

# A Robust Estimation for the Composite Lognormal-Pareto Model

Ro Jin Pak<sup>1,a</sup>

<sup>a</sup>Department of Applied Statistics, Dankook University

---

## Abstract

Cooray and Ananda (2005) proposed a composite lognormal-Pareto model to analyze loss payment data in the actuarial and insurance industries. Their model is based on a lognormal density up to an unknown threshold value and a two-parameter Pareto density. In this paper, we implement the minimum density power divergence estimation for the composite lognormal-Pareto density. We compare the performances of the minimum density power divergence estimator (MDPDE) and the maximum likelihood estimator (MLE) by simulations and an example. The minimum density power divergence estimator performs reasonably well against various violations in the distribution. The minimum density power divergence estimator better fits small observations and better resists against extraordinary large observations than the maximum likelihood estimator.

**Keywords:** Lognormal distribution, maximum likelihood estimation, minimum distance estimation, Pareto distribution.

---

## 1. Introduction

Insurance payment data in actuarial industries are typically highly skewed with a large upper tail. A lognormal distribution, a Pareto distribution or a generalized Pareto distribution is often used to model large claim data in actuarial industries (Klugman *et al.*, 1998; Hogg and Klugman, 1984; Resnick, 1997).

Researchers often use a Pareto model with a longer and thicker tail in order to model the larger loss data, while larger data with lower frequencies as well as smaller data with higher frequencies, are often modeled using a lognormal model. Cooray and Ananda (2005) developed a single composite lognormal-Pareto model and argued that it could be applied to model the entire range of loss data. In this article, we estimate the model of loss data by the minimum density power divergence estimation by Basu *et al.* (1998).

In Section 2, we introduce the minimum density power divergence estimator as well as its definition and robustness properties. We also illustrate how to implement this method for the lognormal-Pareto model. In Section 3, we compare the performance of the maximum likelihood estimator with the minimum density power divergence estimator through simulations and well-known Danish fire insurance loss data. In Section 4, conclusions are drawn.

---

<sup>1</sup> Professor, Department of Applied Statistics, Dankook University, Jukjeon-Dong, Suji-Gu, Yongin, 448-701, Korea.  
E-mail: [rjpak@dankook.ac.kr](mailto:rjpak@dankook.ac.kr)

## 2. Methodologies

### 2.1. Minimum density power divergence estimation

Basu *et al.* (1998) introduced the density power divergence between density functions  $g$  and  $f$  as

$$d_\alpha(g, f) = \int \left\{ f^{1+\alpha}(x) - \left(1 + \frac{1}{\alpha}\right) g(x) f^\alpha(x) + \frac{1}{\alpha} g^{1+\alpha}(x) \right\} dx \quad (\alpha > 0).$$

When  $\alpha = 0$ , the divergence  $d_0(g, f)$  is defined as

$$d_0(g, f) = \lim_{\alpha \rightarrow 0} d_\alpha(g, f) = \int g(x) \log \left\{ \frac{g(x)}{f(x)} \right\} dx.$$

Note that  $d_0(g, f)$  represents the Kullback-Leibler divergence.

Consider a parametric family of models  $\{F_\theta\}$ , indexed by the unknown parameter  $\theta$  (or, a vector of parameters), possessing densities  $\{f_\theta\}$  with respect to the Lebesgue measure. Given a random sample  $X_1, \dots, X_n$  from a distribution  $G$  ( $G$  may not belong to  $\{F_\theta\}$ ), the minimum density divergence estimator  $\hat{\theta}$  is generated by minimizing the estimated divergence

$$H_\alpha(\theta) = \int f_\theta^{1+\alpha}(x) dx - \left(1 + \frac{1}{\alpha}\right) n^{-1} \sum f_\theta^\alpha(X_i) \quad (2.1)$$

with respect to  $\theta$  over the parameter space. Then the estimating equations have the form

$$U_n(\theta) \equiv n^{-1} \sum u_\theta(X_i) f_\theta^\alpha(X_i) - \int u_\theta(z) f_\theta^{1+\alpha}(z) dz = 0,$$

where  $u_\theta(z) = \partial \log f_\theta(z) / \partial \theta$  is the score function.

The minimum density power divergence estimators are in fact M-estimators, and the corresponding  $\psi$  function is

$$\psi(x, \theta) = u_\theta(x) f_\theta^\alpha(x) - \int u_\theta(x) f_\theta^{1+\alpha}(x) dx.$$

Basu *et al.* (1998) show that under certain regularity conditions the minimum density power divergence estimator is a consistent estimator of the true parameter  $\theta_0$  and that its distribution, properly normalized, is asymptotically normal. The proposed class of density power divergences is indexed by a single parameter  $\alpha$  that controls the trade-off between robustness and efficiency. Choices of  $\alpha$  near zero afford considerable robustness while retaining efficiency close to that of maximum likelihood. That is, we can get the maximum likelihood estimator with

$$\psi(x, \theta) = u_\theta(x).$$

### 2.2. Composite lognormal-Pareto model

The composite model takes a lognormal density up to an unknown threshold value and a two-parameter Pareto density thereafter (Cooray and Ananda, 2005). The composite lognormal-Pareto model was constructed in the following manner. Let  $X$  be a random variable with the probability density function

$$f(x) = \begin{cases} c f_1(x), & \text{if } 0 < x \leq \theta, \\ c f_2(x), & \text{if } \theta < x < \infty, \end{cases}$$

where  $c$  is the normalizing constant,  $f_1(x)$  has the form of the two parameter lognormal density, and  $f_2(x)$  has the form of the two-parameter Pareto density, i.e.,

$$f_1(x) = \frac{(2\pi)^{-\frac{1}{2}}}{x\sigma} \exp\left[-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2\right], \quad x > 0$$

and

$$f_2(x) = \frac{\alpha\theta^\alpha}{x^{\alpha+1}}, \quad x > \theta,$$

where  $\theta, \mu, \sigma, \alpha$  are unknown parameter such that  $\theta > 0, \sigma > 0, \alpha > 0$ . Let us impose the continuity and differentiability conditions at  $\theta$ ,

$$f_1(\theta) = f_2(\theta), \quad f'_1(\theta) = f'_2(\theta),$$

where  $f'(\theta)$  is the first derivative of  $f(x)$  evaluated at  $\theta$ . These restrictions reduce the total of unknown parameters from four to two, and we get  $\ln \theta - \mu = \alpha\sigma^2$ , and  $\exp(-\alpha^2\sigma^2) = 2\pi\alpha^2\sigma^2$ . The composite density can be re-parameterized and re-written as

$$f(x) = \begin{cases} f_1(x), & \text{if } 0 < x \leq \theta, \\ f_2(x), & \text{if } \theta < x < \infty, \end{cases} \quad (2.2)$$

with

$$f_1(x) = \frac{\alpha\theta^\alpha}{(1 + \Phi(k))x^{\alpha+1}} \exp\left\{-\frac{\alpha^2}{2k^2} \ln^2\left(\frac{x}{\theta}\right)\right\}, \quad x > 0$$

and

$$f_2(x) = \frac{\alpha\theta^\alpha}{(1 + \Phi(k))x^{\alpha+1}}, \quad x > \theta,$$

where  $\Phi(\cdot)$  is the cumulative distribution function of the standard normal distribution and  $k$  is the known constant which is given by the positive solution of the equation  $\exp(-k^2) = 2\pi k^2$ . This value is  $k = 0.372238898$ . Here  $\alpha\sigma = k$  and  $c = 1/(1 + \Phi(k))$ . Therefore, this composite lognormal-Pareto density has only two unknown parameters  $\theta, \alpha > 0$ .

Let  $X_1, X_2, \dots, X_n$  be a random sample from the two-parameter composite lognormal-Pareto model described in equation (2.2). Then, the likelihood function is given by

$$L(\alpha, \theta) = \begin{cases} C_0\alpha^n\theta^{n\alpha} \left(\prod_{i=1}^n x_i^{-\alpha}\right) \exp\left[-\frac{\alpha^2}{2k^2} \sum_{i=1}^n \ln^2\left(\frac{x_i}{\theta}\right)\right], & \text{if } 0 < x \leq \theta, \\ C_0\alpha^n\theta^{n\alpha} \left(\prod_{i=1}^n x_i^{-\alpha}\right), & \text{if } \theta < x < \infty, \end{cases} \quad (2.3)$$

where  $C_0 = 1/[(\prod_{i=1}^n x_i)(1 + \Phi(k))^n]$ . The maximum likelihood estimators of  $\theta$  and  $\alpha$ ,  $\hat{\theta}_{ML}$  and  $\hat{\alpha}_{ML}$  respectively, can be obtained numerically by maximizing (2.3).

**Remark 1.** Both the divergence  $d_\alpha(g, f)$  and the composite lognormal-Pareto density happened to use the same letters  $\alpha$  and  $\theta$  as an index and a parameter, respectively. For notational convenience, let us use  $\beta$ , in stead of  $\alpha$  in the case of the composite lognormal-Pareto density in (2.2).

### 2.3. Application of minimum density power divergence estimation for the composite lognormal-Pareto model

In order to obtain the minimum density power divergence estimator for the composite lognormal-Pareto density, we need a corresponding  $H_\alpha(\theta)$ , defined in (2.1), which is

$$H_\alpha(\theta) = \begin{cases} \int_0^\theta f_1^{1+\alpha}(x)dx + \left(1 + \frac{1}{\alpha}\right)n^{-1} \sum f_1^\alpha(X_i), & \text{if } 0 < x \leq \theta, \\ \int_\theta^\infty f_2^{1+\alpha}(x)dx + \left(1 + \frac{1}{\alpha}\right)n^{-1} \sum f_2^\alpha(X_i), & \text{if } \theta < x < \infty, \end{cases} \quad (2.4)$$

where  $f_1$  and  $f_2$  are in (2.2).

The minimum density power divergence estimator of the composite lognormal-Pareto model is obtained by minimizing the above  $H_\alpha(\theta)$  with respect to  $\theta$ .

We can rewrite in detail  $H_\alpha(\theta)$  in (2.4) with

$$\begin{aligned} \int_0^\theta f_1^{1+\alpha}(x)dx &= \int_0^\theta \left\{ \frac{\beta\theta^\beta}{(1 + \Phi(k))x^{1+\beta}} \right\}^{1+\alpha} \exp \left\{ \frac{(1 + \alpha)\beta^2}{2k^2} \ln^2 \left( \frac{x}{\theta} \right) \right\} dx \\ &= \frac{(\beta\theta^\beta)^{1+\alpha}}{(1 + \Phi(k))^{1+\alpha}} \int_{-\infty}^0 \{ \theta \exp(y) \}^{-(1+\alpha)(1+\beta)} \exp \left\{ -\frac{(1 + \alpha)\beta^2}{2k^2} y^2 \right\} \theta \exp(y) dy \\ &\quad \text{with } y = \ln \left( \frac{x}{\theta} \right) \\ &= \frac{\beta^\alpha \theta^{-\alpha}}{(1 + \Phi(k))^{1+\alpha}} \left( \frac{2\pi k^2}{1 + \alpha} \right)^{\frac{1}{2}} \exp \left\{ \frac{(\alpha + \beta + \alpha\beta)^2 k^2}{2(1 + \alpha)\beta^2} \right\} \Phi \left( \frac{(\alpha + \beta + \alpha\beta)k^2}{(1 + \alpha)\beta^2} \right) \end{aligned}$$

and

$$\begin{aligned} \int_\theta^\infty f_2^{1+\alpha}(x)dx &= \int_\theta^\infty \left\{ \frac{\beta\theta^\beta}{(1 + \Phi(k))x^{\beta+1}} \right\}^{1+\alpha} dx \\ &= \frac{\beta^{1+\alpha} \theta^{-\alpha}}{(1 + \Phi(k))^{1+\alpha}} \frac{1}{(\alpha + \beta + \alpha\beta)}. \end{aligned}$$

After having taken the derivatives of  $H_\alpha(\theta)$  with respect to  $\beta$  and  $\theta$ , we can get the  $\psi$ -functions for  $\beta$  and  $\theta$  which are plotted in Figure 1. The plots are drawn by setting the true values of  $\beta$  and  $\theta$  being 0.5 and 50 (similar to Cooray and Ananda (2005)), respectively. The  $\psi$ -functions for maximum likelihood estimator are bounded on the right-hand side of the real line but are unbound near zero; however, these functions for minimum density power divergence estimator are bounded everywhere. Hampel, *et al.* (1986) witnessed that bounded and continuous  $\psi$ -functions result in qualitative robustness for ordinary M-estimates at certain distributions. Therefore, both maximum likelihood estimator and a minimum density power divergence estimator are robust against an outlier; however, a maximum likelihood estimator may not be robust against an inlier (smaller observations).

The Pareto model was originally designed to handle large observations, so  $\beta$  could cope with a large observation. The  $\theta$  is a threshold parameter so that the influence of an observation beyond that  $\theta$  should be bounded or limited. Robustness of both the maximum likelihood estimator and the minimum density power divergence estimator for larger observations is intrinsic; however, the maximum likelihood estimator would be influenced by a small observation as the  $\psi$ -functions of maximum likelihood estimator for both  $\beta$  and  $\theta$  are unbounded near zero.

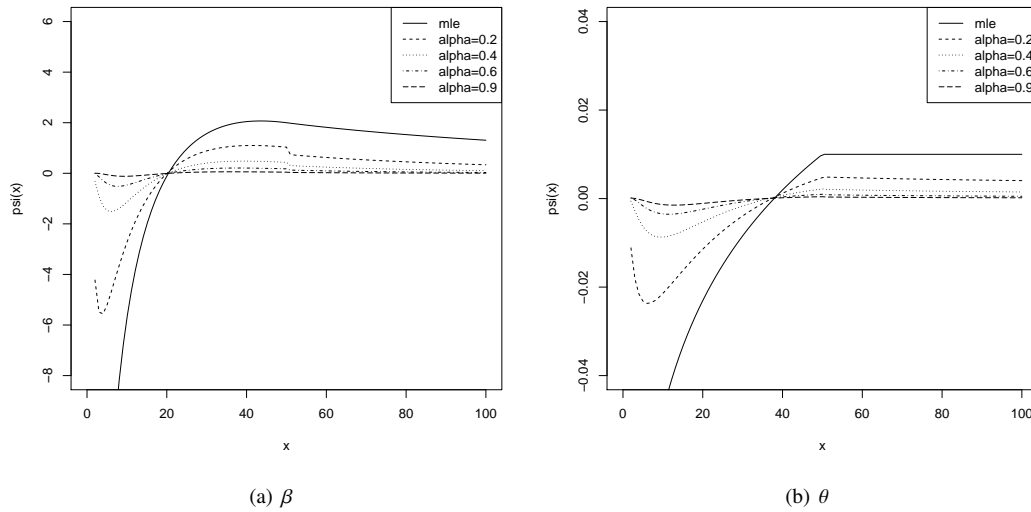


Figure 1:  $\psi$ -function for  $\beta$ (left) and  $\theta$ (right) for various choices of  $\alpha$ , assuming  $\beta = 0.5$  and  $\theta=50$ .

Table 1: Mean squared errors for the estimators under various contamination

Contamination m.s.e.	Uncontaminated		Contaminated at 1		Contaminated at 5	
	$\beta$	$\theta$	$\beta$	$\theta$	$\beta$	$\theta$
m.l.e.	0.0094	76.84	0.0442	309.83	0.0252	302.89
$\alpha = 0.1$	0.0132	83.81	0.0543	119.58	0.0268	258.19
$\alpha = 0.2$	0.0160	91.21	0.0656	145.50	0.0196	220.74
$\alpha = 0.3$	0.0207	99.74	0.0752	353.13	0.0113	209.97
$\alpha = 0.4$	0.0265	108.48	0.0699	476.35	0.0126	209.17
$\alpha = 0.5$	0.0325	116.81	0.0557	381.80	0.0136	211.23
$\alpha = 0.6$	0.0386	124.72	0.0460	233.30	0.0160	215.77
$\alpha = 0.7$	0.0448	132.26	0.0409	157.58	0.0205	223.40
$\alpha = 0.8$	0.0519	139.51	0.0400	136.58	0.0261	228.91
$\alpha = 0.9$	0.0592	145.97	0.0426	131.68	0.0321	230.97

### 3. Simulations and an Example

DATA 1. (Artificial data) We conducted a simulation experiment to investigate the effect of minimum density power divergence estimator under various model conditions. We generate 1000 samples of size 50 from the composite lognormal-Pareto model with true parameter values of  $\beta = 0.5$  and  $\theta = 50$  as we did for Figure 1. We also generated four sets of contaminated samples by replacing 10% of each sample generated above by 1 and 5, respectively. We intentionally contaminated the data with both small and large numbers. The maximum likelihood estimator and the minimum density power divergence estimator with  $\alpha = 0.1, 0.2, \dots, 0.8, 0.9$  are calculated, and their mean squared errors (m.s.e.) are computed (Table 1). When contamination is not assumed, the maximum likelihood estimator is superior to the minimum density power divergence estimator in terms of mean squared error. When contamination is assumed at 1 and 5, the minimum density power divergence estimators perform better than maximum likelihood estimators as  $\alpha$  near 0.6~0.8 (when contaminated at 1) or 0.3~0.6 (when contaminated at 5). This was expected due to the shape of the  $\psi$ -functions in Figure 1. Recall that a Bounded and continuous  $\psi$ -functions result in qualitative robustness for ordinary M-estimates at certain distributions (Hampel *et al.*, 1986).

Table 2: Estimates and goodness-of-fit statistics for Danish data

Method	$\hat{\beta}$	$\hat{\theta}$	CVM	K-S test
maximum likelihood estimator	1.4363	1.3851	1.0001	0.0273
$\alpha = 0.1$	1.5007	1.3624	1.0003	0.0241
$\alpha = 0.2$	1.5305	1.3468	1.0008	0.0197
$\alpha = 0.5$	1.5924	1.3085	1.0021	0.0051
$\alpha = 0.9$	1.6459	1.2707	1.0043	0.0011

Table 3: Chi-square test for goodness of fit:  $p$ -values

Upper limit	Maximum likelihood estimator	$\alpha = 0.2$	$\alpha = 0.5$	$\alpha = 0.9$
1.75	0.200	0.042	0.010	0.000
2.25	0.312	0.086	0.021	0.000
2.75	0.339	0.141	0.044	0.000
3.25	0.494	0.241	0.083	0.000
3.75	0.526	0.326	0.141	0.000
4.25	0.170	0.044	0.012	0.000
4.75	0.196	0.044	0.009	0.000
5.25	0.243	0.043	0.008	0.000
5.75	0.259	0.036	0.007	0.000
6.25	0.230	0.027	0.003	0.000
$\infty$	0.299	0.016	0.000	0.000

DATA 2. (Danish fire insurance loss data) The data set consists of 2492 large fire insurance claims in Denmark from January 3rd 1980 until December 31st 1990. The data set is retrieved from the data depository in R. In order to compare the models, we used (a) Kolmogorov-Smirnov test statistics (K-S) based on the statistic:

$$D_n = \sup_{1 \leq i \leq n} |F(x_i) - F_n(x_i)|$$

and (b) Cramer-von Mises test (CVM) based on the statistic:

$$W^2 = \int_{-\infty}^{\infty} (F_n(x) - F(x))^2 f(x) dx,$$

where  $f(x)$ ,  $F_n(x)$  and  $F(x)$  are the model density, empirical distribution, and model distribution, respectively. Both test statistics signify the level of fitness as long as the statistics are sufficiently small. Both statistics for Danish data are recorded in Table 2. Additionally, the data is binned with upper limits (in millions) 1.25, 1.75, 2.25, 2.75, 3.25, 3.75, 4.25, 4.75, 5.25, 5.75, 6.25 and  $\infty$ , and the  $p$ -values of the chi-square test for goodness-of-fit using expected and observed frequencies in every bin until each upper limit, are recorded in Table 3. The larger the  $p$ -value, the better a model fits a data set. Table 4 lists the empirical and fitted models quantiles for the Danish data. The estimates were calculated in R using the `nlm` function and the test statistics are computed based on various add-on packages in Ricci (2005). Some statistics in this example are slightly different from those of Cooray and Ananda (2005) and McNeil (1997). McNeil (1997) analyzed the upper portion of this data, which consist of 2156 losses of over one million Danish Krone.

We can summarize the results as follows.

Table 2: We use the Kolmogorov-Smirnov (K-S) test statistic as a base line to compare the models. The Smaller value of the K-S test statistic implies the better fit for a given model. Critical values of the K-S test statistic are  $1.22/\sqrt{2492} \approx 0.0244$ ,  $1.36/\sqrt{2492} \approx 0.0272$  and  $1.63/\sqrt{2492} \approx 0.0326$ , respectively, at the 10, 5, and 1% levels of significance. All models we consider fit the data well at

Table 4: Empirical and fitted models quantiles for the Danish data

Quantile	Empirical	Maximum likelihood estimator	$\alpha = 0.2$	$\alpha = 0.5$	$\alpha = 0.9$
.90	5.0804	4.8492	4.4867	4.1260	3.7653
.95	8.4537	7.7336	6.8914	6.1700	5.8093
.99	24.5785	24.3251	19.6360	17.3516	15.4279
.999	144.6576	120.2710	88.7699	73.3801	62.5592
.9999	263.2504	596.874	399.4516	310.9603	253.0080

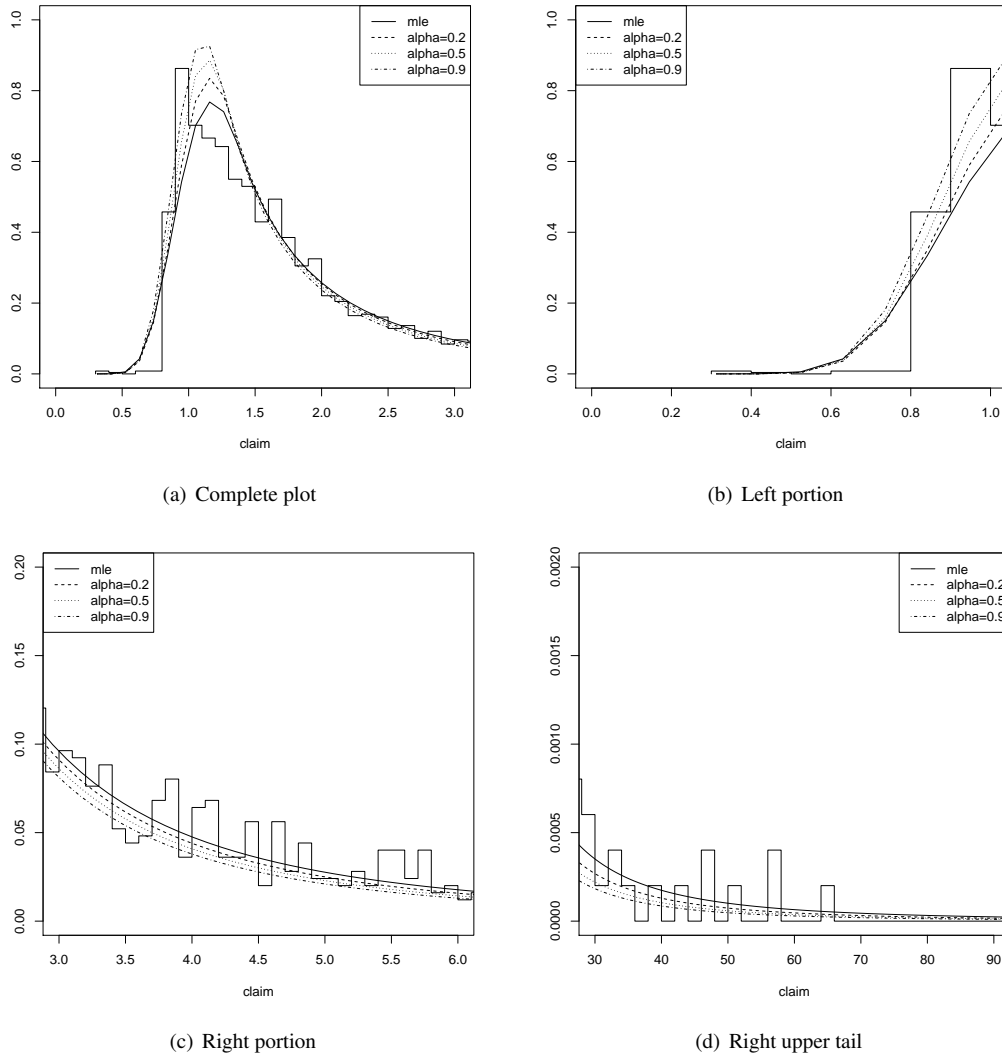


Figure 2: Histogram and estimated densities for various  $\beta$ .

5% and 1% levels of significance; in addition, no significant difference in the CVM test statistics is observed.

Figure 2, Table 3: The models by the minimum density power divergence estimator better fit smaller claims and can identify the peak more clearly than the model by the maximum likelihood

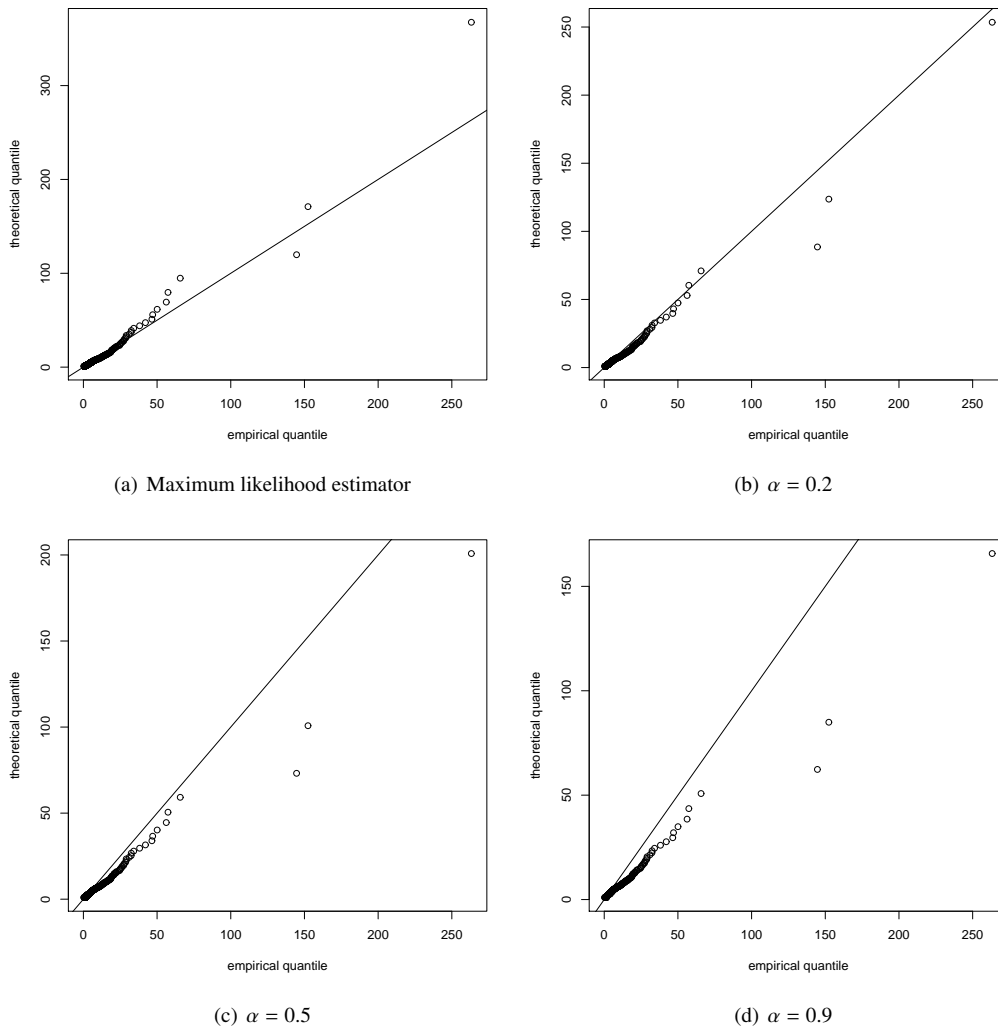


Figure 3: Q-Q plots Danish data to the composite lognormal-Pareto model by maximum likelihood estimator and minimum density power divergence estimator with various  $\alpha$ .

estimator. As a whole, the models by the maximum likelihood estimator and by the minimum density power divergence estimator with  $\alpha = 0.2$  fit 1% level of significance the data better than others. The minimum density power divergence estimator with high values of  $\alpha$  somehow over-fit the peak so that  $\chi^2$  test statistics are inflated ( $p$ -value becomes too small).

Figure 2, Table 4: We admit that the models by the minimum density power divergence estimator underestimate the tail on the right. However, it can be said in reverse that the model by the maximum likelihood estimator overestimates the right upper tail (it produces the highest quantile value of all; it has a longer tail), which should be estimated as exactly as possible in order for insurance companies to predict future losses. This occurs because the minimum density power divergence estimator is less affected by larger and smaller observations than is the maximum likelihood estimator as the nature of a  $\psi$ -function (Figure 1).



Figure 3: The estimated quantiles are plotted against the ordered observations. The estimated  $p^{\text{th}}$  quantile is estimated from the fitted model, and here  $p = (r - 0.5)/n$ ,  $r = 1, 2, \dots, n$ . We can observe that the model by the minimum density power divergence estimator with  $\alpha = 0.2$  fits both the main body of data and the extreme quantiles better than other models.

In conclusion, we recommend the model by the minimum density power divergence estimator with  $\alpha = 0.2$  due to small value of K-S and CVM, the quantiles and the quantile plot.

#### 4. Conclusions

The maximum likelihood estimator and the minimum density power divergence estimator for the composite lognormal-Pareto model are studied and compared by simulations and an example. Based on the composite lognormal-Pareto model, both the maximum likelihood estimator and the minimum density power divergence estimator fit the data well, but the minimum density power divergence estimator is better than the maximum likelihood estimator in terms of avoiding the influence of extremely small and large observations. Since the maximum likelihood estimation underestimates small observations, the estimated density by the maximum likelihood estimator should overestimates larger observations in order for the area under the density curve to be 1. For the purpose of predicting large losses reasonably, we claim that the minimum density power divergence estimator is useful when considered alongside the maximum likelihood estimator. There are two more additional contributions on the composite lognormal-Pareto model by Scollnik (2007) and Pigeon and Denuit (2011). The comparison of estimation methods for those models should have been attempted for a contribution to the field of actuarial science or insurance; however, we have focused on the applicability of the minimum density power divergence estimator on composite models that considered only the original and basic model by Cooray and Ananada (2005).

#### References

- Basu, A., Harris, I. R., Jort, N. L. and Jones, M. C. (1998). Robust and efficient estimation by minimizing a density power divergence, *Biometrika*, **85**, 549–559.
- Cooray, K. and Ananda, M. M. A. (2005). Modeling actuarial data with a composite lognormal-Pareto model, *Scandinavian Actuarial Journal*, **5**, 321–334.
- Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions*, John Wiley and Sons, New York.
- Hogg, R. V. and Klugman, S. A. (1984). *Loss Distribution*, John Wiley and Sons, New York.
- Klugman, S. A., Panjer, H. H. and Willmot, G. E. (1998). *Loss Models From Data to Decisions*, John Wiley and Sons, New York.
- McNeil, A. (1997). Estimating the tails of loss severity distributions using extreme value theory, *ASTIN Bulletin*, **27**, 117–137.
- Pigeon, M. and Denuit, M. (2011). Composite Lognormal-Pareto model with random threshold, *Scandinavian Actuarial Journal*, **3**, 177–192.
- Resnick, S. I. (1997). Discussion of the Danish data on large fire insurance losses, *ASTIN Bulletin*, **27**, 139–151.
- Ricci, V. (2005). Fitting Distributions with R, <http://cran.r-project.org>.
- Scollnik, D. P. (2007). On composite lognormal-Pareto models, *Scandinavian Actuarial Journal*, **1**, 20–33.