Bayesian Predictive Modeling for Exponential-Pareto Composite Distribution

by M. S. Aminzadeh and Min Deng

ABSTRACT

Composite distributions have well-known applications in the insurance industry. In this paper, 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 an inverse gamma prior distribution for the parameter θ, and the density is used to estimate the value at risk (VaR). Goodness of fit of the composite model is verified for a generated data set. The accuracy of the Bayes and VaR estimates is assessed via simulation studies. The "best" value for hyperparameters of the inverse gamma prior distribution are found via an upper bound on the variance of the prior distribution. Simulation studies indicate that when the "best" values of hyperparameters are used in the Bayes estimator, the estimator is consistently more accurate than maximum likelihood estimation.

KEYWORDS

Maximum likelihood estimate, inverse gamma, Bayes estimate, exponential-Pareto composite density, predictive density, value at risk

1. Introduction

Two important operations in actuarial science—using data to construct a model and estimating parameters of interest based on the model—are used in the insurance industry. In this paper, both processes are presented in the context of a composite distribution.

Because insurance data is skewed via a fat tail, it is very difficult to find a classical parametric distribution to model 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, many researchers have developed composite models for insurance data or data with similar characteristics. Klugman, Panjer, and Willmot (2012) discussed how to use data to build a model in the actuarial science field and the insurance industry, including many important concepts, such as value at risk (VaR). VaR, one of the most important risk measures in the business world, is the percentile of the distribution of losses. It lets actuaries and risk managers look at "the chance of an adverse outcome" and helps them to make a decision. This paper uses 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. They derived a maximum likelihood estimator for the unknown parameter θ, which represents the boundary between small and large losses in a data set. In a subsequent paper, the same authors worked on different types of exponential-Pareto composite models (Teodorescu and Vernic 2009). Both models have one unknown parameter. Researchers have proposed many other composite distributions. Preda and Ciumara (2006), for example, introduced the Weibull-Pareto and lognormal-Pareto composite models, pointing out that they could be used to model actuarial data collected from the insurance industry. The authors compared the models using different parameters, developed algorithms to find the maximum likelihood estimates for two unknown parameters, and compared the two models' accuracy.

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 small losses with high frequencies. That is why many other distributions, such as exponential, lognormal, and Weibull, have been combined with the Pareto distribution to model losses with small values in a data set. Cooray and Ananda (2005) discuss modeling actuarial data based on a composite lognormal-Pareto composite model. Cooray and Cheng (2015) provide Bayes estimates for parameters of a lognormal-Pareto composite model. Scollnik and Sun (2012) discuss several Weibull-Pareto composite models.

The aim of this paper is to develop a Bayesian predictive density based on the composite exponential-Pareto distribution. In the Bayesian framework, a predictive density is developed via the composite density and then, based on a random sample, a Bayes estimate and the VaR are estimated. The following section develops a posterior probability density function (pdf) via an inverse gamma prior for θ. It also explains how to use a search method to compute the Bayes estimate for θ based on a sample. Section 3 derives the predictive density for Y, where its realization, y, is considered as a future observation from the composite distribution. The first moment, $E[Y|x]$, of the predictive pdf is undefined because $E[X]$ is also undefined for the composite pdf. Section 4 investigates the accuracy of VaR and shows, through a summary of simulation studies, that Bayes estimation of θ under a squared error loss function is consistently more accurate than maximum likelihood estimation (MLE). The section also gives a method for choosing the "best" values for hyperparameters of the prior distribution to get an accurate Bayes estimate. Section 5 contains a numerical example for computational purposes. Three Mathematica codes are given in the Appendix, one for computing the MLE of θ, the second for computing a Bayes estimate and VaR using a single sample, and the third for a simulation that can be used to search for the "best" values of hyperparameters 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 pdf

$$f_x(x) = \begin{cases} cf_1(x) & 0 < x \le \theta \\ cf_2(x) & \theta \le x \le \infty \end{cases},$$

where

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

and

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

$f_1(x)$ is the pdf of an exponential distribution with parameter λ, and $f_2(x)$ is the pdf of a 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

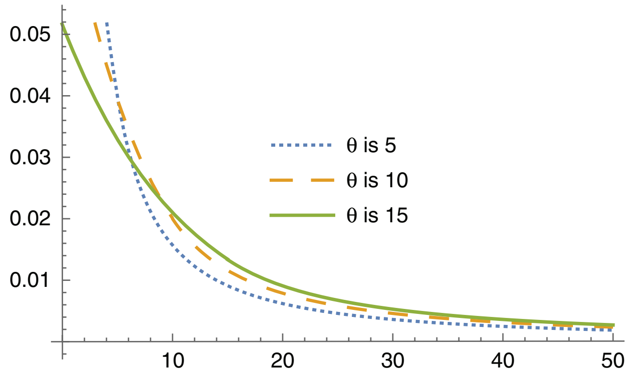
$$\lambda\theta = 1.35, \alpha = 0.35, c = 0.574.$$

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

$$f_x(x|\theta) = \begin{cases} \dfrac{.775}{\theta} e^{\frac{-1.35x}{\theta}} & 0 < x \le \theta \\ \dfrac{.2\theta^{.35}}{x^{1.35}} & \theta \le x < \infty \end{cases}. \tag{1.1}$$

Figure 1.1 provides graphs of the composite pdf for different values of θ, revealing that as θ increases, the tail of the pdf becomes heavier. This implies that the percentile at a specific level, say 0.99, is increasing in θ. Teodorescu and Vernic (2006) compared the exponential distribution with the composite exponential-Pareto distribution and derived an MLE for θ via an ad hoc procedure using a search method. They concluded that when $\theta = 10$, the composite distribution fades to 0 more slowly than the exponential distribution. The implication of this result is that the composite distribution could be a better choice to model insurance data containing large losses. In this situation, we could be able to avoid charging insufficient premiums to cover potential future losses.

Figure 1.1. Composite pdf for different values of θ



2. Derivation of posterior density and Bayes estimator

Let $x_1, \ldots, x_n$ be a random sample for the composite pdf in (1.1) and, without loss of generality, assume that $x_1 < x_2 < \ldots < x_n$ is an ordered random sample from the pdf in (1.1). The likelihood function, also given in Teodorescu and Vernic (2006), is written as

$$L(x|\theta) = k\theta^{0.35n-1.35m} e^{-1.35\sum_{i=1}^{n} x_1/\theta}, \tag{2.1}$$

where $k = \dfrac{0.2^{n-m}(0.775)^m}{\prod_{i=m+1}^{n} x_i^{1.35}}$. To formulate the likelihood function, it is assumed that there is an $m (m = 1, 2, \ldots, n-1)$ such that in the ordered sample $x_m \le \theta \le x_{m+1}$.

To derive a posterior distribution for θ, we use a conjugate prior inverse gamma distribution for θ with the pdf

$$\rho(\theta) = \frac{b^a \theta^{-a-1} e^{-b/\theta}}{\Gamma(a)}, b > 0, a > 0. \tag{2.2}$$

From (2.1) and (2.2), the posterior pdf can be written as

$$f(\theta|x) = L(\theta|x) * \rho(\theta) \propto$$

$$e^{-\frac{b+1.35\sum_{i=1}^{m} x_i}{\theta}} \theta^{-(a-0.35n+1.35m)-1}. \tag{2.3}$$

It can be seen from (2.3) that the expression on the right-hand side is the kernel of inverse gamma (A, B), where $A = (a - 0.35n + 1.35m)$ and $B = (b + 1.35\sum_{i=1}^{m} x_i)$. Therefore, under a squared error loss function, the Bayes estimator for θ is

$$\hat{\theta}_{Bayes} = E[\theta|x] = \frac{B}{A-1} = \frac{b + 1.35\sum_{i=1}^{m} x_i}{a - 0.35n + 1.35m - 1}.$$

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

1. Start with $j = 1$, and check to see if $x_1 \le \frac{B}{A-1} \le x_2$; if yes, then $m = 1$. Otherwise, go to step 2.
2. For $j = 2$, check to see if $x_2 \le \frac{B}{A-1} \le x_3$; if yes, then $m = 2$. Otherwise, let $j = 3$ and continue until the correct value for m has been found.

The idea is to find the correct value for j so that $x_j \le \frac{B}{A-1} \le x_{j+1}$. 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 x, we are interested in deriving the predictive density $f(y|x)$. In the Bayesian framework, predictive density is used to estimate measures such as $E[Y|x]$ or Var$[Y|x]$, or other measures, such as VaR, which is considered in this paper.

$$f(y|x) = \int_0^\infty f(\theta|x) f_Y(y|\theta)\, d\theta,$$

where

$$f_Y(y|\theta) = \begin{cases} \dfrac{0.775}{\theta} e^{\frac{-1.35y}{\theta}} & 0 < y \le \theta \\ \dfrac{0.2\theta^{.35}}{y^{1.35}} & \theta \le y < \infty \end{cases}.$$

As a result, we get

$$f(\theta|x) f(y|\theta)$$
$$= \begin{cases} \dfrac{0.775}{\Gamma(A)\theta} e^{\frac{-1.35y}{\theta}} \theta^{-(A+1)} B^A e^{-B/\theta} & y < \theta < \infty \\ \dfrac{0.2\theta^{.35}}{\Gamma(A) y^{1.35}} \theta^{-(A+1)} B^A e^{-B/\theta} & 0 < \theta < y \end{cases},$$

which reduces to

$$f(\theta|x) f(y|\theta)$$
$$= \begin{cases} \dfrac{0.775 B^A \Gamma(A+1)}{\Gamma(A)(B+1.35y)^{A+1}} h_1(\theta|A+1, B+1.35y) \\ \qquad\qquad\qquad y < \theta < \infty \\ \dfrac{0.2 B^A \Gamma(A-.35)}{\Gamma(A) y^{1.35} B^{A-0.35}} h_2(\theta|A-0.35, B) \\ \qquad\qquad\qquad 0 < \theta < y \end{cases},$$

where $h_1(\theta|A + 1, B + 1.35y)$ is the pdf of the inverse gamma with parameters $A + 1$ and $B + 1.35y$. Also, $h_2(\theta|A - 0.35, B)$ is the pdf of the inverse gamma distribution with parameters $A - .35$ and B. Using the above results, the predictive density, $f(y|x)$, is given by

$$f(y|x) = \int_0^y f(\theta|x) f_Y(y|\theta)\, d\theta + \int_y^\infty f(\theta|x) f_Y(y|\theta)\, d\theta$$
$$= K_2(y) H_2(y|A - 0.35, B)$$
$$+ K_1(y)(1 - H_1(y|A+1, B+1.35y))$$

$$K_1(y) = \frac{0.775A * B^A}{(B + 1.35y)^{A+1}},$$

$$K_2(y) = \frac{0.2B^{.35}\Gamma(A - 0.35)}{\Gamma(A)\, y^{1.35}}. \tag{3.1}$$

H_2 is the cumulative density function (cdf) of the inverse gamma distribution with parameters $(A - 0.35, B)$, and H_1 is the cdf of the inverse gamma distribution with parameters $(A + 1, B + 1.35y)$. Similar to the composite density, for which $E[X]$ is undefined, $E[Y|x]$ is also undefined for the predictive pdf.

4. Simulation

This section describes simulation studies conducted to assess the accuracy of $\hat{\theta}_{Bayes}$ as well as VaR. For selected values of n, θ and the "best" values of hyperparameters (a, b), $N = 300$ samples from the composite density (1.1) are generated.

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

For the inverse gamma prior distribution, we have

$$E[\theta] = \frac{b}{a-1}, \ \mathrm{Var}(\theta) = \frac{b^2}{(a-1)^2(a-2)}. \tag{4.1}$$

Also, note that $A = (a - 0.35n + 1.35m)$ because one of the parameters of the posterior distribution must be positive, and that m takes values 1, 2, . . . , $(n - 1)$. Therefore, $a > 0.35n - 1.35m$, and as a result, the value for a should be at least $0.35n - 1.35$ 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 estimates. However, as a increases, for a desired level of variance in (4.1), b would need to increase. And this causes the expected value in (4.1) to increase. Simulation results, as shown in Table 4.1 for θ = 5,10, indicate that the accuracy of the 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 a large θ produces larger sample points; as a result both the Bayes estimate of θ and its mean squared error (MSE) increase. To overcome this, we propose an upper bound on Var(θ) and let the upper bound be a decreasing function of n. Let

$$\mathrm{Var}(\theta) = \frac{b^2}{(a-1)^2(a-2)} \le \frac{1}{n^{1/3}}. \tag{4.2}$$

The idea is to have a decreasing function of n on the right-hand side of (4.2). Other functions for an upper bound that are decreasing in n may also work, but simulations show that the above choice works

Table 4.1. MSE of Bayes estimate as a function of α and θ

n	θ	a	$\xi(\hat{\theta}_{Bayes})$	n	θ	a	$\xi(\hat{\theta}_{Bayes})$
20	5	17	3.590	20	10	17	43.370
20	5	30	1.800	20	10	30	38.010
20	5	50	0.264	20	10	50	29.080
20	5	70	0.063	20	10	70	21.910
20	5	80	0.331	20	10	80	18.980
—	—	—	—	20	10	250	0.106
—	—	—	—	20	10	260	0.020
—	—	—	—	20	10	280	0.051
50	5	50	0.491	50	10	50	28.670
50	5	70	0.037	50	10	70	24.460
50	5	80	0.019	50	10	80	20.17
50	5	90	0.120	50	10	70	24.46
—	—	—	—	50	10	340	0.043
—	—	—	—	50	10	350	0.006
—	—	—	—	50	10	370	0.043
100	5	60	0.095	100	10	350	0.876
100	5	70	0.039	100	10	410	0.072
100	5	80	0.033	100	10	420	0.027
100	5	90	0.077	100	10	450	0.032

very well. For a selected value of a, (4.2) is used to find b by solving

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

As shown in Table 4.1, for selected values of n and θ, the MSE of the Bayes estimate has a minimum when a takes its "best" value. For selected values of n and θ, Table 4.1 lists the turning points of the MSE. The Mathematica code for the simulation given in the Appendix can be used to make a list of "best" values of a as a function of n and θ. In this paper, $n = 20, 50, 100$, and $\theta = 5, 10$ are considered.

Note that since, in practice, the true value of θ is unknown, it can be guessed by using its MLE. In particular, the MLE would be a good guess if n is large. Note that in simulation studies, in each iteration, the MLE was not used to find a and b because the variability in the MLE would create more variability in the Bayes estimate. For example, in simulation studies for each generated sample and a selected value of a, we tried to find b via $b = (a-1)\hat{\theta}_{MLE}$, due to the $E[\theta]$ formula in (4.1), but under this method, the variability of $\hat{\theta}_{Bayes}$ becomes extremely large and the results are not desirable.

As mentioned earlier, VaR at the 0.90 level is a useful measure for big losses in the insurance industry. We used 0.70 to compute the Bayes estimate of VaR. Even with 0.70 and $N = 300$ interations, it takes several hours for the program to run. Computation of VaR via (3.1) cannot be done analytically, and therefore a numerical method is required. The idea is to find the value of y in (3.1) so that $\int_0^y f(y|x)dy = 0.70$. Mathematica is used to find an estimated value of y based on selected input parameters and a generated sample from the composite density (1.1). As mentioned above, computation of VaR at the 0.90 level is possible, but an extended computer memory is required (in particular when a is large) to solve equations in the code numerically.

Teodorescu and Vernic (2006) derived the MLE of θ. The MLE should satisfy $x_m \le \dfrac{1.35\sum_{i=1}^{m} x_i}{1.35m - 0.35n} \le x_{m+1}$, and it can be found via a search method for $m = 1, 2, \ldots, (n-1)$. Simulation studies indicate that among many generated samples ($k = 300$), only a few (1 or 2 out of 300) lead to extremely large MLE values, and as result, the MSE of the MLE is inflated. Even when such outliers do not occur, the Bayes estimator still outperforms MLE.

Table 4.2. Accuracy of Bayes estimator, MLE, and VaR, $\theta = 5$

n	a	$\overline{\hat{\theta}}_{Bayes}$	$\xi(\hat{\theta}_{Bayes})$	$\overline{VaR}$	Std(VaR)	$\overline{\hat{\theta}}_{MLE}$	$\xi(\hat{\theta}_{MLE})$
20	70	5.24	0.06	32.65	1.01	8.03	25.82
50	80	5.15	0.02	30.22	1.21	7.17	18.45
100	80	5.07	0.04	27.99	1.36	6.20	3.28

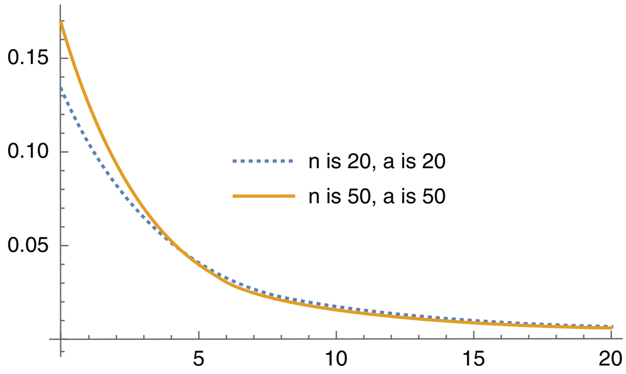
Tables 4.2 and 4.3 summarize simulation studies carried out to assess the accuracy of the Bayes estimator, MLE, and VaR. Both tables use "best" values of a from Table 4.1 to compare the Bayes estimate with the MLE. Tables 4.2 and 4.3 reveal that both $\hat{\theta}_{MLE}$ and $\hat{\theta}_{Bayes}$ are more accurate for larger values of n than for smaller ones. The tables also reveal that $\overline{\hat{\theta}}_{Bayes}$, the average Bayes estimate, is closer to the actual value of θ than is $\overline{\hat{\theta}}_{MLE}$, the average of MLEs. For all values of n in Tables 4.2 and 4.3, we note that MSE(Bayes) = $\xi(\hat{\theta}_{Bayes})$ is considerably smaller than MSE(MLE) = $\xi(\hat{\theta}_{MLE})$.

Figures 4.1 through 4.3 provide graphs of the predictive density for different values of θ and a, revealing that as n increases, the tail of the predictive density becomes heavier, and as a result, at a specific level, say 0.99, VaR increases. Simulation studies confirm this conclusion. Figure 4.3 is a graph of the predictive density for fixed values $n = 50$, $a = 50$, and for three different values of θ based on which samples from the composite pdf are generated. Similar to what is shown in Figure 1.1, as θ increases, the tail of the predictive density becomes heavier, causing VaR at a specific level, say 0.99, to increase.

Table 4.3. Accuracy of Bayes estimator, MLE, and VaR, $\theta = 10$

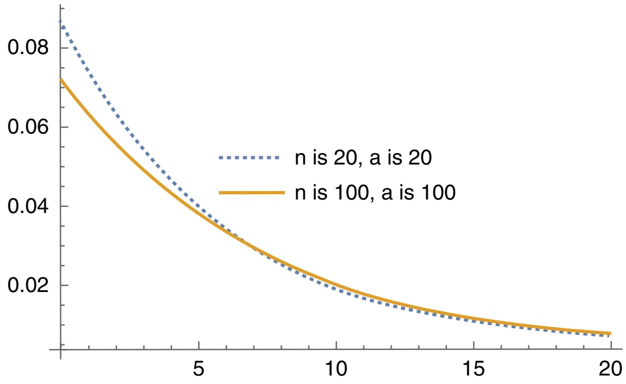
n	a	$\overline{\hat{\theta}}_{Bayes}$	$\xi(\hat{\theta}_{Bayes})$	$\overline{VaR}$	Std(VaR)	$\overline{\hat{\theta}}_{MLE}$	$\xi(\hat{\theta}_{MLE})$
20	260	9.89	0.08	63.36	0.50	13.71	71.72
50	350	9.92	0.07	62.43	0.54	12.30	14.70
100	420	9.94	0.02	61.93	0.58	12.45	12.33

Figure 4.1. Predictive density for selected values of *n* and *a*, θ = 5



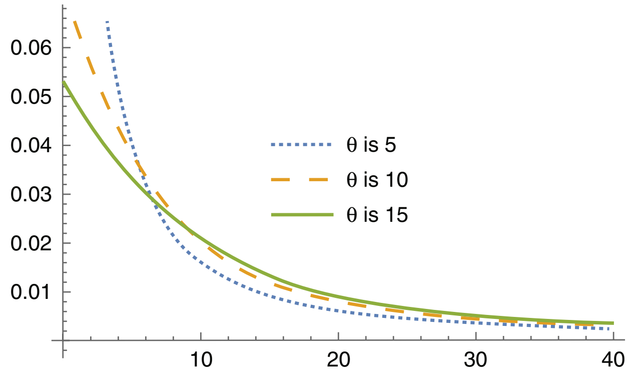
Graphs are based on generated samples from composite pdf.

Figure 4.2. Predictive density for selected values of *n* and *a*, θ = 10



Graphs are based on generated samples from composite pdf.

Figure 4.3. Predictive density with various θ values, *n* = 50 and *a* = 50



Graphs are based on generated samples from composite pdf.

5. Numerical example

The data set in Table 5.1 is a random sample of size 200 generated from the exponential-Pareto composite pdf (1.1) with parameter $\theta = 5$. For this data set, the value of m is 86. The Bayes estimate is 4.9878, and the MLE is 4.935488. Simulation studies indicate that the Bayes estimator is more accurate than the MLE. Due to the long-tailed composite distribution, based on a sample of size $n = 200$, nine intervals of unequal widths are used to assess the goodness of fit of the data via a chi-square test. The observed and expected frequencies for the proportion of observations within each class-interval are given in Table 5.1.

In Table 5.2, p_k denotes the observed percentage of observations in the class-interval k, $k = 1, 2, \dots, 9$. e_k denotes the expected proportion of observations in the class-interval k. Note that to get e_k, a Bayes estimate of 4.98786 for θ is used in the cdf of the composite distribution. At the 0.05 level of significance with 7 degrees of freedom, the critical value for the chi-square test is 14.067. As a result, there is not enough evidence to reject the null hypothesis that $\hat{\theta}_{Bayes} = 5$.

6. Conclusion

A Bayes estimator via inverse gamma prior for the boundary parameter θ, which separates large losses from 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 the variance of the inverse gamma prior distribution. Simulation studies indicate that even for a large sample size, the Bayes estimate outperforms MLE if the "best" values of hyperparameters a and b are used in computations. Having values for the hyperparameters, the predictive density is used along with a numerical method in Mathematica to compute VaR at the 70% level.

Table 5.1. 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	1.720	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
8.899	10.260	10.438	10.967	11.822	12.6443	12.999	14.093
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	1,094.7	1,234.6	1,462.7	1,547.6	2,035.5	2,318.3	3,125.4
4,054.9	7,531.3	7,613.1	17,133.2	28,138.0	31,874.1	87,341.1	189,075.8

Table 5.2. Chi-square goodness-of-fit test for the generated sample

Interval	n_k	p_k	e_k	$\left(\frac{n_k}{p_k}\right)\left(\frac{(p_k - e_k)^2}{e_k}\right)$
[0,4)	79	0.395	0.3795840	0.125218
[4,8)	29	0.145	0.1338970	0.184136
[8,25)	27	0.135	0.1600050	0.781538
[25,100)	25	0.125	0.1255210	0.000433
[100,600)	19	0.095	0.0936368	0.003969
[600,2500)	12	0.060	0.0422084	1.499896
[2500,10000)	4	0.020	0.0250448	0.203236
[10000,100000)	4	0.020	0.0221899	0.043224
[100000, ∞)	1	0.005	0.0179136	1.861838
Total	200	1	1	4.703487

References

Cooray, K., and M. A. Ananda. "Modeling Actuarial Data with a Composite Lognormal-Pareto Model," *Scandinavian Actuarial Journal* 5, 2005, pp. 321–334.

Cooray, K., and C. Cheng, "Bayesian Estimators of the Lognormal-Pareto Composite Distribution," *Scandinavian Actuarial Journal* 6, 2015, pp. 500–515.

Klugman, S. A., H. Panjer, and G. E. Willmot, *Loss Models from Data to Decisions* (3rd ed.), New York: John Wiley, 2012.

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

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

Teodorescu, S., and R. Vernic, "A Composite Exponential-Pareto Distribution," *Annals of the Ovidius University of Constanta, Mathematics Series* 14:1, 2006, pp. 99–108.

Teodorescu, S., and R. Vernic, "Some Composite Exponential-Pareto Models for Actuarial Prediction," *Romanian Journal of Economic Forecasting* 12(4), 2009, pp. 82–100,

Appendix

Mathematica code 1 computes the 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 Mathematica code 3. Code 3 uses a 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";
```

$$g[\beta]: = \frac{\beta^2}{(a-1)^2\,(a-2)};$$

$$\text{sol} = N\left[\text{Solve } g[\beta] == \frac{1}{n^{1/3}}\right];$$

```
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 takes a few minutes
   to find r.
VaR=Sol[[1]];
Print[VaR]
```

Mathematica code 3 can be used to search for the "best" value of *a*. The user needs to provide a "good" guess (e.g., the MLE based on a large sample) for θ.

Start with an initial choice for *a* with at least $0.35n - 1.35 + 1$. After each run, increase *a* until the MSE of the Bayes estimate starts to increase. Simulations used in the paper indicate that for given *n* and θ, the MSE has a minimum value that is attained at the "best" value of *a*. Once we use the "best" value of *a* in Mathematica code 2, we get an accurate Bayes estimate.

Mathematica code 3: Finds the "best" value of *a*

```
θ = "a good guess such as MLE based on large
   sample";
n = ;
a = "starts with at least .35 n − 1.35 + 1";
g[β]: = β^2/((a − 1)^2(a – 2));
sol = N[Solve[g[β] == 1/n^(1/3), β];
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]
```