Bayesian Predictive Modeling for Exponential-Pareto Composite Distribution

by M. S. Aminzadeh and Min Deng

ABSTRCT

Composite distributions have well-known applications in the insurance industry. In this article a composite Exponential-Pareto distribution is considered and the Bayes estimator under the squared error loss function is derived for the parameter θ, which is the boundary point for the supports of the two distributions. A predictive density is constructed under Inverse-Gamma prior for the parameter θ and the density is used to estimate Value at risk(VaR). Goodness-of-fit of the composite model is verified for a generated data set . Accuracies of the Bayes and VaR estimates are assessed via simulation studies. The "best" value for hyper-parameters of IG prior distribution are found via an upper bound on the variance of the prior distribution. Simulation studies indicate that the Bayes estimator is consistently more accurate than MLE when the "best" values of hyper-parameters are used in the Bayes estimator.

KeyWords: MLE, Inverse-Gamma, Bayes estimate, Exponential-Pareto composite density, predictive density, VaR

INTRODUCTION

There are two important topics in Actuarial Science that are used in the insurance industry: using data to construct a model and estimating parameters of interest based on the model. In this paper, both topics for a composite distribution are presented.

Due to the fact that insurance data is skewed with the fat tail, it is very difficult to find a classical parametric distribution to model the insurance data. The central limit theorem is not very useful for the insurance industry because insurance data usually has high frequencies for small losses and large losses occur with small probabilities. Therefore, in order to address modeling insurance data or data with similar characteristics, many researchers developed composite models. Klugman, Panjer, and Willmot (2012) discussed how to use data to build a model in actuarial science field and insurance industry. In this book, they discussed many important concepts, including Value at Risk (VaR). VaR is one of the most important risk measures in the business world. It is the percentile of the distribution of the losses. It gives actuaries and risk managers "the chance of an adverse outcome" and helps them to make a decision. In this article we use the Bayesian inference to estimate VaR based on a predictive model.

Teodorescu and Vernic (2006) introduced the composite Exponential-Pareto distribution, which is a one-parameter distribution. In the article, they derived maximum likelihood estimator for the unknown parameter θ, which presents the boundary for small and large losses in a data set. The authors in a subsequent article (2009) worked on different types of Exponential-Pareto composite models. Both models have one unknown parameter. Many other composite distributions, such as Weibull-Pareto and Lognormal-Pareto, to name a couple, were proposed by researchers. Preda and Ciumara (2006) introduced the Weibull-Pareto and Lognormal-Pareto composite models. They pointed out that these models could be used to model actuarial data collected from the insurance industry. They compared the models with different parameters. In the article, algorithms were developed to find the maximum likelihood estimates for two unknown parameters and the accuracy of the two models were compared.

Pareto distribution has a fatter tail than Normal distribution. Therefore, Pareto is a good model to capture large losses in insurance data, but it is not good for the small losses with high frequencies. That is why many other distributions, such as Exponential, Lognormal, and Weibull, were combined with the Pareto distribution to model losses with small values in a data set.

The aim of this paper is to develop a Bayesian predictive density based on the composite Exponential-Pareto distribution. In Bayesian framework, a predictive density is developed via the composite density and then based on a random sample, Bayes estimate and Value at Risk (VaR) are estimated . Section 2 provides the development of a posterior pdf via Inverse-Gamma prior for θ. In this section it is explained how to use a search method to compute the Bayes estimate for θ based on a sample. In Section 3, the predictive density is derived for Y, where its realization y is considered as a future observation form the composite distribution. It is noted that the first moment $E[Y|\underline{x}]$ of the predictive pdf is undefined. This is due to the fact that for the composite pdf, $E[X]$ is also undefined. Section 4 provides a summary of simulation studies. In this section, the accuracy of VaR is investigated and it is shown through simulation studies that Bayes estimator of θ under squared error loss function is consistently more accurate than MLE. Also, a method for choosing "best" values of hyper parameters of the prior distribution to get an accurate Bayes estimate is given. Section 5 contains a numerical example for computational purposes. Two Mathematica codes are given in Appendix. One for computation of Bayes estimate and VaR using a single sample . The other code is for simulation which can be used to search for "best" values of hyper parameters to find an accurate Bayes estimate.

Teodorescu and Vernic (2006) developed the composite Exponential-Pareto model as follows:

Let X be a random variable with the probability density function,

$$f_X(x) = \begin{cases} cf_1(x) & 0 < x \leq \theta \\ cf_2(x) & \theta \leq x < \infty \end{cases}$$

where

$$f_1(x) = \lambda e^{-\lambda x}, \quad x > 0, \lambda > 0$$

and

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

$f_1(x)$ is the pdf of Exponential distribution with parameter λ and $f_2(x)$ is the pdf of Pareto distribution with parameters θ and α.

In order to make the composite density function smooth, it is assumed that the pdf is continuous and differential at θ. That is,

$$f_1(\theta) = f_2(\theta), f_1^{'}(\theta) = f_2^{'}(\theta)$$

Solving the above equations simultaneously gives,

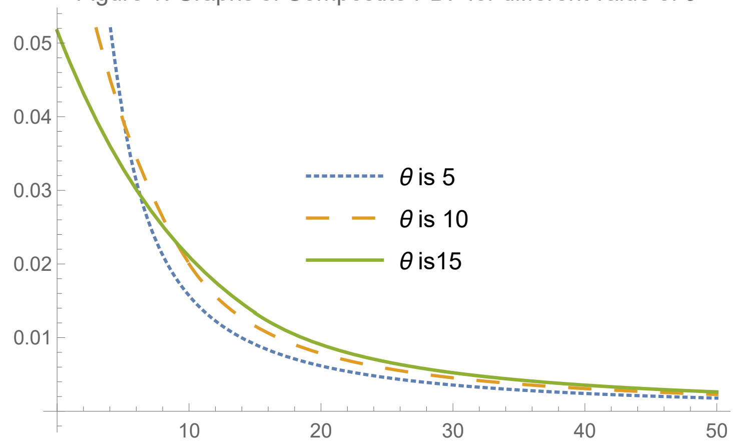
$$\lambda\theta = 1.35, \quad \alpha = 0.35, c = .574$$

As a result the initial three parameters are reduced to only one parameter θ for the composite Exponential-Pareto distribution whose pdf is given by

$$f_X(x|\theta) = \begin{cases} \frac{.775}{\theta} e^{\frac{-1.35x}{\theta}} & 0 < x \leq \theta \\ \frac{.2\theta^{.35}}{x^{1.35}} & \theta \leq x < \infty \end{cases} \tag{1}$$

Figure 1 provides graphs of the composite pdf for different values of θ. Figure 1 reveals that as θ increase, the tail of the pdf becomes heavier. This implies that the percentile at a specific level, say .99, is increasing in θ.

Figure 1: Graphs of Compostite PDF for different value of θ



Teodorescu and Vernic (2006) compared the Exponential distribution with the composite Exponential-Pareto distribution and MLE for θ was derived via an ad-hoc procedure which uses a search method. It was concluded that when $\theta = 10$, the composite distribution fades to zero slower than the Exponential distribution. The implication of this result is that the composite distribution could be a better choice to model insurance data due to large losses. In this situation, we could be able to avoid charging insufficient premiums to cover potential future losses.

2. Derivation of posterior density and Bayes estimator

Let $x_1, ..., x_n$ be a random sample for the composite pdf in (1) and without loss of generality assume that $x_1 < x_2 < ... < x_n$ is an ordered random sample from the pdf in (1). The likelihood function, also given in Teodorescu and Vernic(2006), is written as

$$L(\underline{\mathrm{x}}|\theta) = k\theta^{.35n-1.35m}e^{-1.35\sum_{i=1}^{n} x_i/\theta} \tag{2}$$

where $k = \frac{.2^{n-m}(.775)^m}{\prod_{i=m+1}^{n} x_i^{1.35}}$. To formulate the likelihood function it is assumed that there is an $m(m = 1, 2, ..., n-1)$ such that in the ordered sample $x_m \leq \theta \leq x_{m+1}$.

To derive a posterior distribution for θ, we use a conjugate prior Inverse-Gamma distribution for θ with the pdf

$$\rho(\theta) = \frac{b^a \theta^{-a-1} e^{-b/\theta}}{\Gamma(a)}, b > 0, a > 0. \tag{3}$$

From (2) and (3), the posterior pdf can be written as

$$f(\theta|\underline{x}) = L(\underline{x}|\theta) * \rho(\theta) \propto e^{-\frac{b+1.35\sum_{i=1}^{m} x_i}{\theta}} \theta^{-(a-.35n+1.35m)-1}. \tag{4}$$

It can be seen from (4) that the expression on the right side is the kernel of Inverse-Gamma(A, B), where $A = (a-.35n+1.35m)$ and $B = (b+1.35\sum_{i=1}^{m} x_i)$. Therefore under squared error loss function, the Bayes estimator for θ is

$$\hat{\theta}_{Bayes} = E[\theta|\underline{x}] = \frac{B}{A-1} = \frac{b + 1.35\sum_{i=1}^{m} x_i}{a - .35n + 1.35m - 1}. \tag{5}$$

Given an ordered sample $x_1 < x_2 < ... < x_n$, we need to identify the correct value of m in order to compute $\hat{\theta}_{Bayes}$. We use the following algorithm:

1. Start with $j = 1$, check to see if $x_1 \leq \frac{B}{A-1} \leq x_2$, if yes, then $m = 1$. otherwise go to step 2.

2. For $j = 2$, check to see if $x_2 \leq \frac{B}{A-1} \leq x_3$, if yes, then $m = 2$, otherwise let $j = 3$ and continue until to find the correct value for m. The idea is to find the value for j so that $x_j \leq \frac{B}{A-1} \leq x_{j+1}$. The Mathematica code used for simulation studies in this article is based on the above algorithm to compute $\hat{\theta}_{Bayes}$.

3. **Derivation of predictive density**

Let y be a realization of the random variable Y from the composite density. Based on the observed sample data $\underline{x}$, we are interested in deriving the predictive density $f(y|\underline{x})$. In the context of Bayesian framework, predictive density is used to estimate measures such as $E[Y|\underline{x}]$ or Var$[Y|\underline{x}]$ or other measures such as VaR that is considered in this article.

$$f(y|\underline{x}) = \int_0^\infty f(\theta|\underline{x}) f_Y(y|\theta) d\theta$$

where

$$f_Y(y|\theta) = \begin{cases} \frac{.775}{\theta} e^{\frac{-1.35y}{\theta}} & 0 < y \leq \theta \\ \frac{.2\theta^{.35}}{y^{1.35}} & \theta \leq y < \infty \end{cases}$$

as a result we get

$$f(\theta|\underline{x}) f(y|\theta) = \begin{cases} \frac{.775}{\Gamma(A)\theta} e^{\frac{-1.35y}{\theta}} \theta^{-(A+1)} B^A e^{-B/\theta} & y < \theta < \infty \\ \frac{.2\theta^{.35}}{\Gamma(A) y^{1.35}} \theta^{-(A+1)} B^A e^{-B/\theta} & 0 < \theta < y \end{cases}$$

which reduces to

$$f(\theta|\underline{x}) f(y|\theta) = \begin{cases} \frac{.775 B^A \Gamma(A+1)}{\Gamma(A)(B+1.35y)^{A+1}} h_1(\theta|A+1, B+1.35y) & y < \theta < \infty \\ \frac{.2 B^A \Gamma(A-.35)}{\Gamma(A) y^{1.35} B^{A-.35}} h_2(\theta|A-.35, B) & 0 < \theta < y \end{cases}$$

where $h_1(\theta|A+1, B+1.35y)$ is the pdf of Inverse-gamma with parameters $A+1$ and $B+1.35y$. Also $h_2(\theta|A-.35, B)$ is the pdf of Inverse-gamma distribution with parameters $A-.35$ and B. Using the above results the predictive density $f(y|\underline{x})$ is given by

$$f(y|\underline{x}) = \int_0^y f(\theta|\underline{x}) f_Y(y|\theta) d\theta + \int_y^\infty f(\theta|\underline{x}) f_Y(y|\theta) d\theta$$
$$= K_2(y) H_2(y|A-.35, B) + K_1(y)(1 - H_1(y|A+1, B+1.35y) \quad (6)$$

$$K_1(y) = \frac{.775A * B^A}{(B+1.35y)^{A+1}}, \quad K_2(y) = \frac{.2B^{.35}\Gamma(A-.35)}{\Gamma(A) y^{1.35}}.$$

H_2 is the cdf of Inverse-gamma distribution with parameters $(A-.35, B)$ and H_1 is the cdf of inverse-gamma distribution with parameters $(A+1, B+1.35y)$. Similar to the composite density for which $E[X]$ is undefined, $E[Y|\underline{x}]$ is also undefined for the predictive pdf.

4. **Simulation**

To assess accuracy of θ_{Bayes} as well as VaR, simulation studies are conducted.

For selected values of n, θ and the "best" values of hyper-parameters (a, b), $N = 300$ samples from the composite density (1) are generated .

a. For each generated sample, observations are ranked and the correct value for m through a search method is identified so that $x_m \leq \frac{B}{A-1} \leq x_{m+1}$. Simulation results indicate that accuracy of Bayes estimate does depend on the selected values for hyper-parameters a and b. Below is a method proposed to produce an accurate Bayes estimate.

For the Inverse-gamma prior distribution we have

$$E[\theta] = \frac{b}{a-1}, \quad \text{Var}(\theta) = \frac{b^2}{(a-1)^2(a-2)} \tag{7}$$

Also it is noted $A = (a - .35n + 1.35m)$ as one of the parameters of posterior distribution must be positive, and m takes values 1,2,...,$(n-1)$. Therefore, $a > .35n - 1.35m$ and as a result the value for a should be at least $.35n - 1.35$ to avoid computational errors. For example for $n = 50$, a should be at least 17. Simulation studies indicate that for a given sample size n, generally larger values of a provide more accurate Bayes estimate. However, as a increases, for a desired level of variance in (7), b would need to increase. And this causes the expected value in (7) to increase. Simulations results as shown in Table 1 for $\theta = 5, 10$ indicate that accuracy of Bayes estimate depends on the choice of a (b is found via a) which depends on n as well as θ. To identify a "good" value for a we propose the following method. Simulations indicate that large θ produces larger sample points as a results both Bayes estimate of θ and its MSE increase. To overcome this we propose an upper bound on $\text{Var}(\theta)$ and let the upper bound be a decreasing function of n. Let

$$\text{Var}(\theta) = \frac{b^2}{(a-1)^2(a-2)} \leq \frac{1}{n^{1/3}} \tag{8}$$

The idea is to have a decreasing function of n on RHS of (8). Other functions for an upper bound that are decreasing in n may also work, but simulations show that the above choice works very well. For a selected value a, (8) is used to find b by solving

$$\frac{b^2}{(a-1)^2(a-2)} = \frac{1}{n^{1/3}}$$

As it can be seen from Table 1, for selected values of n and θ, MSE of the Bayes estimate has a minimum when a takes its "best" value. For selected values of n and θ, Table 1 lists the turning points of the MSE. The Mathematica code for simulation given in Appendix can be used to make a list of "best" values of a as a function of n and θ. In this article n= 20, 50, 100 & θ=5,10 are considered. Note that since in practice the true value of θ is unknown, it can be guessed by using its MLE. In particular MLE would be a good guess if n is large. Note that in simulation studies, in each iteration, MLE was not used to to find a and b, as the variability of MLE would create more variability in the Bayes estimate. For example, in simulations studies for each generated sample and a selected value of a, we tried to find b via $b = (a-1)\hat{\theta}_{MLE}$ $\big($due to the $E[\theta]$ formula in (7)$\big)$, but this way variability of $\hat{\theta}_{Bayes}$ becomes extremely large and the results are not desirable.

b. As mentioned earlier, VaR at .90 level is a useful measure for big losses in insurance industries. We used .70 to compute Bayes estimate of Var. Even with .70 and N =300 interations, it takes several hours for the program to run. Computation of VaR via (6) cannot be done analytically and one has to use a numerical method. The idea is to find the value of y in (6) so that $\int_0^y f(y|\underline{\mathrm{x}})dy = .70$. Mathematica is used to find an estimated value of y based on selected input parameters and a generated sample from the composite density (1). As mentioned above, computation of VaR at .90 level is possible but it requires an extended computer memory (in particular when a is large) to solve equations in the code numerically. MLE of θ is derived in Teodorescu and Vernic (2006). The MLE should satisfy $x_m \leq \frac{1.35\sum_{i=1}^m x_i}{1.35m - .35n} \leq x_{m+1}$, and it can be found via a search method for $m = 1, 2, ..., (n-1)$. Simulation studies indicate that among many generated samples ($k = 300$), only a few samples (1-2 out of 300) lead to extremely large MLE values and as result, MSE of MLE is inflated. Even when such outlier do not occur, Bayes estimator still outperforms MLE.

A summary of simulation studies for accuracy of Bayes estimator, MLE, and VaR is given in Tables 2 and 3. Examination of Tables 2 and 3 reveals that for larger n, $\hat{\theta}_{MLE}$ and $\hat{\theta}_{Bayes}$ are more accurate. In both Tables 2 and 3 "best" values of a from Table 1 are used to compare the Bayes estimate with MLE. Tables 2 and 3 also reveal that $\bar{\hat{\theta}}_{Bayes}$, average Bayes estimates is more closer to the actual value of θ than $\bar{\hat{\theta}}_{MLE}$, average of MLEs. For all values of n in Tables 2, 3, we note that MSE(Bayes)=$\xi(\hat{\theta}_{Bayes})$ is considerably smaller than MSE(MLE) =$\xi(\hat{\theta}_{MLE})$.

Table 1: MSE of Bayes estimate as a function of a and θ

n	θ	a	$\xi(\hat{\theta}_{Bayes})$	n	θ	a	$\xi(\hat{\theta}_{Bayes})$
20	5	17	3.59	20	10	17	43.37
20	5	30	1.80	20	10	30	38.01
20	5	50	.264	20	10	50	29.08
20	5	70	**.063**	20	10	70	21.91
20	5	80	.331	20	10	80	18.98
-	-	-	-	20	10	250	.106
-	-	-	-	20	10	260	**.020**
-	-	-	-	20	10	280	.051
50	5	50	.491	50	10	50	28.67
50	5	70	.037	50	10	70	24.46
50	5	80	**.019**	50	10	80	20.17
50	5	90	.120	50	10	70	24.46
-	-	-	-	50	10	340	.043
-	-	-	-	50	10	350	**.006**
-	-	-	-	50	10	370	.043
100	5	60	.095	100	10	350	.876
100	5	70	.039	100	10	410	.072
100	5	80	**.033**	100	10	420	**.027**
100	5	90	.077	100	10	450	.032

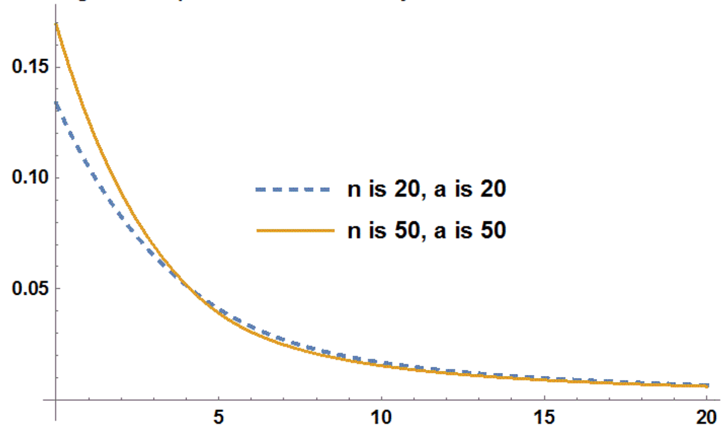
Table 2: Accuracy of Bayes Estimator, MLE, and VaR

n	a	$\overline{\hat{\theta}_{Bayes}}$	$\xi(\hat{\theta}_{Bayes})$	$\overline{VaR}$	Std(VaR)	$\overline{\hat{\theta}_{MLE}}$	$\xi(\hat{\theta}_{MLE})$
				θ=5			
20	70	5.24	.064	32.65	1.01	8.03	25.82
50	80	5.15	.015	30.22	1.21	7.17	18.450
100	80	5.07	.036	27.99	1.36	6.20	3.28

Table 3: Accuracy of Bayes Estimator, MLE, and VaR

n	a	$\overline{\hat{\theta}_{Bayes}}$	$\xi(\hat{\theta}_{Bayes})$	$\overline{VaR}$	Std(VaR)	$\overline{\hat{\theta}_{MLE}}$	$\xi(\hat{\theta}_{MLE})$
				θ=10			
20	260	9.89	.081	63.36	.503	13.71	71.72
50	350	9.92	.07	62.43	.535	12.30	14.7
100	420	9.94	.0233	61.93	.584	12.45	12.33

Figure 2:Graphs of Predictive Density for Selected Values of n and a

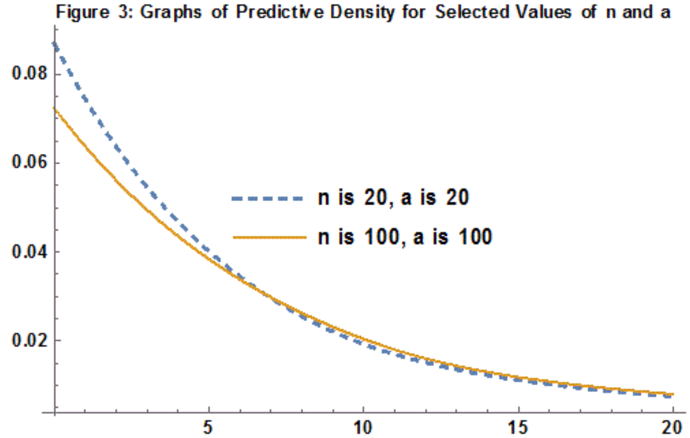


Graphs are based on generated samples from composite PDF(θ = 5)

Figures 2-4 provide graphs of the predictive density for different values on θ and a. Figures 2 and 3 reveal that as n increase, the tail of the predictive density becomes heavier and a result at a specific level, say .99, VaR increases.

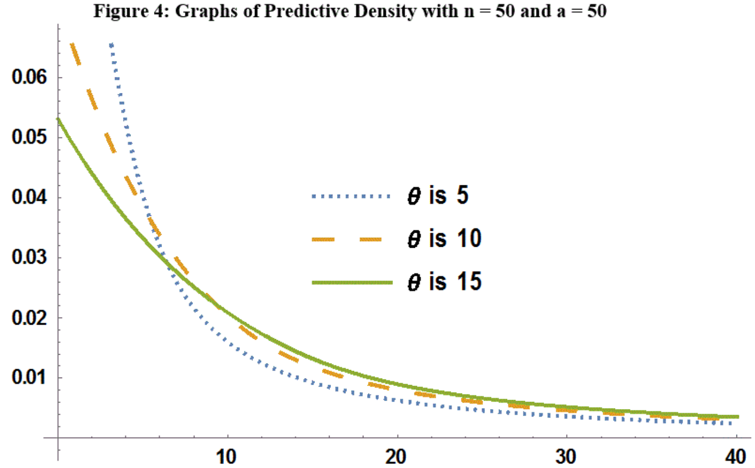
Simulation studies confirm the conclusion. Figure 4 provides graphs of the predictive density for fixed values $n = 50, a = 50$ and three different values of θ based on which samples from the composite pdf are generated. Similar to Figure 1, as θ increase the tail of the predictive density becomes heavier and this causes VaR at a specific level, say .99 increase.

Figure 3: Graphs of Predictive Density for Selected Values of n and a



Graphs are based on generated samples from composite PDF(θ = 10)

Figure 4: Graphs of Predictive Density with n = 50 and a = 50



Graphs are based on generated samples from composite PDF for different θ

5. Numerical Example

The data set in Table 4 is a random sample of size 200 generated from Exponential-Pareto composite pdf (1) with parameter $\theta = 5$. For this data set, the value of m is 86. The Bayes estimate is 4.9878 and MLE is 4.935488. Simulation studies indicate that Bayes estimator is more accurate than MLE. Due to the long-tail composite distribution, based on sample of size n=200, 9 unequal width intervals are used to assess goodness-of-fit of data via Chi–Squared test. The observed and expected frequencies for proportions of observation within each class-interval are given in Table 4.

Table 4: Generated Sample, n=200, θ=5

0.036	0.038	0.046	0.052	0.105	0.115	0.140	0.425
0.469	0.471	0.489	0.540	0.559	0.583	0.617	0.632
0.693	0.728	0.728	0.746	0.880	0.905	0.938	0.940
0.949	1.0176	1.024	1.048	1.068	1.122	1.176	1.207
1.224	1.229	1.253	1.324	1.538	1.540	1.551	1.588
1.689	1.714	1.720	1.72	1.7536	1.757	1.766	1.945
1.956	2.017	2.073	2.073	2.107	2.183	2.194	2.208
2.293	2.333	2.370	2.487	2.660	2.703	2.844	2.870
3.105	3.1628	3.337	3.389	3.393	3.401	3.644	3.645
3.690	3.715	3.756	3.813	3.816	3.845	3.988	4.261
4.364	4.367	4.578	4.645	4.646	4.848	5.025	5.127
5.207	5.227	5.347	5.486	5.503	5.612	5.784	5.877
6.126	6.209	6.301	6.388	6.466	6.494	6.867	6.935
7.525	7.722	7.756	7.906	8.415	8.565	8.933	9.449
9.899	10.260	10.438	10.967	11.822	12.6443	12.999	14.093

Table 4: (Continued) Generated Sample, n=200, θ=5

14.11	16.77	16.89	17.41	18.22	19.943	21.52	21.99
23.19	23.38	23.48	23.9	24.17	24.27	24.82	25.52
27.99	28.56	30.42	31.14	31.42	34.50	37.38	37.56
39.39	42.16	43.66	46.66	49.06	52.56	57.4	61.82
70.87	71.06	72.95	74.16	76.19	77.96	88.36	94.17
102.4	106.0	127.5	134.0	137.4	161.04	163.6	178.9
241.3	258.7	330.3	342.1	345.3	388.1	389.4	448.5
487.5	551.8	573.9	645.2	669.1	708.6	738.9	836.4
957.6	1094.7	1234.6	1462.7	1547.6	2035.5	2318.3	3125.4
4054.9	7531.3	7613.1	17133.2	28138.0	31874.1	87341.1	189075.8

Table 5: Chi-Squared goodness-of-test for the generated sample

Interval	n_k	p_k	e_k	$(\frac{n_k}{p_k})(\frac{(p_k-e_k)^2}{e_k})$
[0,4)	79	.395	.379584	.125218
[4,8)	29	.145	.133897	.184136
[8,25)	27	.135	.160005	.781538
[25,100)	25	.125	.125521	.000433
[100,600)	19	.095	.0936368	.003969
[600,2500)	12	.06	.0422084	1.499896
[2500,10000)	4	.02	.0250448	.203236
[10000,100000)	4	.02	.0221899	.043224
[100000, ∞)	1	.005	.0179136	1.861838
Total	200	1.0	1.0	4.703487

In Table 5, p_k denotes observed percentage of observations in the class-interval k, $k = 1, 2, ..., 9$. e_k denotes expected proportion of observations in the class interval k. Note that to get e_k, Bayes estimate, 4.98786 of θ is used in the CDF of the composite distribution. At .05 level of significant with degrees of freedom 7, the critical value for the Chi-Squared test is 14.067, as a result there is no enough evidence to reject the null hypothesis that $\hat{\theta}_{Bayes} = 5$.

Conclusion

A Bayes estimator via Inverse-gamma prior for the boundary parameter θ, that separates large losses and small losses in insurance data is derived based on the Exponential-Pareto composite model. A Bayesian predictive density is derived via the posterior pdf for θ. The "best" value for a is selected through an upper limit (a decreasing function of n) on variance of the Inverse-Gamma prior distribution. Simulation studies indicate that even for large sample size, Bayes estimate outperforms MLE if the "best" values of hyper-parameters a and b are used in computations. Having values of the hyper-parameters, the predictive density is used along with a numerical method in Mathematica to compute VaR at 70% level.

References

Preda, V. and Ciumara, R. "On Composite Models: Weibull-Pareto and Lognormal-Pareto - A comparative Study" Romanian Journal of Economic Forecasting, 2006, 8(4): pp. 32-46.

Cooray, K. and Cheng, C. "Bayesian estimators of the lognormal-Pareto composite distribution" Scandinavian Actuarial Journal, 2015, Issue 6 pp. 500-515.

Cooray, K. and Ananda, M.A. "Modeling actuarial data with a composite lognormal-Pareto model" Scandinavian Actuarial Journal, 2005, Issue 5 pp. 321-334.

Klugman, S.A., Panjer, H. & Willmot, G.E. Loss models from data to decisions, 3rd edition, New York; John Wiley. 2012.

Scollnik, D. and Sun, C. "Modeling with Weibull-Pareto Models" North American Actuarial Journal, 2012, 16:2, pp. 260-272.

Teodorescu, S. and Vernic, R. "A composite Exponential-Pareto distribution" The Annals of the "Ovidius" University of Constanta, Mathematics Series, 2006, Vol. XIV, 1: pp. 99-108.

Appendix A

Mathematica Code 1 below computes MLE of θ. Mathematica Code 2 computes $\hat{\theta}_{Bayes}$ and Var using the "best" value of a provided by a user. The "best" values for a (based on n and a good guess for θ (MLE in particular when n is large)) can be obtained by using the Mathematica Code 3. Code 3 uses simulation to find " the "best"value of a.

Mathematica Code 1 finds MLE for a given sample:

```
data={};
n = ;
Array[data, n];
Array[Rankdata, n];
Array[ML, n];
Array[U2, n];
Rankdata = Sort[data];
While[j < n + 1,
ML[j] = (1.35 Sum[Rankdata[[d]],{d, 1, j}])/(1.35 j - .35 n);
j++; ];
U2 = Table[ML[j], {j, 1, n}];
U2 = Sort[U2];
w1 = 1;
While[w1 < n,
If[Rankdata[[w1]] ≤ U2[[w1]] ≤ Rankdata[[w1 + 1]],
f2 = U2[[w1]]];
w1++; ];
MLE= f2;
Print[MLE];
```

Mathematica Code 2 computes Bayes estimate and Var based on the data set used in Code 1:

```
data=;
n = ;
a = " user inputs "best" value . Use Mathematica Code 3";
g[β] := β^2/((a−1)^2(a−2));
sol = N[Solve g[β] == 1/n^(1/3)];
b = β/. sol[[2]];
Array[data, n];
Array[Rankdata, n];
Array[BE, n];
Array[U1, n];
Rankdata = Sort[data];
j = 1;
While[j < n + 1,
B = (b+ 1.35*Sum[Rankdata[[d]], {d, 1, j}]);
A = (a − .35n + 1.35j);
BE[j] = B/(A − 1);
ML[j] = (1.35 Sum[Rankdata[[d]],{d, 1, j}])/(1.35 j - .35 n);
j++; ];
U1 = Table[BE[j], {j, 1, n}];
U1 = Sort[U1];
s = 1;
While[s < n,
If[Rankdata[[s]] ≤ U1[[s]] ≤ Rankdata[[s + 1]], {m = s, f1 = U1[[s]]}];
s++;
];
Bayes = f1;
Ap = (a − .35n + 1.35 ∗ m);
```

$Bp = (b + 1.35$*Sum[Rankdata[[v]],{v,1,m}]);

k2=((.775 $* Ap(Bp)^{Ap})/(Bp + 1.35y)^{(Ap+1)}$)*

(1-CDF[InverseGammaDistribution[$Ap + 1, Bp + 1.35y], y$]);

k1=($1/y^{1.35}) * (.20 * Bp^{(.35)}$*Gamma[$Ap - .35$])*

(CDF[InverseGammaDistribution[$Ap - .35, Bp], y$])/(Gamma[Ap]);

f[r-?NumberQ]:=NIntegrate[k1+k2,{y,0,r}];

Sol=r/.NSolve[f[r]==.70,r]; **It take a few minutes to find r.**

VaR=Sol[[1]];

Print[VaR]

Mathematica Code 3 can be used to search for "best" value of a. The user needs to provide a "good" guess (Ex: MLE based on a large sample) for θ.

Start an initial choice for a with at least .35 n-1.35 +1, after each run increase a until MSE of Bayes estimate starts to increase. Simulations used in the article indicated that for given n and θ, MSE has a minimum value attained at "best" value of a. Once we use the "best" value of a in Mathematica Code 2, we get an accurate Bayes estimate.

θ=" a good guess such as MLE based on large sample";

n = ;

a ="starts with at least .35 n-1.35 +1 ;

g[β] := $\beta^2/((a-1)^2(a-2))$;

sol = N[Solve[g[β] == $1/n^(1/3), \beta$];

b = β /. sol[[2]];

k = 300;

Array[Bayes, k];

Array[data, n];

Array[Rankdata, n];

Array[BE, n];

dist = ProbabilityDistribution[Piecewise[{{(0.775/θ)*

```
Exp[−1.35x/θ], {0 < x ≤ θ}, {.2 ∗ (θ^.35)/x^1.35, x > θ}}], {x, 0, ∞}];
data = RandomVariate[dist, n];
data = Flatten[data];
Rankdata = Sort[data];
i = 1;
While[i < k + 1,
j = 1;
While[j <n + 1,
B = (b + 1.35*Sum[Rankdata[[d]], d, 1, j]);
A = (a − .35 ∗ n + 1.35 ∗ j);
BE[j] = B/(A − 1);
j++; ];
U1 = Table[BE[j], j, 1, n];
U1 = Sort[U1];
s = 1;
While[s < n,
If[Rankdata[[s]] ≤ U1[[s]] ≤ Rankdata[[s + 1]], m = s, f1 = U1[[s]]];
s++; ];
Bayes[i] = f1;
i++;
];
EstB = Table[Bayes[i], i, 1, k];
g1 = Mean[EstB];
g2 = Mean[(EstB -θ^2];
Print[g1,g2]
```