any function, then the σ - algebra $\mathcal{H}_X$ generated by X is the smallest σ - algebra on Ω containing all the sets

$$X^{-1}(U) \ ; \ U \subset \mathbb{R}^n \text{ open.}$$

That is $\mathcal{H}_X = \{X^{-1}(B); B \in \mathcal{B}\}$, where $\mathcal{B}$ is the Borel σ - algebra on $\mathbb{R}^n$.

A random variable X is an $\mathcal{F}$ - measurable function mapping Ω to the real numbers, i.e., $X : \Omega \to \mathbb{R}$ is such that

$$X^{-1}((-\infty, x]) \in \mathcal{F} \quad \text{for any } x \in \mathbb{R},$$

where $X^{-1}((-\infty, x]) = \{\omega \in \Omega \mid X(\omega) \le x\}$. Every random variable induces a probability measure μ_X on $\mathbb{R}$, defined by

$$\mu_X(B) = P(X^{-1}(B)).$$

μ_X is called *the distribution of* X.

The actuary deals with objects such as random variables. An example of a random variable is the amount of a claim associated with the occurrence of an automobile accident.

2.2 Distribution Functions

To each random variable X is associated a function F_X called *the distribution function of* X or the cumulative distribution function (CDF) of X. The distribution F_X does not indicate what is the actual outcome of X, but shows how the possible values for X are distributed. The CDF of the random variable X is defined as

$$F_X(x) = P[X^{-1}((-\infty, x])] \equiv P[X \le x],\ x \in \mathbb{R}.$$

$F_X(x)$ represents the probability that the random variable X assumes a value that is less than or equal to x. If X is the total amount of claims generated by some policyholder, $F_X(x)$ is the probability that this policyholder produces a total claim amount of at most x Thai Baht.

Any distribution function F has the following properties:

(i) F is nondecreasing , i.e., If $x < y$ then $F(x) \le F(y)$.

(ii) $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to +\infty} F(x) = 1$.

(iii) F is right-continuous, that is, $\lim_{h \to 0^+} F(x+h) = F(x)$ for all $x \in \mathbb{R}$.

Definition 2.2. A random variable X is called *discrete* if it takes values in some countable subset $\{x_1, x_2, ...\}$ of $\mathbb{R}$. The discrete random variable X has *probability mass function* $f : \mathbb{R} \to [0,1]$ given by

$$f(x) = P(X = x).$$

Definition 2.3. A random variable X is called *continuous* if its distribution function can be expressed as

$$F(x) = \int_{-\infty}^{x} f(u)du \quad ; \quad x \in \mathbb{R},$$

for some integrable function $f : \mathbb{R} \to [0,1]$ called *the probability density function* (PDF) of X.

Definition 2.4. Suppose that $X_i, i = 1, 2, ..., n$ are random variables on a probability space $(\Omega, \mathcal{F}, P)$. They can be composed to a random vector in $\mathbb{R}^n$ is defined by

$$\mathbf{X} = (X_1, X_2, \ ..., \ X_n)'.$$

Definition 2.5. The expectation of a continuous random variable X with density function f is given by

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx$$

whenever this integral exists.

Definition 2.6. The variance of a continuous random variable X with density function f is given by

$$Var[X] = E[(X - E[X])^2].$$

We can rewrite as $$Var[X] = E[X^2] - (E[X])^2.$$

Theorem 2.1. If X has density function f with $f(x) = 0$ when $x < 0$, and distribution function F, then the expected value of X is

$$E[X] = \int_0^\infty [1 - F(x)]\, dx.$$

Proof:

$$\int_0^\infty \left[1 - F(x)\right] dx = \int_0^\infty P(X > x)\,dx$$

$$= \int_0^\infty \left(\int_{y=x}^\infty f(y)\,dy \right) dx$$

$$= \int_0^\infty \left(\int_0^y f(y)\,dx \right) dy$$

$$= \int_0^\infty (y - 0) f(y)\,dy$$

$$= \int_0^\infty y f(y)\,dy$$

Conclusion that

$$E[X] = \int_0^\infty [1 - F(x)]\, dx. \qquad \square$$

Definition 2.7. Let X be a continuous random variable with density function f. The moment generating function (MGF) of the random variable X is the function $M:\mathbb{R}\to[0,\infty)$ given by $M_X(t)=E(e^{tX})$. That is,

$$M_X(t)=E\left[e^{tX}\right]=\int_{-\infty}^{\infty} e^{tx}\,dF(x)=\int_{-\infty}^{\infty} e^{tx}\,f(x)\,dx\,.$$

Example. If $X \sim N(\mu,\sigma^2)$ then $E\left[e^{rX}\right]=\exp\left(\mu r+\frac{1}{2}r^2\sigma^2\right)$. In the special case when $X \sim N\ 0,1$ we have $M_X(t)=E\left[e^{tX}\right]=e^{t^2/2}$.

2.3 Lognormal Distribution

Lognormal distribution is useful as a model for the claim size distributions. A random variable X is said to have the Lognormal distribution with parameters μ and σ if $Y=\ln X$ has the normal distribution with mean μ and standard deviation σ. We assume that the random variable X representing claim size has the Lognormal distribution with parameters μ and σ.

Assume that $X \sim$ Lognormal(μ,σ), abbreviated $X \sim LN(\mu,\sigma)$.

CDF : $F_X(x)=\Phi\left(\dfrac{\ln x-\mu}{\sigma}\right)$; $\mu\in\mathbb{R}$, $\sigma>0$ and $x>0$.

PDF : $f_X(x)=\dfrac{1}{x\sigma\sqrt{2\pi}}\exp\left(-\dfrac{\ln x-\mu^{\ 2}}{2\sigma^2}\right)$

Moment : $E[X^k]=\exp\left(k\mu+\frac{1}{2}k^2\sigma^2\right)$

Mean : $\exp\left(\mu + \frac{1}{2}\sigma^2\right)$

Median : $\exp(\mu)$

Variance : $\left[\exp(\sigma^2) - 1\right]\left[\exp(2\mu + \sigma^2)\right]$

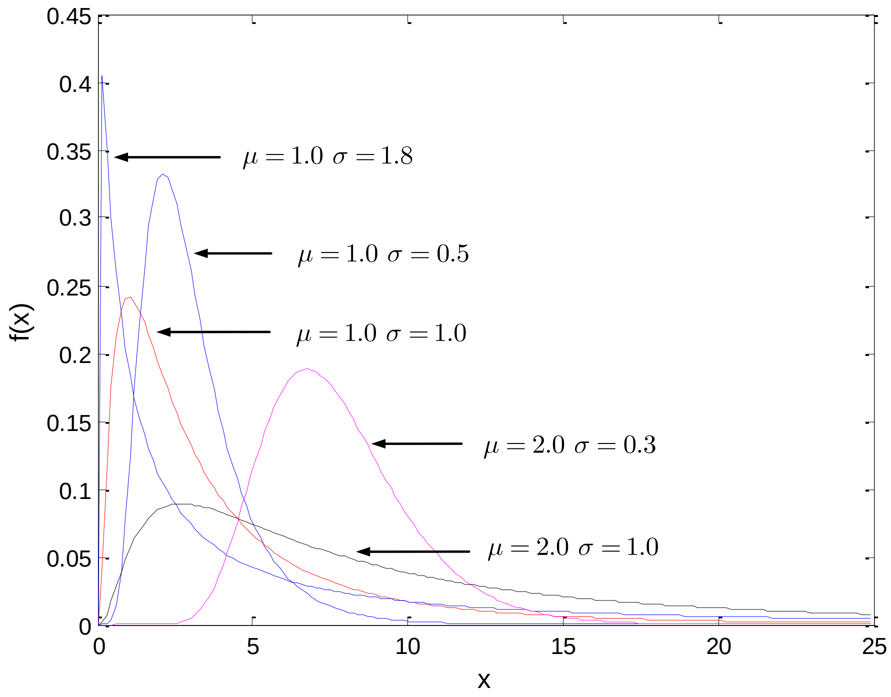

Figure 2.1 The PDF of the Lognormal distribution.

2.4 Uniform Distribution

The random variable X has the uniform distribution with parameters α and β, abbreviated $X \sim Uni(\alpha, \beta)$, if its density function is given as follows:

PDF : $$f_X(x) = \begin{cases} \dfrac{1}{(\beta - \alpha)} & , \alpha \le x \le \beta \\ 0 & \text{elsewhere.} \end{cases} \quad , \quad \alpha < \beta.$$

Example: $X \sim Uni(0,1)$.

$$\text{PDF} \quad : \quad f_X(x) = \begin{cases} 1 & , x \in (0,1) \\ 0 & \text{elsewhere.} \end{cases}$$

$$\text{CDF} \quad : \quad F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } 0 \le x < 1 \\ 1 & \text{if } x \ge 1. \end{cases}$$

Lemma 2.1. Suppose X has a continuous and strictly increasing CDF F. Then $F(X)$ has the uniform distribution,

$$F(X) \sim Uni(0,1).$$

Proof:

Let $u \in (0,1)$.

$$\begin{aligned} P[F(X) \le u] &= P[F^{-1}F(X) \le F^{-1}(u)] \\ &= P[X \le F^{-1}(u)] \\ &= F(F^{-1}(u)) \\ &= u. \end{aligned}$$

The lemma has been proved. □

Note that above we have used:

(1) F is strictly increasing and continuous $\Rightarrow F^{-1} : (0,1) \to \mathbb{R}$ exists.

(2) $F^{-1}(F(x)) = x$, $\forall x \in \mathbb{R}$.

(3) $F(F^{-1}(x)) = x$, $\forall x \in (0,1)$.

Corollary 2.1. Let X be a random variable with continuous and strictly increasing CDF F and Φ be the standard normal distribution. If $V = \Phi^{-1}\left[F(X)\right]$, then V has distribution Φ, i.e.,

$$P(V \leq x) = \Phi(x).$$

Proof:

Let $x \in \mathbb{R}$, one has:

$$P(V \leq x) = P[\Phi^{-1}(F(X) \leq x]$$

$$= P[F(X) \leq \Phi(x)].$$

By Lemma 2.1, $F(X) \sim Uni(0,1)$.

Conclusion that

$$P(V \leq x) = \Phi(x), \;\; V \sim N(0,1). \qquad \square$$

2.5 Mixture Models

A mixture model is a discrete or continuous weighted combination of distributions and represents a heterogeneous population comprised of two or more distinct subpopulations. The source of heterogeneity could be gender, age, mode of benefit payment, etc.

2.5.1 The Finite Mixture Models

A finite mixture model allows us to combine two or more characteristics into one model. It can be represented by a probability density function (PDF) of the form:

$$f(x) = \tau_1 f_1(x) + \cdots + \tau_k f_k(x)$$

with $x \in \mathbb{R}$, $\tau_j > 0$ for $j = 1, \ldots, k$ and $\tau_1 + \cdots + \tau_k = 1$.

All $f_k(\cdot)$ are PDF (either continuous or discrete). The τ_k are called *the mixing weights* (mixing values) and the $f_k(x)$ are called *the components*, k is the number of component distributions of the mixture. In most situations, the $f_k(\cdot)$ have specified parametric forms:

$$f(x) = \tau_1 f_1(x \mid \theta_1) + \cdots + \tau_k f_k(x \mid \theta_k),$$

where θ_j denotes the vector of parameters in density $f_j(\cdot)$ for $j = 1, \ldots, k$.

2.6 Random Vector and Covariance

Definition 2.8. The joint distribution function of random variables X and Y is the function $F : \mathbb{R}^2 \to [0,1]$ given by

$$F(x, y) = P(X \leq x, Y \leq y).$$

Definition 2.9. The random variables X and Y are (jointly) continuous with joint probability density function $f : \mathbb{R}^2 \to [0, \infty]$ if

$$F(x, y) = \int_{v=-\infty}^{y} \int_{u=-\infty}^{x} f(u, v)\, du\, dv, \text{ for each } x, y \in \mathbb{R}.$$

From here on, let X, Y be random variables with joint PDF $f(x, y)$. Then the marginal distribution functions of X and Y are

$$F_X(x) = P(X \leq x) = \lim_{y \to \infty} F(x, y) \text{ and } F_Y(y) = P(Y \leq y) = \lim_{x \to \infty} F(x, y),$$

respectively. Hence,

$$F_X(x) = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(x,y)\,dy\,dx\ ,\ F_Y(y) = \int_{-\infty}^{y} \int_{-\infty}^{\infty} f(x,y)\,dxdy$$

and it follows that the marginal density functions of X and Y are

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy \text{ and } f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx\text{, respectively.}$$

Definition 2.10. Suppose that $g : \mathbb{R}^2 \to \mathbb{R}$ is a continuous function. If X and Y are continuous random variables with joint probability density function f, then the expected value of the random variable $g(X,Y)$ is given by

$$E\left[g(X,Y)\right] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y)\,f(x,y)\,dx\,dy\ .$$

Definition 2.11. If X and Y are random variables, the covariance of X and Y is

$$Cov\,[X,Y] = E\Big[(X - E[X])(Y - E[Y])\Big].$$

It can be rewritten as

$$Cov[X,Y] = E[XY] - E[X]E[Y].$$

The correlation (coefficient) of X and Y is

$$Corr[X,Y] = \frac{Cov[X,Y]}{\sqrt{Var[X]\,Var[Y]}} = \frac{Cov[X,Y]}{\sigma_X \sigma_Y}$$

as long as the variances are non-zero.

Lemma 2.2. Let V be a random variable which has the standard normal distribution, $V \sim N(0,1)$. Then for every $\theta \in \mathbb{R}$, $Cov[V, -\theta V] = -\theta$.

Proof:

$$Cov(V, -\theta V) = E[V(-\theta V)] - E[V]E[-\theta V]$$

$$Cov[V, -\theta V] = -\theta E[V^2] + \theta E[V]E[V]$$

$$= -\theta[E[V^2] - (E[V])^2]$$

$$= -\theta Var[V]$$

$$= -\theta \ . \qquad \square$$

Theorem 2.2. Suppose that X_1 and X_2 are normal and independent. Then $X_1 + X_2$ is normal.

Lemma 2.3. For $j = 1, \ldots, k$, suppose that random variables X_j are independent and let $g_j : \mathbb{R} \to \mathbb{R}$, be continuous functions. Then the random variables $g_j(X_j)$, $j = 1, \ldots, k$ are also independent.

Definition 2.12. Let random variables (X, Y) have the joint PDF

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right\},$$

where $-\infty < x < \infty$, $-\infty < y < \infty$, $-\infty < \mu_X < \infty$, $-\infty < \mu_Y < \infty$, σ_X, $\sigma_X > 0$ and $-1 < \rho < 1$. Then X, Y are said to have a bivariate normal distribution, and $E[X] = \mu_X$, $E[Y] = \mu_Y$, $Var[X] = \sigma_X^2$, $Var[Y] = \sigma_Y^2$, $Cov[X,Y] = \rho\sigma_X\sigma_Y$ and $Corr[X,Y] = \rho$.

Definition 2.13. The joint moment generating function of (X, Y) is defined by

$$M_{X,Y}(t_1, t_2) = E\left[e^{t_1X + t_2Y}\right]$$

and the moment generating function (MGF) for the bivariate normal distribution is

$$M_{X,Y}(t_1,t_2)=\exp\left(t_1\mu_X+t_2\mu_Y+\frac{1}{2}(t_1^2\sigma_X^2+2\rho t_1t_2\sigma_X\sigma_Y+t_2^2\sigma_Y^2)\right),$$

where $E[X]=\mu_X$, $E[Y]=\mu_Y$, $Var[X]=\sigma_X^2$, $Var[Y]=\sigma_Y^2$, $Cov[X,Y]=\rho\sigma_X\sigma_Y$

and $Corr[X,Y]=\rho$.

Lemma 2.4. Suppose X,Y is bivariate normal then

$$M_{X,Y}(s,-1)=E\left[e^{-Y}\right]\exp\left(sE[X]+\frac{s^2}{2}Var[X]-sCov[X,Y]\right).$$

Proof:

By MGF for the bivariate normal distribution, one gets

$$M_{X,Y}(s,t)=\exp\left(s\mu_X+t\mu_Y+\frac{1}{2}(s^2\sigma_X^2+2\rho st\sigma_X\sigma_Y+t^2\sigma_Y^2)\right).$$

$$M_{X,Y}(s,-1)=\exp\left(s\mu_X-\mu_Y+\frac{1}{2}(s^2\sigma_X^2-2\rho s\sigma_X\sigma_Y+\sigma_Y^2)\right)$$

$$=\exp\left(sE[X]+\frac{s^2}{2}Var[X]-E[Y]+\frac{1}{2}Var[Y]-\rho s\sigma_X\sigma_Y)\right)$$

$$=\exp\left(sE[X]+\frac{s^2}{2}Var[X]-E[Y]+\frac{1}{2}Var[Y]-sCov[X,Y]\right)$$

$$=\exp\left(-E[Y]+\frac{1}{2}Var[Y]\right)\exp\left(sE[X]+\frac{s^2}{2}Var[X]-sCov[X,Y]\right).$$

The MGF of the univariate random variable of normal distribution is

$$\eta(s)=M_Y(s)=\exp\left(s\mu_Y+\frac{1}{2}s^2\sigma_Y^2\right). \tag{2.5}$$

If $s=-1$, then $\eta(-1)=M_Y(-1)=\exp\left(-\mu_Y+\frac{1}{2}\sigma_Y^2\right)=E[e^{-Y}]$.

Conclusion that

$$M_{X,Y}(s,-1) = E\left[e^{-Y}\right]\exp\left(sE[X] + \frac{s^2}{2}Var[X] - sCov[X,Y]\right). \qquad \square$$

Lemma 2.5. Suppose that (X,Y) is jointly normally distributed. Then

$$E\left[e^{-Y}f(X)\right] = E\left[e^{-Y}\right]E\left[f(X - Cov[X,Y])\right]$$

for any $f(x)$ for which the above expectation exists.

Proof:

Let $\xi(x,y)$ be the joint density of (X,Y) and define

$$\xi_X(x) = \int_{-\infty}^{\infty} e^{-y}\xi(x,y)\,dy \quad , -\infty < x < \infty.$$

Then

$$E\left[e^{-Y}f(X)\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-y}f(x)\xi(x,y)\,dx\,dy \;=\; \int_{-\infty}^{\infty} f(x)\xi_X(x)\,dx\,.$$

Denoting the MGF of (X,Y) by

$$\eta(s,t) = E\left[e^{sX+tY}\right]$$

one obtains that

$$\eta(s,-1) = E\left[e^{sX-Y}\right] = \int_{-\infty}^{\infty} e^{sx}\,\xi_X(x)\,dx\ . \qquad (2.6)$$

Since

$$E\left[e^{sX-Y}\right] = \eta(s,-1) = M_{X,Y}(s,-1) = E\left[e^{sX}e^{-Y}\right]$$

and as (X,Y) is bivariate normally distributed, applying Lemma 2.4 it follows that

$$\eta(s,-1) = E\left[e^{-Y}\right]\exp\left(sE[X] + \frac{s^2}{2}Var[X] - sCov[X,Y]\right). \qquad (2.7)$$

Next, we consider

$$\exp\left(sE[X] + \frac{s^2}{2}Var[X] - sCov[X,Y]\right) \text{ of Eq. 2.7.}$$

For any random variable $X - Cov[X,Y]$, its mean and variance are

$$E[X - Cov[X,Y]] = E[X] - Cov[X,Y]$$

and

$$Var[X - Cov[X,Y]] = Var[X] = \sigma_X^2.$$

Since

$$M_{X-Cov[X,Y]}(s) = E\left[e^{s\{X-Cov[X,Y]\}}\right] = \exp\left(s(E[X] - Cov[X,Y]) + \frac{1}{2}s^2 Var[X]\right)$$

then Eq. 2.7 can be written as

$$\eta(s,-1) = E\left[e^{-Y}\right]E\left[e^{s\{X-Cov[X,Y]\}}\right]. \qquad (2.8)$$

Consider Eq. 2.6 and Eq. 2.8, one gets

$$E\left[e^{-Y}\right]E\left[e^{s\{X-Cov[X,Y]\}}\right] = \int_{-\infty}^{\infty} e^{sx}\,\xi_X(x)\,dx$$

$$E\left[e^{s\{X-Cov[X,Y]\}}\right] = \int_{-\infty}^{\infty} e^{sx}\,\frac{\xi_X(x)}{E[e^{-Y}]}\,dx.$$

Let $Cov[X,Y] = a$ and $x = u - a$.

Then we get that

$$E\left[e^{s(X-a)}\right] = \int_{-\infty}^{\infty} e^{s(u-a)}\,\frac{\xi_X(u-a)}{E[e^{-Y}]}\,du.$$

Thus, the density function of the random variable $(X-a)$ is

$$\frac{\xi_X(u-a)}{E\left[e^{-Y}\right]}.$$

We have seen that

$$E\left[e^{-Y}f(X)\right] = \int_{-\infty}^{\infty} f(x)\,\xi_X(x)\,dx\,.$$

Then we obtain that

$$E\left[e^{-Y}f(X)\right] = E\left[e^{-Y}\right]\int_{-\infty}^{\infty} f(x)\,\frac{\xi_X(x)}{E\left[e^{-Y}\right]}\,dx$$

$$= E\left[e^{-Y}\right]\int_{-\infty}^{\infty} f(u-a)\,\frac{\xi_X(u-a)}{E[e^{-Y}]}\,du$$

$$= E\left[e^{-Y}\right]E\left[f(X-a)\right].$$

We conclude that

$$E\left[e^{-Y}f(X)\right] = E\left[e^{-Y}\right]E\left[f(X-Cov[X,Y])\right]. \qquad \square$$

2.7 Equilibrium Price

2.7.1 A Model for the Market

The economic premiums are not only depending on the risk but also on market conditions. We can describe the risk by a random variable X and the market conditions by a random variable Z; such as an aggregate risk, collective wealth, correlation and etc.

In the market we are considering agents $j = 1, 2, \ldots, n$. They constitute buyers of insurance, insurance companies or reinsurance companies. Each agent j is characterized by his

(i) utility function $u_j(x)$ with first derivative and second derivative of $u_j(x)$ are $u_j'(x) > 0$ and $u_j''(x) < 0$, respectively, and

(ii) initial wealth w_j.

The risk aspect is modeled by a finite (for simplicity) probability space with states $s = 1,\ 2, \ldots,\ S$ and probabilities π_s of state s happening, i.e.,

$$\sum_{s=1}^{S} \pi_s = 1 .$$

The states s can be described as follows:

(a) Consider a whole insurance business; states are lines of insurance business such as the insurance of fire, motor, automobile, marine, health and etc. The amount of claims are produced from each line of business.

(b) Consider one line of business. For example, in automobile insurance; states may be the type of coverage such as type 1 (comprehensive cover), type 2 (third party fire and theft cover) and type 3 (third party cover).

(c) Consider one type of coverage. For example, in type 1 (comprehensive cover) of automobile insurance, states are loss of properties, accidental benefits and third party coverage.

Each agent j in the market has an original risk function $X_j(s)$; the payment caused to j if s is happening. He is buying an exchange function $Y_j(s)$; payment received by j if s is happening. The notion of price for this purchase is given by a vector

$$\mathbf{p} = (p_1,\ p_2, \ldots,\ p_S)'$$

and

$$\text{Price}[Y_j] = \sum_{s=1}^{S} p_s Y_j(s).$$

Hence p_s is the price for one unit of conditional money and $\sum_{s=1}^{S} p_s = 1$.

Definition 2.14. $\mathbf{Y} = (Y_1,\ Y_2, \ldots,\ Y_n)'$ is *a risk exchange (REX)* if $\sum_{j=1}^{n} Y_j \quad s \ = 0$ for all $s = 1, 2, \ldots, S$.

2.7.2 Equilibrium Price

Definition 2.15. The pair $(\mathbf{p}, \mathbf{Y})$ is called in *equilibrium* of the market if

(i) For all j, $\sum_{s=1}^{S} \pi_s u_j \left[w_j - X_j(s) + Y_j(s) - \sum p_s Y_j(s) \right]$ = max for all possible choices of exchange functions Y_j .

(ii) $\sum_{j=1}^{n} Y_j(s) = 0$ for all $s = 1, 2, \ldots, S$.

If condition (i) and (ii) are satisfied, $\mathbf{p}$ is called an *equilibrium price* and $\mathbf{Y}$ is called an *equilibrium risk exchange (REX).*

The notion of equilibrium price can be extended to an arbitrary probability space $(\Omega, \mathcal{F}, P)$ where the risk function $X_j(s)$ and exchange function $Y_j(s)$ will be represented by the random variables $X_j(\omega)$ and $Y_j(\omega)$, $\omega \in \Omega$, respectively. The notion of price is given by a function $\varphi : \Omega \to \mathbb{R}$ and the price $[Y_j]$ is defined by

$$\text{Price } [Y_j] = \int_{\Omega} Y_j(\omega)\, \varphi(\omega)\, dP(\omega)\,.$$

Definition 2.16. The pair (Y_j, φ) is called in *equilibrium* if

(i) For all j, $E[u_j(w_j - X_j + Y_j - \text{Price}(Y_j))]$ is a maximum among all possible choices of the exchange variables Y_j and

(ii) $\sum_{j=1}^{n} Y_j(\omega) = 0$ for all $\omega \in \Omega$.

In the equilibrium, Y_j is called *the equilibrium risk exchange* and φ is called *the equilibrium price density*.

2.7.3 Bühlmann's Equilibrium Pricing Model

Definition 2.17. (Bühlmann's equilibrium pricing model).

Each agent j has an exponential utility function

$$u_j(x) = \frac{1}{\lambda_j}\left[1 - \exp(-\lambda_j x)\right].$$

So that $u_j'(x) = \exp(-\lambda_j x)$, λ_j stands for the risk aversion and $\frac{1}{\lambda_j}$ stands for the risk tolerance unit. Then the equilibrium price density satisfies:

$$\varphi_e(\omega) = \frac{e^{(\lambda Z(\omega))}}{E[e^{\lambda Z}]},$$

where $Z(\omega) = \sum_{j=1}^{n} X_j(\omega)$ is the aggregate risk (the sum of original risk functions in the market) and λ satisfies

$$\frac{1}{\lambda} = \sum_{j=1}^{n} \frac{1}{\lambda_j}.$$

The parameters λ_j can be seen as *the risk aversion index* of the j^{th} agent.

Lemma 2.6. The equilibrium price for any risk X of Bühlmann's equilibrium pricing model is

$$H_B(X,Z) = \frac{E[Xe^{\lambda Z}]}{E[e^{\lambda Z}]},$$

where $Z(\omega) = \sum_{j=1}^{n} X_j(\omega)$ is the aggregate risk and λ satisfies

$$\frac{1}{\lambda} = \sum_{j=1}^{n} \frac{1}{\lambda_j}.$$

Proof:

The price of any risk X is

$$H_B(X,Z) := \text{Price } [X]$$

$$= \int_{\Omega} X(\omega)\, \varphi(\omega)\, dP(\omega)$$

$$= \int_{\Omega} X(\omega)\, \frac{e^{(\lambda Z(\omega))}}{E[e^{\lambda Z}]}\, dP(\omega)$$

$$= \frac{1}{E[e^{\lambda Z}]} \int_{\Omega} X(\omega)\, e^{(\lambda Z(\omega))}\, dP(\omega)$$

$$H_B(X,Z) = \frac{1}{E[e^{\lambda Z}]} E\left[Xe^{(\lambda Z)}\right].$$

We conclude that

$$H_B(X,Z) = \frac{E[Xe^{\lambda Z}]}{E[e^{\lambda Z}]}. \qquad \square$$

2.8 Wang Transform

Definition 2.18. Let Φ denote the standard normal cumulative distribution function, i.e., $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2} ds$, and let θ be a real valued parameter. By definition, the *Wang transform* transforms a CDF $F(x)$ to a function $F^*(x)$:

$$F^*(x) = \Phi[\Phi^{-1}(F(x)) + \theta], \tag{2.9}$$

It is obvious that $F^*(x)$ is also a CDF.

The key parameter θ in the Wang transform of Eq. 2.9 has a positive sign as the random variable X is kept in asset. On the other hand, in the insurance business, a liability of loss variable X is viewed as a negative asset. Thus, the Wang transform of our study has a negative sign in front of θ. That is

$$F^*(x) = \Phi[\Phi^{-1}(F(x)) - \theta], \tag{2.10}$$

where θ is a positive constant that is relevant to the market price of risk.

For a liability with loss variable X, the Wang transform in Eq. 2.9 has an equivalent representation.

$$S^*(x) = \Phi[\Phi^{-1}(S(x)) + \theta], \tag{2.11}$$

where $S(x) = 1 - F(x)$.

Lemma 2.7. For any θ, $S^*(x) = 1 - F^*(x)$. That is, transform Eq. 2.10 and Eq. 2.11 are equivalent.

Proof:

As $S(x) = 1 - F(x)$ and $S^*(x) = \Phi\left[\Phi^{-1}(S(x)) + \theta\right]$.

That is,

$$S^*(x) = \Phi\left[\Phi^{-1}(S(x)) + \theta\right]$$

$$= \Phi\left[\Phi^{-1}(1 - F(x)) + \theta\right]$$

$$= \Phi[-\Phi^{-1}(F(x)) + \theta)]$$

$$= \Phi[-(\Phi^{-1}(F(x)) - \theta)]$$

$$= 1 - \Phi[(\Phi^{-1}(F(x)) - \theta)]$$

$$= 1 - F^*(x).$$

Thus, the lemma has been proved. □

Note that above we have used:

(1) $1 - \Phi(x) = \Phi(-x)$

(2) $\Phi^{-1}(1 - u) = -\Phi^{-1}(u)$

Lemma 2.8. Let F be the Lognormal cumulative distribution function of a loss X with μ and σ, i.e., $X \sim LN(\mu, \sigma)$. Then the Wang transform F^* is a Lognormal CDF with $\mu + \theta\sigma$ and σ corresponding to some loss X' i.e., $X' \sim LN(\mu + \theta\sigma, \sigma)$.

Proof:

As $X \sim LN(\mu,\sigma)$ then $\dfrac{\ln X - \mu}{\sigma} \sim N(0,1)$.

By the Wang transform, for any constant θ, one has:

$$F^*(x) = \Phi\left[\Phi^{-1}(F(x)) - \theta\right]$$

$$= \Phi\left[\Phi^{-1}\left[\Phi\left(\frac{\ln x - \mu}{\sigma}\right)\right] - \theta\right]$$

$$= \Phi\left(\frac{\ln x - \mu}{\sigma} - \theta\right)$$

$$= \Phi\left(\frac{\ln x - \mu - \theta\sigma}{\sigma}\right)$$

$$= \Phi\left(\frac{\ln x - (\mu + \theta\sigma)}{\sigma}\right).$$

The proof is completed, one obtains that

$$\ln X \sim N(\mu + \theta\sigma, \sigma),$$

that is

$$X \sim LN(\mu + \theta\sigma, \sigma). \qquad \square$$

CHAPTER III

CLAIM MODELING

In this chapter, the finite mixture of Lognormal distributions is presented for the modeling of insurance claims. The EM algorithm is used to perform a parametric fit of given data to a mixture of Lognormal distributions. We have performed numerical experiments to fit data simulated by mixtures of various loss distributions to finite mixture Lognormal distributions, and also mathed an actual set of insurance claim data to a finite mixture of Lognormal distributioins.

We consider individual claim policies, and the claim amount X_i is paid for the i^{th} policy. Some assumptions and restrictions are specified as below.

Assumption 1: (Policy independence): Consider n different policies. Let X_i denote the response for policy i. Then $X_1,...,X_n$ are independent.

Assumption 2: Severity losses are non-catastrophic losses.

Assumption 3: There are no deductibles and no reinsurance agreement.

Assumption 4: A recorded claim is equal to an actual claim (observation).

Assumption 5: The loss distributions are skewed to the right.

The right skewness of loss distributions are considered for this study. We assume that the portfolio claim amount is arising from different loss distributions, e.g., the empirical data are generated by mixing of Lognormal, Gamma, Pareto and Weibull distributions. We have performed numerical experiments by simulation, see section A. 3 of Appendix A for details. The probability density function (PDF) and cumulative distribution function (CDF) of loss distributions are specified in Appendix A.

3.1 Single Parametric Distribution

On the basis of the analyst's knowledge, experience and statistical tests, the Lognormal distribution is our selection for modeling and fitting to the data set. The maximum likelihood estimate (MLE) is used for parameter estimation, as explained below.

3.1.1 The Model

Assume that $X \sim \text{Lognormal}(\mu, \sigma)$, abbreviated $X \sim LN(\mu, \sigma)$, with density

$$f_X(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right) \ ; \quad \mu \in R, \quad \sigma > 0, \quad x > 0. \qquad (3.1)$$

3.1.2 Estimation for the Model

Let a vector $\boldsymbol{x} = (x_1, \ldots, \ x_n)'$ be an independent observation. Consider the amount x_i paid for the i^{th} contract. We fit the Lognormal distribution in Eq. 3.1 to the data set by MLE. The likelihood function is $L = \prod_{i=1}^{n} f_X(x_i)$; $i = 1, 2, \ldots, n$.

Then $\ln L = \ln \prod_{i=1}^{n} f_X(x_i)$

$$= \sum_{\mathrm{i}=1}^{n} \ln f_X(x_i)$$

$$\ln L = \sum_{\mathrm{i}=1}^{n} \ln\left[\frac{1}{x_i \sigma \sqrt{2\pi}} \exp\left(-\frac{(\ln x_i - \mu)^2}{2\sigma^2}\right)\right]$$

$$= \sum_{\mathrm{i}=1}^{n}\left[-(\ln\sigma + \ln x_i) - \frac{1}{2}\ln 2\pi - \frac{1}{2\sigma^2}(\ln x_i - \mu)^2\right] .$$

We estimate $\hat{\mu}$ and $\hat{\sigma}$ for μ and σ respectively by $\frac{\partial}{\partial\mu}\ln L = 0$ and $\frac{\partial}{\partial\sigma}\ln L = 0$.

We obtain maximum likelihood estimates for the parameter μ and the parameter σ as follows:

$$\hat{\mu} = \frac{\sum_{i=1}^{n}\ln x_i}{n} \quad \text{and} \quad \hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n}(\ln x_i - \hat{\mu})^2}{n}}, \text{ respectively.} \tag{3.2}$$

3.2 Finite Mixture Models

Next, second-order and higher-order finite mixture models are considered. In this section, we aim to find the mixing weights according to the number of Lognormal distributions and estimated parameters by the MLE via EM algorithm.

3.2.1 The Model

The PDF of finite mixture Lognormal distributions is

$$f(x) = \tau_1 f_1(x) + \cdots + \tau_k f_k(x)$$

$$= \frac{1}{x\sqrt{2\pi}}\left(\tau_1 \frac{1}{\sigma_1}\exp\left(-\frac{(\ln x - \mu_1)^2}{2\sigma_1^2}\right) + \cdots + \tau_k \frac{1}{\sigma_k}\exp\left(-\frac{(\ln x - \mu_k)^2}{2\sigma_k^2}\right)\right), \tag{3.3}$$

$\mu_j \in \mathbb{R}$, $\sigma_j > 0$, $x > 0$, where $0 < \tau_j < 1$ for $j = 1, \ldots, k$ and $\tau_1 + \cdots + \tau_k = 1$.

The likelihood function can be written as follows:

$$L = \prod_{i=1}^{n} \frac{1}{x_i\sqrt{2\pi}} \left(\tau_1 \frac{1}{\sigma_1} \exp\left(-\frac{(\ln x_i - \mu_1)^{\;2}}{2\sigma_1^2} \right) + \cdots + \tau_k \frac{1}{\sigma_k} \exp\left(-\frac{(\ln x_i - \mu_k)^2}{2\sigma_k^2} \right) \right)$$

and the log-likelihood function is in the form

$$\ln L = \sum_{i=1}^{n} \ln \left[\sum_{j=1}^{k} \tau_j \frac{1}{x_i\sqrt{2\pi}\sigma_j} \exp\left(-\frac{(\ln x_i - \mu_j)^2}{2\sigma_j^2} \right) \right].$$

3.2.2 Estimation for the Model

Here, we construct the complete data set which is composed of observed data (incomplete data) and unobservable (latent) data. The EM algorithm is a powerful algorithm for parameter estimation of data arising from mixtures. The details of MLE via EM algorithm are as follows.

Let a sample $\boldsymbol{x} = (x_1, x_2, \; ..., \; x_n)'$ be observed data to be matched to the mixture of Eq. 3.3 and having a postulated PDF as

$$f(\boldsymbol{x}, \psi),$$

where ψ is a vector of unknown parameters; $\psi = (\theta, \boldsymbol{\tau})$, $\boldsymbol{\tau} = (\tau_1, ..., \; \tau_{k-1})'$ and $\theta = (\mu_1, ..., \; \mu_k, \; \sigma_1, ..., \; \sigma_k)'$.

Let $\boldsymbol{z}$ be the unobservable data matrix; denoted by

$$\boldsymbol{z} = (z_{ij}, \; i = 1, ..., n; \; j = 1, ..., k)$$

The values z_{ij} are indicators defined as

$$z_{ij} = \begin{cases} 1\;, & \text{observation } x_i \text{ comes from the distribution } f_j \\ 0\;, & \text{elsewhere} \end{cases}$$

The unobservable matrix $\boldsymbol{z}$ tell us, where the i^{th} observation x_i comes from.

Let $\boldsymbol{Z}$ be a random matrix whose realization is the unobservable matrix $\boldsymbol{z}$.

Let $k(\boldsymbol{z} \mid \psi, \boldsymbol{x})$ denote the conditional PDF of the unobserved data and define the PDF as

$$k(\boldsymbol{z} \mid \psi, \boldsymbol{x}) = t_{ij},$$

where

$$t_{ij} = \frac{\tau_j f_j(x_i \mid \mu_j, \sigma_j)}{\sum_{j=1}^{k} \tau_j f_j(x_i \mid \mu_j, \sigma_j)} = \frac{\tau_j f_j(x_i \mid \mu_j, \sigma_j)}{f(x_i)}.$$

Note that t_{ij} is the probability of the i^{th} observation coming from the j^{th} component.

We obtain that

$$E(Z_{ij} \mid \boldsymbol{x}) = P(Z_{ij} = 1 \mid \boldsymbol{x}) = t_{ij}.$$

Assume that $\mathbf{X}$ and $\boldsymbol{Z}$ are independent. Then the complete likelihood takes form;

$$L_c(\psi \mid \boldsymbol{x}, \boldsymbol{z}) = \prod_{i=1}^{n} \prod_{j=1}^{k} \left[\tau_j \frac{1}{x_i \sqrt{2\pi} \sigma_j} \exp\left(-\frac{(\ln x_i - \mu_j)^2}{2\sigma_j^2} \right) \right]^{z_{ij}}.$$

The complete log-likelihood function is

$$\ln L_c(\psi \mid \boldsymbol{x}, \boldsymbol{z}) = \ln \left[\prod_{i=1}^{n} \prod_{j=1}^{k} \left[\tau_j \frac{1}{x_i \sqrt{2\pi} \sigma_j} \exp\left(-\frac{(\ln x_i - \mu_j)^2}{2\sigma_j^2} \right) \right]^{z_{ij}} \right]$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{k} \ln \left[\tau_j \frac{1}{x_i \sqrt{2\pi} \sigma_j} \exp\left(-\frac{(\ln x_i - \mu_j)^2}{2\sigma_j^2} \right) \right]^{z_{ij}}$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{k} z_{ij} \ln \left[\tau_j \frac{1}{x_i \sqrt{2\pi} \sigma_j} \exp\left(-\frac{(\ln x_i - \mu_j)^2}{2\sigma_j^2} \right) \right]$$

We obtain that

$$\ln L_c(\psi|\boldsymbol{x},\boldsymbol{z}) = \sum_{i=1}^{n}\sum_{j=1}^{k} z_{ij}\left[\ln\tau_j - \ln x_i - \ln\sigma_j - \frac{1}{2}\ln(2\pi) - \frac{1}{2\sigma_j^2}(\ln x_i - \mu_j)^2\right]. \quad (3.4)$$

Note that: $\psi = (\theta, \boldsymbol{\tau})$, $\boldsymbol{\tau} = (\tau_1,\ldots,\ \tau_{k-1})'$ and $\theta = (\mu_1,\ldots,\ \mu_k,\ \sigma_1,\ldots,\ \sigma_k)'$.

For each k components, there are $3k-1$ unknown parameters that will be estimated by the EM algorithm. We use a computer for the calculation of the parameters and visualization as a way to see its modeling. The proper number of components to be included in the mixture model will be considered.

Expectation Step (E-step):

Replacing z_{ij} in Eq. 3.4 by its expected value, $\hat{t}_{ij}$, yields the expected complete log-likelihood,

$$E[\ln L_c(\psi|\boldsymbol{x},\boldsymbol{z})] = \sum_{i=1}^{n}\sum_{j=1}^{k} \hat{t}_{ij}\left[\ln\tau_j - \ln x_i - \ln\sigma_j - \frac{1}{2}\ln(2\pi) - \frac{1}{2\sigma_j^2}(\ln x_i - \mu_j)^2\right], \quad (3.5)$$

where $\hat{t}_{ij}$ is the estimated value of t_{ij}.

Note that: t_{ij} is given by

$$t_{ij} = P(Z_{ij} = 1|X_i = x_i, \psi) = \frac{\tau_j f_j(x_i \mid \mu_j, \sigma_j)}{\sum_{j=1}^{k}\tau_j f_j(x_i \mid \mu_j, \sigma_j)} = \frac{\tau_j f_j(x_i \mid \mu_j, \sigma_j)}{f(x_i)}.$$

Maximization Step (M-step):

We maximize Eq. 3.5 to estimate ψ. Firstly, we solve the first order condition:

$$\frac{\partial}{\partial\tau_j} E[\ln L_c(\psi|\boldsymbol{x},\boldsymbol{z})] = 0,$$

with constraint

$$\tau_1 + \cdots + \tau_k = 1.$$

$$\frac{\partial}{\partial \tau_j} \sum_{i=1}^{n} \sum_{j=1}^{k} \hat{t}_{ij} \left[\ln \tau_j - \ln x_i - \ln \sigma_j - \frac{1}{2} \ln(2\pi) - \frac{1}{2\sigma_j^2} (\ln x_i - \mu_j)^2 \right] = 0 .$$

Without loss of generality (w.l.g.), we consider

$$\frac{\partial}{\partial \tau_j} \sum_{i=1}^{n} \sum_{j=1}^{k} \hat{t}_{ij} \left[\ln \tau_j \right] = 0$$

$$\frac{\partial}{\partial \tau_j} \left[\sum_{j=1}^{k} \left(\sum_{i=1}^{n} \hat{t}_{ij} \right) \ln \tau_j \right] = 0 \ .$$

This has the same form as the MLE for the multinomial distribution, for details see multinomial distribution and MLE in Appendix B. We get that

$$\hat{\tau}_j = \frac{\sum_{i=1}^{n} \hat{t}_{ij}}{\sum_{j=1}^{k} \left(\sum_{i=1}^{n} \hat{t}_{ij} \right)} = \frac{\sum_{i=1}^{n} \hat{t}_{ij}}{\sum_{i=1}^{n} \left(\sum_{j=1}^{k} \hat{t}_{ij} \right)} = \frac{1}{n} \sum_{i=1}^{n} \hat{t}_{ij} \quad . \tag{3.6}$$

Secondly, we solve the equation $\frac{\partial}{\partial \theta_j} E[\ln L_c(\psi | \boldsymbol{x}, \boldsymbol{z})] = 0$ for estimated parameters of $\theta_j = (\mu_j, \sigma_j)$, $j = 1, 2, \ldots, k$.

Consider $\theta_1 = (\mu_1, \sigma_1)$.

We will estimate θ_1 by solving;

$$\frac{\partial}{\partial \mu_1} E[\ln L_c(\psi | \boldsymbol{x}, \boldsymbol{z})] = 0 \text{ and } \frac{\partial}{\partial \sigma_1} E[\ln L_c(\psi | \boldsymbol{x}, \boldsymbol{z})] = 0 .$$

Note that the relation $\frac{\partial}{\partial \mu_1} E[\ln L_c(\psi | \boldsymbol{x}, \boldsymbol{z})] = 0$ and equation (3.6) imply

$$\sum_{i=1}^{n}\sum_{j=1}^{k}\hat{t}_{ij}\frac{\partial}{\partial\mu_1}\left[\ln\tau_j-\ln x_i-\ln\sigma_j-\frac{1}{2}\ln(2\pi)-\frac{1}{2\sigma_j^2}(\ln x_i-\mu_j)^2\right] = 0$$

$$\sum_{i=1}^{n}\hat{t}_{i1}\frac{\partial}{\partial\mu_1}\left[\ln\tau_1-\ln x_i-\ln\sigma_1-\frac{1}{2}\ln(2\pi)-\frac{1}{2\sigma_1^2}(\ln x_i-\mu_1)^2\right] = 0$$

$$\sum_{i=1}^{n}\hat{t}_{i1}(\ln x_i-\mu_1) = 0$$

$$\sum_{i=1}^{n}\hat{t}_{i1}\ln x_i-\sum_{i=1}^{n}\hat{t}_{i1}\mu_1 = 0$$

$$\hat{\mu}_1 = \frac{\sum_{i=1}^{n}\hat{t}_{i1}\ln x_i}{\sum_{i=1}^{n}\hat{t}_{i1}}.$$

$$\frac{\partial}{\partial\sigma_1}E[\ln L_c(\psi|\boldsymbol{x},\boldsymbol{z})] = 0$$

$$\sum_{i=1}^{n}\sum_{j=1}^{k}\hat{t}_{ij}\frac{\partial}{\partial\sigma_1}\left[\ln\tau_j-\ln x_i-\ln\sigma_j-\frac{1}{2}\ln(2\pi)-\frac{1}{2\sigma_j^2}(\ln x_i-\hat{\mu}_j)^2\right] = 0$$

$$\sum_{i=1}^{n}\hat{t}_{i1}\frac{\partial}{\partial\sigma_1}\left[\ln\tau_1-\ln x_i-\ln\sigma_1-\frac{1}{2}\ln(2\pi)-\frac{1}{2\sigma_1^2}(\ln x_i-\hat{\mu}_1)^2\right] = 0$$

$$\sum_{i=1}^{n}\hat{t}_{i1}\left[-\frac{1}{\sigma_1}+\frac{1}{\sigma_1^3}(\ln x_i-\hat{\mu}_1)^2\right] = 0$$

$$\sum_{i=1}^{n}\hat{t}_{i1}\left[-1+\frac{1}{\sigma_1^2}(\ln x_i-\hat{\mu}_1)^2\right] = 0$$

$$\frac{1}{\sigma_1^2}\sum_{i=1}^{n}\hat{t}_{i1}(\ln x_i-\hat{\mu}_1)^2 = \sum_{i=1}^{n}\hat{t}_{i1}$$

$$\hat{\sigma}_1 = \sqrt{\frac{\sum_{i=1}^{n} \hat{t}_{i1}(\ln x_i - \hat{\mu}_1)^2}{\sum_{i=1}^{n} \hat{t}_{i1}}} \quad .$$

Similarly, one can show that

$$\hat{\mu}_{\mathrm{j}} = \frac{\sum_{i=1}^{n} \hat{t}_{ij} \ln x_i}{\sum_{i=1}^{n} \hat{t}_{ij}} \quad \text{and} \quad \hat{\sigma}_j = \sqrt{\frac{\sum_{i=1}^{n} \hat{t}_{ij}(\ln x_i - \hat{\mu}_j)^2}{\sum_{i=1}^{n} \hat{t}_{ij}}} \quad , j = 1,\ 2,\ldots,\ k.$$

In summary, we obtain that

$$\hat{\tau}_j = \frac{1}{n}\sum_{i=1}^{n} \hat{t}_{ij}\ ,\ \hat{\mu}_{\mathrm{j}} = \frac{\sum_{i=1}^{n} \hat{t}_{ij} \ln x_i}{\sum_{i=1}^{n} \hat{t}_{ij}} \quad \text{and} \quad \hat{\sigma}_j = \sqrt{\frac{\sum_{i=1}^{n} \hat{t}_{ij}(\ln x_i - \hat{\mu}_j)^2}{\sum_{i=1}^{n} \hat{t}_{ij}}} \quad ;$$

$j = 1,\ 2,\ldots,\ k.$

Note that the expected complete log-likelihood function is given by

$$E[\ln L_c(\psi | \boldsymbol{x}, \boldsymbol{z})] = \sum_{i=1}^{n}\sum_{j=1}^{k} \hat{t}_{ij}\left[\ln \tau_j - \ln x_i - \ln \sigma_j - \frac{1}{2}\ln(2\pi) - \frac{1}{2\sigma_j^2}(\ln x_i - \mu_j)^2\right].$$

For a given set of parameters ψ, i.e. $\hat{\theta}_j = (\hat{\mu}_j,\ \hat{\sigma}_j)$, $j = 1,\ 2,\ldots,\ k$ and $\hat{\boldsymbol{\tau}} = (\hat{\tau}_1,\ldots,\ \hat{\tau}_{k-1})'$, the E-step consists of calculating $\hat{t}_{ij}$ and $\hat{\tau}_j$ for M-step. Given $\hat{\tau}_j$, the M-step consists of maximizing the expected complete log-likelihood function. The E-step and M-step are repeated in an alternating fashion until the expected complete log-likelihood fails to increase. At this point, we conduct a final M-step in which the set of parameters ψ is estimated. Otherwise, we return to the E-step for the next iteration. In the final step after the m^{th} iteration, the EM algorithm is produced as below:

E-step: Given our current estimation of the parameters $\psi^{(m)}$ after the m^{th} iteration. Thus the E-step results in the function:

$$Q(\psi \mid \psi^{(m)}) = \sum_{i=1}^{n}\sum_{j=1}^{k}\hat{t}_{ij}^{(m)}\left[\ln\hat{\tau}_j^{(m)} - \ln x_i - \ln\hat{\sigma}_j^{(m)} - \frac{1}{2}\ln(2\pi) - \frac{1}{2\hat{\sigma}_j^{(m)^2}}(\ln x_i - \hat{\mu}_j^{(m)})^2\right]. \quad (3.7)$$

M-step: Maximizing ψ . That is

$$\hat{\boldsymbol{\tau}}^{(m+1)} = \arg\max_{\boldsymbol{\tau}} Q(\psi \mid \psi^{(m)}) \text{ and } \hat{\theta}^{(m+1)} = \arg\max_{\theta} Q(\psi \mid \psi^{(m)}).$$

By taking partial derivative Eq. 3.7 with respect to ψ and by equating to zero, one gets

$$\hat{\tau}_j^{(m+1)} = \frac{1}{n}\sum_{i=1}^{n}\hat{t}_{ij}^{(m)}\ , \quad \hat{\mu}_j^{(m+1)} = \frac{\sum_{i=1}^{n}\hat{t}_{ij}^{(m)}\ln x_i}{\sum_{i=1}^{n}\hat{t}_{ij}^{(m)}}$$

and $$\hat{\sigma}_j^{\,m+1} = \sqrt{\frac{\sum_{i=1}^{n}\hat{t}_{ij}^{(m)}(\ln x_i - \hat{\mu}_j^{(m)})^2}{\sum_{i=1}^{n}\hat{t}_{ij}^{(m)}}}\ .$$

Note that $\left|\frac{Q(\psi \mid \psi^{(m+1)}) - Q(\psi \mid \psi^{(m)})}{Q(\psi \mid \psi^{(m)})}\right| \le 10^{-3}$ is applied for our programming.

3.3 Bootstrap Technique

We are interested in the bootstrap sample for observation and residual. We shall recalculate the estimated parameters of the Lognormal distribution by using the bootstrap technique and MLE. One advantage of the bootstrap technique is that we can calculate as many replications of the sample as we want.

3.3.1 Observation Bootstrap

Define

$$\boldsymbol{x}^* = (x_1^*,\ x_2^*, \ldots,\ x_n^*)'. \tag{3.8}$$

The bootstrap data points $x_1^*,\ x_2^*, \ldots,\ x_n^*$ are a random sample of size n with replacement from the observation of n objects $(x_1,\ x_2, \ldots,\ x_n)'$. Then we recalculate the estimated parameters, $\hat{\mu}^*$ and $\hat{\sigma}^*$, by MLE based on $\boldsymbol{x}^*$.

3.3.2 Residual Bootstrap

There are many forms of the residual definition and it is important to use an appropriate residual definition for the determination of each problem. We have already run trials with some forms of residual definitions, such as the unscaled Pearson residual and the unscaled Anscombe residual, but these forms of residual proned not suitable for our data. Instead, we consider the residual form $\hat{\mu}$, that is, we define the form of the residual as follows.

$$\varepsilon_i = \ln x_i - \hat{\mu},$$

where ε_i is the residual ($i = 1, 2, \ldots, n$) and $\hat{\mu}$ comes from Eq. 3.2.

Let $\boldsymbol{\varepsilon} = (\varepsilon_1,\ \varepsilon_2, \ldots,\ \varepsilon_n)'$ and $\boldsymbol{\varepsilon}^* = (\varepsilon_1^*,\ \varepsilon_2^*, \ldots,\ \varepsilon_n^*)'$ be the resample residual. By using the bootstrap technique, we obtain a resample $\boldsymbol{\varepsilon}^*$ and the bootstrap data samples

$$\ln x_i^* = \varepsilon_i^* + \hat{\mu}\ ;\ i = 1, 2, \ldots, n. \tag{3.9}$$

We recalculate the estimated parameters, $\hat{\mu}^*$ and $\hat{\sigma}^*$ by MLE based on $\ln x_i^*$, $i = 1,\ 2, \ldots,\ n$.

3.4 Goodness of Fit Test

The goodness of fit (GOF) test measures the compatibility of a random sample with a theoretical probability distribution function. We use the Kolmogorov-Smirnov test (*K-S* test) and the Anderson-Darling test (*A-D* test) for showing how well the distribution fits our data set.

The *K-S* test is used to decide if a sample comes from a hypothesized continuous distribution. It is based on the empirical cumulative distribution function (ECDF) and denoted by

$$F_n(x) = \frac{1}{n}\left[\text{Number of observations} \le x\right].$$

The *K-S* test statistic is defined by

$$D = \sup_x \left| F_n(x) - F_X^*(x) \right|.$$

The *A-D* test is a general test to compare the fit of an observed cumulative distribution function to an expected cumulative distribution function. This test gives more weight to the tails than the *K-S* test.

The *A-D* test statistic is defined as

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\left[\ln F_X^*(x_i) + \ln \;\; 1 - F_X^*(x_{n-i+1}) \;\; \right],$$

where F_X^* is the theoretical cumulative distribution of the distribution being tested.

The test, for both *K-S* and *A-D*, is defined by:

H_0: The data follow the specified distribution.

H_1: The data do not follow the specified distribution.

Level critical values: The hypothesis regarding the distributional form is rejected at the chosen significance level (alpha, α) if the test statistic, *D* and A^2, is greater than the critical value obtained from Appendix A, Table A.4 and Table A.5 for *D* and A^2, respectively. On the other hand, we can calculate the *P*-value and interpret the result of hypothesis test. The interpretation of the *P*-value is given in Table A.6 of Appendix A.

3.5 The Simulation

We assume that the insurance portfolio is heterogeneous, due to variability in the parameters and distributions, and thus cannot be fitted to any single parametric distribution. For this reason, we have performed numerical experiments matching simulated data to finite mixtures of Lognormal distribution. The simulated heterogeneous data was generated by applying various combinations of loss distributions. The programming for this study is in MATLAB.

The data is generated by simulations that are under the following assumptions.

1) Sample size

n : 100, 300, 500, 800 and 1,000 for 2 and 4 mixed components.

n : 150, 300, 600, 900 and 1,200 for 3 mixed components.

2) The empirical data

2.1) The loss distributions: Lognormal, Gamma, Pareto and Weibull.

2.2) The empirical data: The x_i is simulated by loss distributions, due to variability in the parameters and distributions as detailed in Table 3.1.

Table 3.1 The variability of mixed components.

Components	Variability	
	Parameters	Distributions
2	Lognormal Gamma Pareto Weibull	Lognormal/Gamma Lognormal/Pareto Lognormal/Weibull Gamma/Pareto Gamma/Weibull Pareto /Weibull
3	Lognormal Gamma Pareto Weibull	Lognormal/Gamma/Weibull Gamma/Weibull/Pareto Weibull/Pareto/Lognormal
4	-	Lognormal/Gamma/Weibull/Pareto

The proportion of mixing is the same for each component mixed. The empirical data are simulated according to assumed parameters for each component mixed, see the imposed parameters for details in Table A.1, Table A.2 and Table A.3 of appendix A. The simulations span 90 cases.

2.3) The compound Poisson-mixed loss distributions: the frequency distribution is Poisson and the severity distributions are loss distributions. For $i = 1, 2, ..., n$, the claim X_i occurs at time t_i and is to be discounted at time zero with the risk free of interest rate j per annum. The claim amount at time zero is defined by

$$X_i^* = X_i (1 + j)^{-t_i}.$$

The j are assumed as 0.5%, 1% , 2%, 3%, 4% and 5% per annum.

3) The model of finite mixture distributions

The models for fitting to the empirical data is the finite mixture of Lognormal distributions. The k components depend on the sample size n. The total number of calculated components is 752 for 2, 3 and 4 components are 410, 301 and 41 cases, respectively. The single parametric distribution of Lognormal is used as a control to compare how well the finite mixture Lognormal distributions perform.

4) The bootstrap

The bootstrap process is a tool for fitting and it is not complicated to implement. We apply the bootstrap technique to reproduce pseudo data; reproduce from empirical data, then recalculate the estimated parameters by MLE and compare to the finite mixture Lognormal distributions.

The simulations run 200 iterations for the best solution that provide the estimated parameters for model fitting. That is, the average of estimated parameters are rather stable as the number of iterations is 200 times.

$$\left|\frac{\sum_{t=1}^{200}\hat{\mu}_t}{200}-\frac{\sum_{t=1}^{199}\hat{\mu}_t}{199}\right|\leq 10^{-4} \text{ and } \left|\frac{\sum_{t=1}^{200}\hat{\sigma}_t}{200}-\frac{\sum_{t=1}^{199}\hat{\sigma}_t}{199}\right|\leq 10^{-4}.$$

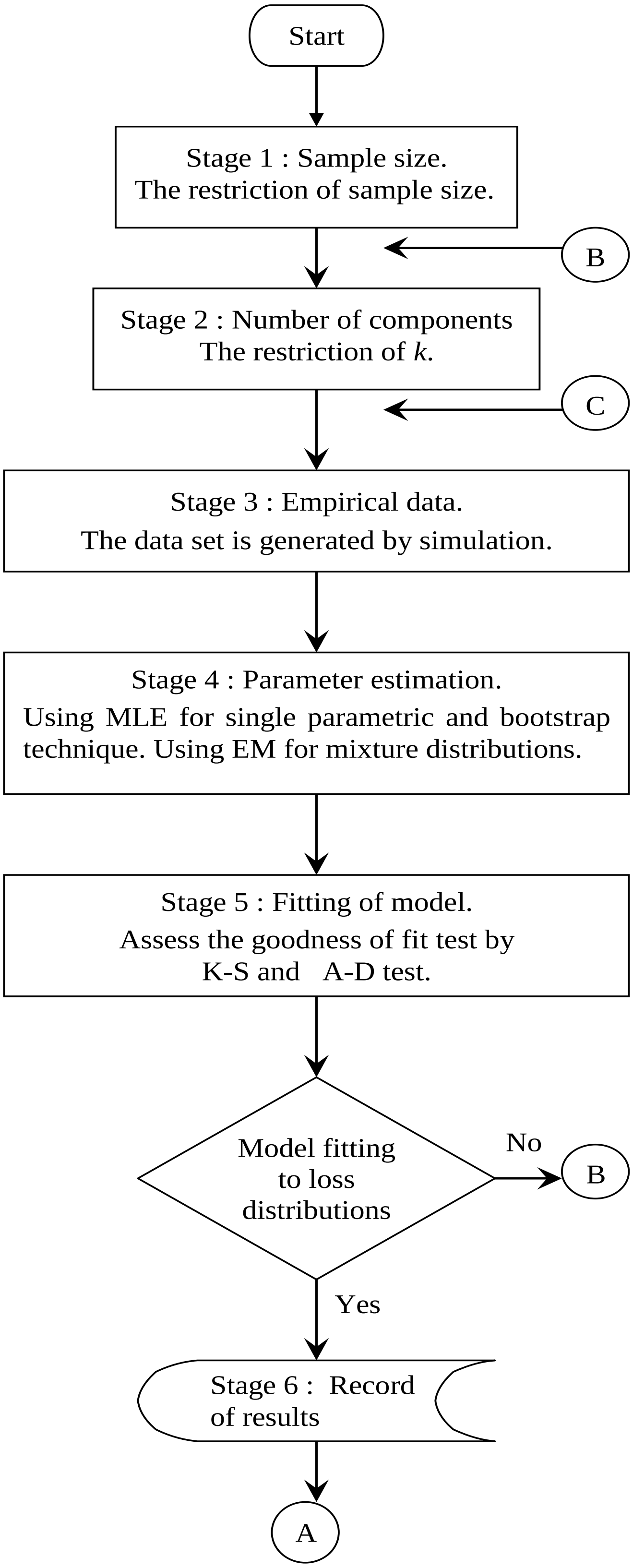

Figure 3.1 Flowchart of the claim modeling process.

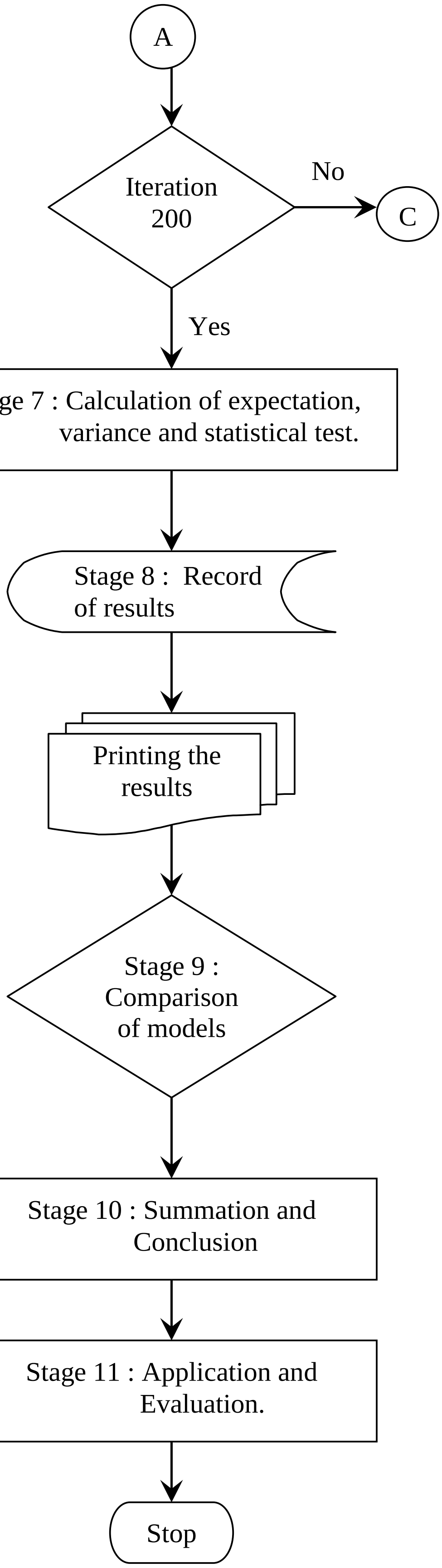

Figure 3.1 Flowchart of the claim modeling process (Continued).

3.6 Simulation Results

The purpose of claim modeling is to investigate the *k* components and summarize what kind of mixed loss data can be fitted by the finite mixture of Lognormal distributions. The empirical data is simulated by mixed components of loss distributions; Lognormal, Gamma, Pareto and Weibull distributions. The methodologies for parameter estimation are the MLE for single parametric Lognormal distribution and the EM for finite mixture Lognormal distributions. The statistical test for model fitting are *K-S* and *A-D* test. Some symbols are defined for easier explanation.

EMD means the empirical data which are simulated by mixed components of loss distributions.

EDP means the empirical data of discounted compound Poisson-mixed loss distributions with interest rate *j* per annum.

SPLD means the fitting of single parametric Lognormal distribution to EMD.

SPLD with Boot means the fitting of single parametric Lognormal distribution to the EMD with the bootstrap technique.

DCP means the fitting of single parametric Lognormal distribution to the EDP.

P-AS means *P*-value based on *A-D* test

P-KS means *P*-value based on *K-S* test

We analyze and present the value of A^2, D, *P-AS*, *P-KS*, $\hat{\mu}$ and $\hat{\sigma}$ on tables. The results are shown as the following tables.

Tables 3.2 - 3.20 show the values of A^2, D, *P-AS*, *P-KS*, $\hat{\mu}$ and $\hat{\sigma}$ of SPLD, SPLD with Boot and DCP for each sample size.

The results: For SPLD, the single parametric Lognormal distribution cannot be fitted to any EMD by *A-D* and *K-S* test. The SPLD with Boot, the single parametric Lognormal distribution is fitted to some sample sizes of EMD respective to *K-S* test only. The DCP, the single parametric Lognormal distribution cannot be fitted to any EDP by *A-D* and *K-S* test. For each sample size, the value of A^2 and D are mostly reduced when interest rate j increases.

Tables 3.21 - 3.39 show the values of A^2, D, *P-AS* and *P-KS* of finite mixture Lognormal distributions for fitting in each sample size. The results show that the finite mixture Lognormal distributions can be fitted to EMD at a significant level of $\alpha = 0.10$, for both *K-S* and *A-D* test. The mixture Lognormal distributions are a better fit to the EMD while k is increased.

Table 3.2 Lognormal distribution fitting to 2 mixed components of Lognormal distributed samples.

n	Item	SPLD	SPLD with Boot	DCP 0.50%	DCP 1%	DCP 2%	DCP 3%	DCP 4%	DCP 5%
100	A^2	3.17131	3.79915	3.17121	3.17111	3.17087	3.17059	3.17028	3.16995
	D	0.15850	0.11702	0.15851	0.15852	0.15854	0.15857	0.15857	0.15856
	P-AS	0.03009	0.01170	0.03010	0.03010	0.03011	0.03011	0.03011	0.03011
	P-KS	0.01667	0.13110	0.01665	0.01664	0.01660	0.01656	0.01656	0.01658
300	A^2	8.73616	12.25058	8.73610	8.73601	8.73578	8.73548	8.73511	8.73469
	D	0.14730	0.11401	0.14730	0.14730	0.14730	0.14728	0.14727	0.14724
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
500	A^2	14.57275	12.89082	14.57307	14.57336	14.57385	14.57422	14.57446	14.57459
	D	0.14465	0.11374	0.14465	0.14465	0.14465	0.14470	0.14472	0.14474
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
800	A^2	22.90827	20.24887	22.90826	22.90819	22.90788	22.90733	22.90656	22.90557
	D	0.14177	0.11357	0.14177	0.14176	0.14173	0.14175	0.14174	0.14173
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
1,000	A^2	28.55152	28.11877	28.55142	28.55125	28.55071	28.54990	28.54882	28.54747
	D	0.14111	0.13422	0.14109	0.14108	0.14109	0.14108	0.14110	0.14110
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01

Table 3.3 Lognormal distribution fitting to 2 mixed components of Gamma distributed samples.

n	Item	SPLD	SPLD with Boot	DCP 0.50%	DCP 1%	DCP 2%	DCP 3%	DCP 4%	DCP 5%
100	A^2	17.65236	17.70023	17.64868	17.63912	17.60802	17.56771	17.52201	17.47268
	D	0.33564	0.33482	0.33561	0.33556	0.33539	0.33516	0.33488	0.33458
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
300	A^2	52.94636	52.91433	52.93683	52.90952	52.81970	52.70469	52.57605	52.43828
	D	0.33670	0.33644	0.33664	0.33647	0.33606	0.33561	0.33515	0.33470
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
500	A^2	88.25355	88.24749	88.23911	88.19881	88.06382	87.88685	87.68613	87.46952
	D	0.33791	0.33643	0.33786	0.33772	0.33731	0.33685	0.33637	0.33588
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
800	A^2	141.18684	141.11608	141.16129	141.08858	140.84636	140.53080	140.17342	139.78778
	D	0.33860	0.33755	0.33853	0.33835	0.33790	0.33742	0.33692	0.33641
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
1,000	A^2	176.49137	176.40968	176.45893	176.36581	176.05765	175.65994	175.21189	174.72965
	D	0.33881	0.33705	0.33872	0.33855	0.33810	0.33762	0.33713	0.33662
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01

Table 3.4 Lognormal distribution fitting to 2 mixed components of Pareto distributed samples.

n	Item	SPLD	SPLD with Boot	DCP					
				0.50%	1%	2%	3%	4%	5%
100	A^2	1.35102	0.93414	1.35050	1.34998	1.34893	1.34787	1.34677	1.34567
	D	0.09820	0.08134	0.09817	0.09815	0.09807	0.09798	0.09788	0.09778
	P-AS	> 0.10	> 0.10	> 0.10	> 0.10	> 0.10	> 0.10	> 0.10	> 0.10
	P-KS	> 0.10	> 0.10	> 0.10	> 0.10	> 0.10	> 0.10	> 0.10	> 0.10
300	A^2	3.25056	2.00082	3.25008	3.24957	3.24847	3.24728	3.24599	3.24461
	D	0.08667	0.06509	0.08665	0.08663	0.08661	0.08660	0.08660	0.08658
	P-AS	0.02777	0.09393	0.02779	0.02780	0.02783	0.02787	0.02790	0.02795
	P-KS	0.02909	> 0.10	0.02914	0.02920	0.02923	0.02928	0.02927	0.02932
500	A^2	5.18852	6.42725	5.18898	5.18940	5.19011	5.19065	5.19105	5.19130
	D	0.08365	0.08358	0.08364	0.08363	0.08365	0.08365	0.08366	0.08367
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
800	A^2	8.05050	10.06196	8.05056	8.05055	8.05030	8.04977	8.04896	8.04787
	D	0.08192	0.08086	0.08193	0.08195	0.08198	0.08201	0.08201	0.08197
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
1,000	A^2	10.21911	9.80539	10.21938	10.21954	10.21959	10.21927	10.21858	10.21753
	D	0.08271	0.07815	0.08274	0.08277	0.08280	0.08281	0.08283	0.08282
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01

Table 3.5 Lognormal distribution fitting to 2 mixed components of Weibull distributed samples.

n	Item	SPLD	SPLD with Boot	DCP					
				0.50%	1%	2%	3%	4%	5%
100	A^2	4.13089	5.22480	4.13035	4.12976	4.12846	4.12699	4.12534	4.12352
	D	0.18545	0.13401	0.18545	0.18545	0.18545	0.18542	0.18537	0.18530
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	0.05710	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
300	A^2	11.62556	10.71413	11.62469	11.62368	11.62122	11.61824	11.61473	11.61072
	D	0.18407	0.16742	0.18405	0.18406	0.18406	0.18404	0.18402	0.18402
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
500	A^2	19.13908	18.09580	19.13954	19.13978	19.13966	19.13873	19.13700	19.13450
	D	0.18334	0.16448	0.18334	0.18331	0.18328	0.18327	0.18326	0.18327
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
800	A^2	30.58181	29.84309	30.58216	30.58212	30.58091	30.57822	30.57410	30.56858
	D	0.18246	0.17097	0.18246	0.18246	0.18244	0.18241	0.18236	0.18230
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
1,000	A^2	37.94584	36.09822	37.94626	37.94618	37.94457	37.94107	37.93574	37.92861
	D	0.18131	0.16808	0.18134	0.18136	0.18141	0.18144	0.18141	0.18137
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01

Table 3.6 Lognormal distribution fitting to mixed components of Lognormal and Gamma distributed samples.

n	Item	SPLD	SPLD with Boot	DCP 0.50%	DCP 1%	DCP 2%	DCP 3%	DCP 4%	DCP 5%
100	A^2	8.36264	10.82667	8.35954	8.35222	8.33006	8.30356	8.27557	8.24709
	D	0.28165	0.26531	0.28157	0.28141	0.28101	0.28055	0.28009	0.27964
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
300	A^2	25.19776	26.46229	25.19049	25.17055	25.10745	25.03090	24.94982	24.86721
	D	0.28719	0.26883	0.28712	0.28698	0.28659	0.28617	0.28571	0.28525
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
500	A^2	41.39709	40.20186	41.38693	41.35875	41.26759	41.15474	41.03406	40.91052
	D	0.28791	0.26741	0.28785	0.28772	0.28733	0.28689	0.28642	0.28593
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
800	A^2	66.47282	64.77281	66.45435	66.40370	66.24287	66.04712	65.83950	65.62779
	D	0.28867	0.27133	0.28858	0.28840	0.28800	0.28758	0.28713	0.28670
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
1,000	A^2	82.91493	83.82496	82.89098	82.82538	82.61786	82.36636	82.10019	81.82904
	D	0.28849	0.27916	0.28842	0.28826	0.28784	0.28739	0.28691	0.28644
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01

Table 3.7 Lognormal distribution fitting to mixed components of Lognormal and Pareto distributed samples.

n	Item	SPLD	SPLD with Boot	DCP 0.50%	DCP 1%	DCP 2%	DCP 3%	DCP 4%	DCP 5%
100	A^2	3.08540	3.39554	3.08529	3.08517	3.08491	3.08463	3.08431	3.08396
	D	0.17810	0.11972	0.17812	0.17814	0.17818	0.17823	0.17827	0.17830
	P-AS	0.03261	0.02352	0.03261	0.03262	0.03263	0.03263	0.03264	0.03265
	P-KS	< 0.01	> 0.10	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
300	A^2	8.34191	7.64943	8.34183	8.34175	8.34153	8.34125	8.34092	8.34053
	D	0.16780	0.14207	0.16779	0.16778	0.16776	0.16776	0.16776	0.16778
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
500	A^2	13.65111	10.17753	13.65103	13.65092	13.65063	13.65023	13.64973	13.64914
	D	0.16685	0.13066	0.16683	0.16682	0.16680	0.16680	0.16680	0.16678
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
800	A^2	21.85418	23.69013	21.85442	21.85460	21.85479	21.85478	21.85455	21.85413
	D	0.16603	0.15096	0.16603	0.16603	0.16603	0.16604	0.16605	0.16601
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
1,000	A^2	27.27300	27.77180	27.27287	27.27268	27.27212	27.27131	27.27026	27.26899
	D	0.16545	0.14315	0.16546	0.16546	0.16546	0.16549	0.16549	0.16551
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01

Table 3.8 Lognormal distribution fitting to mixed components of Lognormal and Weibull distributed samples.

n	Item	SPLD	SPLD with Boot	DCP 0.50%	DCP 1%	DCP 2%	DCP 3%	DCP 4%	DCP 5%
100	A^2	4.26848	4.26702	4.26814	4.26777	4.26692	4.26591	4.26475	4.26347
	D	0.17643	0.15871	0.17643	0.17644	0.17645	0.17642	0.17637	0.17635
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	0.01636	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
300	A^2	12.20158	11.97442	12.20095	12.20020	12.19834	12.19601	12.19324	12.19004
	D	0.16639	0.15430	0.16638	0.16637	0.16637	0.16635	0.16634	0.16630
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
500	A^2	20.05170	16.98177	20.05237	20.05286	20.05332	20.05310	20.05223	20.05075
	D	0.16271	0.15071	0.16271	0.16269	0.16264	0.16263	0.16263	0.16264
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
800	A^2	31.68224	0.00000	31.68178	31.68101	31.67853	31.67486	31.67003	31.66407
	D	0.16061	0.00000	0.16061	0.16063	0.16059	0.16058	0.16060	0.16059
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
1,000	A^2	39.41399	39.61496	39.41407	39.41374	39.41187	39.40843	39.40347	39.39704
	D	0.15915	0.15700	0.15916	0.15917	0.15916	0.15915	0.15909	0.15906
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01

Table 3.9 Lognormal distribution fitting to mixed components of Gamma and Pareto distributed samples.

n	Item	SPLD	SPLD with Boot	DCP 0.50%	DCP 1%	DCP 2%	DCP 3%	DCP 4%	DCP 5%
100	A^2	2.04088	2.38613	2.04071	2.04052	2.04008	2.03956	2.03894	2.03825
	D	0.10608	0.09329	0.10605	0.10604	0.10609	0.10615	0.10615	0.10612
	P-AS	0.09035	0.05947	0.09037	0.09038	0.09042	0.09047	0.09052	0.09059
	P-KS	> 0.10	> 0.10	> 0.10	> 0.10	> 0.10	> 0.10	> 0.10	> 0.10
300	A^2	30.97450	30.52701	30.96801	30.95086	30.89688	30.83150	30.76228	30.69168
	D	0.32702	0.31086	0.32696	0.32684	0.32652	0.32618	0.32583	0.32546
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
500	A^2	51.57491	52.20833	51.56640	51.54279	51.46606	51.37082	51.26872	51.16397
	D	0.32832	0.31407	0.32827	0.32816	0.32786	0.32752	0.32717	0.32681
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
800	A^2	82.23466	79.78941	82.21950	82.17711	82.04146	81.87565	81.69931	81.51910
	D	0.32892	0.31571	0.32886	0.32873	0.32840	0.32805	0.32769	0.32732
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
1,000	A^2	102.82321	103.40175	102.80369	102.74892	102.57454	102.36248	102.13754	101.90794
	D	0.32919	0.31637	0.32913	0.32900	0.32868	0.32832	0.32795	0.32758
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01

Table 3.10 Lognormal distribution fitting to mixed components of Gamma and Weibull distributed samples.

n	Item	SPLD	SPLD with Boot	DCP 0.50%	DCP 1%	DCP 2%	DCP 3%	DCP 4%	DCP 5%
100	A^2	9.91611	8.83166	9.90379	9.86071	9.72101	9.55326	9.37798	9.20191
	D	0.29076	0.26196	0.29037	0.28926	0.28641	0.28312	0.27997	0.27672
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
300	A^2	29.11845	30.72714	29.08552	28.98218	28.63543	28.19967	27.73287	27.25684
	D	0.29388	0.29173	0.29363	0.29299	0.29100	0.28856	0.28606	0.28337
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
500	A^2	48.49356	51.95749	48.42520	48.23654	47.61966	46.85123	46.03050	45.19494
	D	0.29456	0.29122	0.29424	0.29340	0.29125	0.28864	0.28589	0.28323
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
800	A^2	77.32569	72.98106	77.19756	76.83886	75.69606	74.31018	72.85114	71.37711
	D	0.29456	0.29227	0.29419	0.29316	0.29032	0.28728	0.28405	0.28074
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
1,000	A^2	96.61183	93.90908	96.44155	95.99545	94.58801	92.87947	91.07743	89.25504
	D	0.29472	0.28922	0.29444	0.29360	0.29093	0.28801	0.28486	0.28206
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01

Table 3.11 Lognormal distribution fitting to mixed components of Pareto and Weibull distributed samples.

n	Item	SPLD	SPLD with Boot	DCP 0.50%	DCP 1%	DCP 2%	DCP 3%	DCP 4%	DCP 5%
100	A^2	5.75862	5.55956	5.75743	5.75620	5.75361	5.75083	5.74787	5.74475
	D	0.21185	0.16336	0.21182	0.21179	0.21174	0.21170	0.21167	0.21163
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
300	A^2	16.61916	16.88498	16.61876	16.61821	16.61665	16.61448	16.61177	16.60849
	D	0.21225	0.18487	0.21224	0.21224	0.21222	0.21219	0.21216	0.21214
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
500	A^2	27.61244	28.88741	27.61359	27.61451	27.61570	27.61604	27.61556	27.61428
	D	0.21162	0.19110	0.21162	0.21162	0.21159	0.21158	0.21159	0.21160
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
800	A^2	44.08424	45.74890	44.08392	44.08320	44.08056	44.07639	44.07071	44.06359
	D	0.21117	0.19289	0.21116	0.21114	0.21113	0.21112	0.21109	0.21107
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
1,000	A^2	54.94829	51.99963	54.94883	54.94885	54.94733	54.94379	54.93831	54.93094
	D	0.21034	0.19568	0.21034	0.21033	0.21033	0.21031	0.21029	0.21026
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01

Table 3.12 Lognormal distribution fitting to 3 mixed components of Lognormal distributed samples.

n	Item	SPLD	SPLD with Boot	DCP					
				0.50%	1%	2%	3%	4%	5%
150	A^2	4.48931	4.32716	4.48752	4.48572	4.48213	4.47854	4.47497	4.47139
	D	0.14444	0.09437	0.14439	0.14433	0.14424	0.14419	0.14415	0.14409
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	> 0.10	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
300	A^2	8.80618	9.00591	8.80635	8.80651	8.80675	8.80693	8.80704	8.80709
	D	0.13817	0.11239	0.13816	0.13815	0.13814	0.13811	0.13808	0.13805
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
600	A^2	17.37574	17.23807	17.37734	17.37890	17.38192	17.38480	17.38755	17.39016
	D	0.13454	0.11093	0.13455	0.13457	0.13465	0.13471	0.13474	0.13475
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
900	A^2	25.46682	26.30497	25.46961	25.47234	25.47764	25.48272	25.48759	25.49226
	D	0.13125	0.11949	0.13127	0.13128	0.13130	0.13133	0.13136	0.13139
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
1,200	A^2	34.30348	28.44723	34.30430	34.30507	34.30641	34.30750	34.30836	34.30899
	D	0.13092	0.11553	0.13091	0.13093	0.13096	0.13098	0.13097	0.13099
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01

Table 3.13 Lognormal distribution fitting to 3 mixed components of Gamma distributed samples.

n	Item	SPLD	SPLD with Boot	DCP					
				0.50%	1%	2%	3%	4%	5%
150	A^2	20.76018	21.67382	20.76271	20.75568	20.72551	20.68476	20.63839	20.58838
	D	0.34723	0.32125	0.34724	0.34710	0.34664	0.34602	0.34533	0.34461
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
300	A^2	41.51742	42.80644	41.50194	41.46639	41.36120	41.23442	41.09670	40.95219
	D	0.35034	0.32819	0.35023	0.34995	0.34923	0.34845	0.34762	0.34678
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
600	A^2	83.02808	84.60625	82.99630	82.93142	82.74127	82.51110	82.26093	81.99847
	D	0.35176	0.33251	0.35162	0.35135	0.35066	0.34989	0.34910	0.34830
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
900	A^2	124.54406	126.26601	124.49855	124.39603	124.09420	123.73290	123.34305	122.93572
	D	0.35222	0.33586	0.35207	0.35176	0.35096	0.35007	0.34915	0.34820
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
1,200	A^2	166.04475	168.41179	165.99757	165.87944	165.51675	165.07255	164.58778	164.07784
	D	0.35242	0.33565	0.35227	0.35197	0.35114	0.35023	0.34930	0.34836
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01

Table 3.14 Lognormal distribution fitting to 3 mixed components of Pareto distributed samples.

n	Item	SPLD	SPLD with Boot	DCP 0.50%	1%	2%	3%	4%	5%
150	A^2	4.92811	5.00806	4.93000	4.93186	4.93552	4.93908	4.94252	4.94585
	D	0.15115	0.12847	0.15117	0.15120	0.15126	0.15132	0.15137	0.15141
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	0.01838	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
300	A^2	9.47184	9.89068	9.47164	9.47139	9.47078	9.47001	9.46909	9.46802
	D	0.14693	0.13241	0.14693	0.14693	0.14689	0.14687	0.14683	0.14684
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
600	A^2	18.73960	16.96631	18.73810	18.73653	18.73324	18.72974	18.72603	18.72211
	D	0.14474	0.12337	0.14475	0.14475	0.14475	0.14476	0.14476	0.14477
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
900	A^2	28.00193	29.38318	27.99868	27.99533	27.98836	27.98100	27.97329	27.96522
	D	0.14364	0.13599	0.14363	0.14362	0.14358	0.14356	0.14355	0.14356
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
1,200	A^2	37.15138	37.33256	37.14992	37.14833	37.14476	37.14067	37.13609	37.13103
	D	0.14255	0.13263	0.14256	0.14256	0.14254	0.14253	0.14251	0.14251
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01

Table 3.15 Lognormal distribution fitting to 3 mixed components of Weibull distributed samples.

n	Item	SPLD	SPLD with Boot	DCP 0.50%	1%	2%	3%	4%	5%
150	A^2	10.13939	10.34525	10.14507	10.15062	10.16136	10.17162	10.18141	10.19074
	D	0.22568	0.20598	0.22575	0.22582	0.22597	0.22613	0.22625	0.22633
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
300	A^2	20.23490	18.59456	20.23273	20.23038	20.22514	20.21924	20.21266	20.20543
	D	0.22712	0.20239	0.22714	0.22716	0.22716	0.22711	0.22705	0.22695
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
600	A^2	39.84124	41.27956	39.83297	39.82443	39.80661	39.78779	39.76801	39.74732
	D	0.22408	0.20803	0.22405	0.22401	0.22395	0.22392	0.22388	0.22383
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
900	A^2	59.96905	56.94038	59.95949	59.94944	59.92787	59.90442	59.87913	59.85211
	D	0.22474	0.21355	0.22472	0.22470	0.22470	0.22468	0.22461	0.22456
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
1,200	A^2	79.73742	77.15067	79.73257	79.72710	79.71432	79.69917	79.68172	79.66202
	D	0.22363	0.20890	0.22361	0.22359	0.22357	0.22353	0.22352	0.22346
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01

Table 3.16 Lognormal distribution fitting to mixed components of Lognormal, Gamma and Weibull distributed samples.

n	Item	SPLD	SPLD with Boot	DCP 0.50%	DCP 1%	DCP 2%	DCP 3%	DCP 4%	DCP 5%
150	A^2	17.54073	16.98817	17.54576	17.54877	17.55054	17.54924	17.54649	17.54294
	D	0.36507	0.34467	0.36505	0.36499	0.36476	0.36445	0.36410	0.36375
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
300	A^2	34.95202	36.35885	34.94652	34.93584	34.90468	34.86751	34.82806	34.78756
	D	0.36811	0.34885	0.36804	0.36791	0.36756	0.36719	0.36681	0.36644
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
600	A^2	69.83733	69.27511	69.82344	69.80074	69.73789	69.66355	69.58457	69.50348
	D	0.36907	0.35310	0.36901	0.36889	0.36858	0.36823	0.36788	0.36752
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
900	A^2	104.78636	105.27593	104.76816	104.73453	104.63799	104.52344	104.40195	104.27729
	D	0.36930	0.35392	0.36922	0.36908	0.36875	0.36834	0.36790	0.36747
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
1,200	A^2	139.33974	139.79703	139.32499	139.29080	139.18486	139.05507	138.91578	138.77213
	D	0.36927	0.35557	0.36919	0.36904	0.36869	0.36827	0.36788	0.36748
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01

Table 3.17 Lognormal distribution fitting to mixed components of Gamma, Weibull and Pareto distributed samples.

n	Item	SPLD	SPLD with Boot	DCP 0.50%	DCP 1%	DCP 2%	DCP 3%	DCP 4%	DCP 5%
150	A^2	11.27102	12.39009	11.26685	11.26067	11.24421	11.22494	11.20458	11.18375
	D	0.22300	0.20798	0.22293	0.22286	0.22274	0.22261	0.22247	0.22234
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
300	A^2	22.36032	22.81061	22.35977	22.35394	22.33244	22.30516	22.27585	22.24573
	D	0.22243	0.20986	0.22245	0.22246	0.22248	0.22250	0.22253	0.22257
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
600	A^2	44.24601	41.25085	44.24877	44.24250	44.21220	44.17042	44.12413	44.07591
	D	0.22127	0.20776	0.22132	0.22134	0.22137	0.22143	0.22143	0.22142
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
900	A^2	66.22609	65.75394	66.22159	66.20160	66.13236	66.04541	65.95203	65.85607
	D	0.22044	0.21492	0.22044	0.22044	0.22046	0.22046	0.22047	0.22050
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
1,200	A^2	88.41672	89.06076	88.40876	88.38116	88.28839	88.17229	88.04745	87.91913
	D	0.21976	0.21058	0.21978	0.21979	0.21980	0.21977	0.21975	0.21975
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01

Table 3.18 Lognormal distribution fitting to mixed components of Weibull, Pareto and Lognormal distributed samples.

n	Item	SPLD	SPLD with Boot	DCP 0.50%	1%	2%	3%	4%	5%
150	A^2	6.28572	5.70624	6.28668	6.28756	6.28916	6.29050	6.29161	6.29250
	D	0.17972	0.15078	0.17977	0.17980	0.17986	0.17986	0.17986	0.17987
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
300	A^2	11.98518	12.41797	11.98519	11.98510	11.98462	11.98373	11.98244	11.98076
	D	0.17346	0.15978	0.17348	0.17350	0.17351	0.17354	0.17356	0.17360
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
600	A^2	24.56584	20.04620	24.56561	24.56522	24.56397	24.56210	24.55963	24.55658
	D	0.17255	0.15626	0.17255	0.17255	0.17254	0.17251	0.17252	0.17253
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
900	A^2	36.03477	34.90527	36.03336	36.03168	36.02754	36.02236	36.01621	36.00907
	D	0.16922	0.15106	0.16922	0.16921	0.16921	0.16919	0.16922	0.16921
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
1,200	A^2	48.27341	47.72015	48.27226	48.27079	48.26685	48.26165	48.25522	48.24761
	D	0.17022	0.16139	0.17023	0.17022	0.17022	0.17019	0.17020	0.17017
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01

Table 3.19 Lognormal distribution fitting to mixed components of Pareto, Lognormal and Gamma distributed samples.

n	Item	SPLD	SPLD with Boot	DCP 0.50%	1%	2%	3%	4%	5%
150	A^2	6.10410	5.51327	6.10241	6.09708	6.07998	6.05940	6.03776	6.01579
	D	0.21850	0.18465	0.21846	0.21834	0.21800	0.21759	0.21715	0.21666
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
300	A^2	11.67553	11.36463	11.67077	11.65936	11.62441	11.58271	11.53884	11.49427
	D	0.21929	0.20598	0.21924	0.21910	0.21875	0.21838	0.21800	0.21761
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
600	A^2	22.89235	22.31802	22.88289	22.86307	22.80293	22.72968	22.65165	22.57191
	D	0.21523	0.20548	0.21518	0.21504	0.21471	0.21430	0.21382	0.21337
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
900	A^2	34.38764	34.42791	34.37717	34.34828	34.25603	34.14365	34.02461	33.90335
	D	0.21570	0.20334	0.21562	0.21545	0.21509	0.21463	0.21416	0.21370
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
1,200	A^2	45.77310	45.41784	45.75870	45.72216	45.60621	45.46375	45.31202	45.15707
	D	0.21591	0.20954	0.21581	0.21564	0.21528	0.21483	0.21435	0.21389
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01

Table 3.20 Lognormal distribution fitting to mixed components of Lognormal, Gamma, Pareto and Weibull distributed samples.

n	Item	SPLD	SPLD with Boot	DCP					
				0.50%	1%	2%	3%	4%	5%
100	A^2	3.69190	3.34164	3.68765	3.68255	3.67053	3.65724	3.64346	3.62952
	D	0.21151	0.18386	0.21136	0.21116	0.21066	0.21012	0.20958	0.20906
	P-AS	0.01484	0.02510	0.01496	0.01511	0.01546	0.01585	0.01626	0.01667
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
300	A^2	10.07180	8.36302	10.06482	10.05503	10.02959	10.00011	9.96909	9.93749
	D	0.20773	0.17785	0.20763	0.20746	0.20706	0.20660	0.20612	0.20562
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
500	A^2	16.67129	14.67989	16.67164	16.66489	16.63865	16.60514	16.56913	16.53219
	D	0.20721	0.18813	0.20714	0.20695	0.20641	0.20579	0.20523	0.20466
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
800	A^2	26.51817	29.22734	26.46858	26.41662	26.30669	26.19112	26.07247	25.95251
	D	0.20825	0.19084	0.20805	0.20783	0.20730	0.20673	0.20615	0.20555
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
1,000	A^2	33.22040	30.50589	33.15650	33.08963	32.94848	32.80035	32.64848	32.49510
	D	0.20831	0.18550	0.20809	0.20786	0.20731	0.20671	0.20609	0.20551
	P-AS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
	P-KS	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01