## Article

# Bayesian Inference for the Loss Models via Mixture Priors 

Min Deng (D) and Mostafa S. Aminzadeh *<br>Department of Mathematics, Towson University, Towson, MD 21252, USA; mdeng@towson.edu<br>* Correspondence: maminzadeh@towson.edu

Citation: Deng, Min, and Mostafa S. Aminzadeh. 2023. Bayesian Inference for the Loss Models via Mixture Priors. Risks 11: 156. https:// doi.org/10.3390/risks11090156

Academic Editor: Mogens Steffensen
Received: 5 June 2023
Revised: 16 August 2023
Accepted: 18 August 2023
Published: 31 August 2023


Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ $4.0 /$ ).


#### Abstract

Constructing an accurate model for insurance losses is a challenging task. Researchers have developed various methods to model insurance losses, such as composite models. Composite models combine two distributions: one for part of the data with small and high frequencies and the other for large values with low frequencies. The purpose of this article is to consider a mixture of prior distributions for exponential-Pareto and inverse-gamma-Pareto composite models. The general formulas for the posterior distribution and the Bayes estimator of the support parameter $\theta$ are derived. It is shown that the posterior distribution is a mixture of individual posterior distributions. Analytic results and Bayesian inference based on the proposed mixture prior distribution approach are provided. Simulation studies reveal that the Bayes estimator with a mixture distribution outperforms the Bayes estimator without a mixture distribution and the ML estimator regarding their accuracies. Based on the proposed method, the insurance losses from natural events, such as floods from 2000 to 2019 in the USA, are considered. As a measure of goodness-of-fit, the Bayes factor is used to choose the best-fitted model.


Keywords: Bayesian estimation; composite model; mixture prior distribution; mixture posterior distribution; Bayes factor; marginal likelihood

## 1. Introduction

Constructing an accurate loss model for insurance loss is one of the essential topics in actuarial science. Insurance industry data have unique properties. They have a high frequency for small losses and very few significant losses. A traditional distribution, such as normal and others, cannot describe insurance data skewness and fat-tailed properties. Therefore, many researchers explored the other distributions to fit the insurance loss data better. The class of composite distributions is one of them. A composite distribution combines a typical distribution, such as exponential, inverse-gamma, Weibull, and lognormal, for the data with slight losses and the Pareto distribution with extreme losses but low frequencies.

Klugman et al. (2012) provided a detailed discussion on modeling datasets in actuarial science. Teodorescu and Vernic (2006) considered the exponential-Pareto composite model and derive the maximum likelihood estimator for the support parameter $\theta$. Preda and Ciumara (2006) employed the composite models Weibull-Pareto and log-normal-Pareto to model insurance losses. These models have two parameters: one is the support parameter $\theta$ and another is the shape parameter $\alpha$. In the article, they developed algorithms to find and compare the maximum likelihood estimates for two unknown parameters. Cooray and Cheng (2013) estimated the parameters of the log-normal-Pareto composite distribution by using Bayesian methods with both Jeffreys and conjugate priors. They used MCMC methods rather than developing closed mathematical formulas. Scollnik and Sun (2012) developed several composite Weibull-Pareto models and suggested using them in different situations. Aminzadeh and Deng (2017) reconsidered the composite exponential-Pareto distribution and provided the Bayesian estimate of the $\theta$ via inverse-gamma as the prior distribution. Aminzadeh and Deng (2019) developed an inverse-gamma-Pareto composite distribution to model insurance losses and provided Bayesian inference based on gamma
prior distribution. Deng and Aminzadeh (2019) revisited the Weibull-Pareto composite model and derived the Bayesian inference for the model. In Deng and Aminzadeh (2019), both inverse-gamma (IG) and gamma priors were employed to find Bayes estimates of the support parameter of $\theta$ and the shape parameter $\alpha$. They also confirmed via simulation studies that the Bayes estimates of the parameters consistently outperform MLEs in all cases. Bakar et al. (2015) develop several new composite models on the Weibull distribution for heavy-tailed insurance loss data. These models are fitted to two real insurance loss data and their goodness-of-fit is tested.

Mixture distributions have applications in many fields, including insurance, actuarial science, and risk management. Klugman et al. (2012) discussed why mixture distributions have broad applications in the actuarial science field. Miljkovic and Grün (2016) used the mixture distributions to model insurance losses. They compared the mixture model with composite models for Danish Fire data and pointed out that it is better than composite models. Bhati et al. (2019) used a mixture of the Pareto and log-gamma distributions to model the heavy-tailed losses.

Abdul Majid and Ibrahim (2021a) analyzed composite Pareto models for Malaysian household income data. The parameter estimation uses numerical methods based on maximum pseudo-likelihood. The conclusion is that the log-normal-Pareto (II) model provides the best fit compared to other models. Abdul Majid and Ibrahim (2021b) proposed a Bayesian approach to composite Pareto models that involves prior distribution on the proportion of data coming from the Pareto distribution instead of assuming the prior distribution on the threshold $\theta$. They concluded that a uniform prior on the proportion approach is less biased than the point estimates determined when using a uniform prior on the threshold. Deng et al. (2021) provided an analytical Bayesian approach to derive estimators of the log-normal-Pareto composite distribution parameters based on the selected priors. The article compared exponential-Pareto, inverse-gamma-Pareto, and log-normal-Pareto as candidate models for data on natural hazards from 1900 to 2016 in the USA. The conclusion is that the log-normal-Pareto distribution provides the best fit.

To model large losses, the Pareto distribution is the distribution favored by practitioners and researchers for modeling heavy-tailed financial data. However, when losses consist of smaller values with high frequencies and larger losses with low frequencies, the log-normal or the Weibull distributions are preferred. Nevertheless, no ordinary distribution provides an acceptable fit for both small and large losses. On the one hand, as mentioned by Dominicy and Sinner (2017), the Pareto fits the tail well, but on the other hand, log-normal, Weibull, and inverse-gamma produce an overall good fit but fit the tail badly. The purpose of using composite distributions is to overcome the dilemma. Saleem (2010) considers type-I mixtures of the members of a subclass of the one-parameter exponential family distributions, such as exponential, Rayleigh, Pareto, Burr type XII, and power function distributions for censored data. The article provides Bayes estimators of parameters using ML, as well as uniform and Jeffreys priors. To our knowledge, the mixture of the priors' approach has not been considered in the literature for composite distributions. The proposed method in the current article considers two composite distributions. The mixture prior method is based on gamma and inverse-gamma priors, which are good candidates for the positive threshold parameter $\theta$. Furthermore, we propose a data-driven approach to compute optimal values for hyperparameters. For a real dataset where a selected "true" value for $\theta$ (unlike in simulations) is not available, we propose using the MLE of $\theta$ along with the characteristics of the prior distribution to assign optimal values for the hyperparameter values.

In this article, we apply the Bayesian method to the composite models using a mixture of prior distributions instead of a single prior distribution for $\theta$. The motivation comes from the natural loss of data over many years due to many factors, such as floods, fires, storms, and earthquakes. Each of them should have a distribution with its parameters. Therefore, the mixture distribution describes the overall distribution. The organization of the article is as follows. Section 2 discusses the general mixture prior, the general mixture posterior, and
the general predictive distributions with the risk measures. Section 3 provides the formulas for Bayes estimators of $\theta$ via the mixture prior distribution approach for both exponentialPareto and gamma-Pareto composite models. Section 4 summarizes simulation studies based on equality-weighted mixture distributions and compares the accuracy of different methods. Section 5 analyzes the natural disaster loss data to illustrate the computations involved and identifies the best model using the Bayes factor as a goodness-of-fit measure.

## 2. Mixture Distribution

Definition 1. A random variable $Z$ is a $K$-point mixture of the random variables $X_{1}, X_{2}, \ldots, X_{K}$ if its cdf is given by

$$
F_{Z}(z)=k_{1} F_{X_{1}}(z)+k_{2} F_{X_{2}}(z)+\ldots+k_{K} F_{X_{K}}(z)
$$

where $k_{j}>0$ and $\sum_{j=1}^{K} k_{j}=1$.
Therefore, a mixture distribution density is given by

$$
f_{Z}(z)=\sum_{j=1}^{K} k_{j} f_{X_{j}}(z)
$$

The steps to derive the posterior distribution of the random variables with a mixture prior distribution are as follows:

Let $x_{1}, x_{2}, \ldots x_{n}$ be a random sample from the distribution with a parameter $\theta$. The likelihood function $\mathrm{L}(\theta)$ can be written as follows:

$$
L(\theta)=\prod_{i=1}^{n} f\left(x_{i} \mid \theta\right)
$$

Let the prior distribution of the parameter $\theta$ be a $K$-point mixture distribution with the density function given by

$$
\pi(\theta)=\sum_{j=1}^{K} k_{j} \pi_{j}(\theta)
$$

where all $k_{j}>0$, and $\sum_{j=1}^{K} k_{j}=1$, and $\int_{-\infty}^{\infty} \pi_{j}(\theta) d \theta=1$.
Therefore, the joint distribution of $\theta$ and $X$ is

$$
f(\theta, x)=L(\theta) \sum_{j=1}^{K} k_{j} \pi_{j}(\theta)
$$

where $\pi_{j}(\theta), j=1,2,3 \ldots, K$ belong to the same class of distributions. For example, all could be Pareto, gamma, or normal. The marginal distribution of $X$ is given by

$$
f_{X}(x)=\int_{-\infty}^{\infty} f(\theta, x) d \theta
$$

The posterior distribution of $\theta$ is

$$
\begin{equation*}
\pi(\theta \mid x)=\frac{f(\theta, x)}{f_{X}(x)}=\frac{L(\theta) \sum_{j=1}^{K} k_{j} \pi_{j}(\theta)}{\int_{-\infty}^{\infty} f(\theta, x) d \theta} \tag{1}
\end{equation*}
$$

For now, consider only the $j^{t h}$ prior distribution $\pi_{j}(\theta)$ and denote the corresponding joint distribution as $f^{j}(\theta, x)$, then

$$
f^{j}(\theta, x)=\pi_{j}(\theta) L(\theta), \quad j=1,2, \ldots, K .
$$

Let us denote the corresponding marginal distribution of $X$ as $f_{X}^{j}(x)$, then

$$
f_{X}^{j}(x)=\int_{-\infty}^{\infty} f^{j}(\theta, x) d \theta=\int_{-\infty}^{\infty} \pi_{j}(\theta) L(\theta) d \theta
$$

Therefore, the corresponding posterior distribution is

$$
\pi_{X}^{j}(\theta \mid x)=\frac{\pi_{j}(\theta) L(\theta)}{f_{X}^{j}(x)}
$$

which implies

$$
\begin{equation*}
\pi_{j}(\theta) L(\theta)=\pi_{X}^{j}(\theta \mid x) f_{X}^{j}(x) \tag{2}
\end{equation*}
$$

Using (2) and (1), the posterior distribution, based on the K-point mixture prior $\pi_{j}(\theta)$, $(j=1,2, \ldots, K)$, is given by

$$
\begin{align*}
\pi(\theta \mid x) & =\frac{f(\theta, x)}{f_{X}(x)}=\frac{\sum_{j=1}^{K} k_{j} \pi_{j}(\theta) L(\theta)}{f_{X}(x)}=\frac{\sum_{j=1}^{K} k_{j} \pi_{X}^{j}(\theta \mid x) f_{X}^{j}(x)}{\int_{-\infty}^{\infty} f(\theta, x) d \theta} \\
& =\frac{\sum_{j=1}^{K} k_{j} \pi_{X}^{j}(\theta \mid x) f_{X}^{j}(x)}{\sum_{j=1}^{K} \int_{-\infty}^{\infty} k_{j} \pi_{X}^{j}(\theta \mid x) f_{X}^{j}(x) d \theta}=\sum_{j=1}^{K} \beta_{j} \pi^{j}(\theta \mid x) \tag{3}
\end{align*}
$$

where

$$
\beta_{j}=\frac{k_{j} f_{X}^{j}(x)}{\sum_{j=1}^{K} \int_{-\infty}^{\infty} k_{j} \pi_{X}^{j}(\theta \mid x) f_{X}^{j}(x) d \theta}=\frac{k_{j} f_{X}^{j}(x)}{\sum_{j=1}^{K} k_{j} f_{X}^{j}(x)}
$$

because, $\int_{-\infty}^{\infty} \pi_{X}^{j}(\theta \mid x) d \theta=1$. Therefore, $\sum_{j=1}^{K} \beta_{j}=1$. Hence, the form of the posterior pdf in (3) confirms that the posterior distribution based on a mixture prior distribution is also a mixture distribution of the individual posterior distributions.

Now, we consider the predictive distribution of $Y$, given $X$. Let $y$ denote a future realization of the random variable $Y$. We assume that $\theta>0$, which is the case for the composite models.

The predictive density of $y$, given $x$, is formulated as follows:

$$
\begin{gather*}
f(y \mid x)=\frac{f(y, x)}{f_{X}(x)}=\frac{\int_{0}^{\infty} f(y, x, \theta) \pi(\theta) d \theta}{f_{X}(x)}=\frac{\int_{0}^{\infty} f(y \mid x, \theta) f(x \mid \theta) \pi(\theta) d \theta}{f_{X}(x)} \\
=\frac{\int_{0}^{\infty} f(y \mid \theta) f(x, \theta) d \theta}{f_{X}(x)}=\int_{0}^{\infty} f(y \mid \theta) \pi(\theta \mid x) d \theta \tag{4}
\end{gather*}
$$

Using (3) and (4), and noting that $f^{j}(y \mid x)=\int_{0}^{\infty} f(y \mid \theta) \pi^{j}(\theta \mid x) d \theta$, we obtain

$$
\begin{aligned}
f(y \mid x)= & \int_{0}^{\infty} f(y \mid \theta) \sum_{j=1}^{K} \beta_{j} \pi^{j}(\theta \mid x) d \theta \\
& =\sum_{j=1}^{K} \beta_{j} f^{j}(y \mid x) .
\end{aligned}
$$

Recall that we have already shown $\sum_{j=1}^{K} \beta_{j}=1$. Therefore, the predictive distribution of the mixture prior distribution is also the mixture distribution of the individual predictive distributions.

### 2.1. Example: Exponential with a Mixture of Gamma Distributions

Let $X_{1}, X_{2}, \ldots, X_{n}$ be independent identically distributed (iid) random variables from the exponential distribution with parameter $\theta$. The density function is given by

$$
f_{X_{i}}(x \mid \theta)=\theta e^{-\theta x}, x>0, \theta>0, i=1,2, \ldots \ldots n
$$

and the likelihood function is

$$
L(\theta)=\prod_{i=1}^{n} f_{X_{i}}(x \mid \theta)=\theta^{n} e^{-\theta \sum_{i=1}^{n} x_{i}}
$$

Let the prior distribution of $\theta$ be in the class of gamma distributions with parameters $\alpha_{j}>0$ and $\beta_{j}>0, j=1,2, \ldots, K$. Then the mixture prior distribution is

$$
\pi(\theta)=\sum_{j=1}^{K} k_{j} \frac{\theta^{\alpha_{j}-1} e^{-\frac{\theta}{\beta_{j}}}}{\Gamma\left(\alpha_{j}\right) \beta_{j}^{\alpha_{j}}} .
$$

Therefore, the joint distribution is given by

$$
\begin{align*}
f(x, \theta)= & L(\theta) \pi(\theta)=\sum_{j=1}^{K} k_{j} \frac{\theta^{\alpha_{j}-1} e^{-\frac{\theta}{\beta_{j}}}}{\Gamma\left(\alpha_{j}\right) \beta_{j}^{\alpha_{j}}} \theta^{n} e^{-\theta \sum_{i=1}^{n} x_{i}} \\
& =\sum_{j=1}^{K} k_{j} \frac{\theta^{n+\alpha_{j}-1} e^{-\theta\left(\frac{1}{\beta_{j}}+\sum_{i=1}^{n} x_{i}\right)}}{\Gamma\left(\alpha_{j}\right) \beta_{j}^{\alpha_{j}}} \tag{5}
\end{align*}
$$

Hence, the marginal distribution of $X$ is given by

$$
\begin{align*}
& f_{X}(x)=\int_{0}^{\infty} \sum_{j=1}^{K} k_{j} \frac{\theta^{n+\alpha_{j}-1} e^{-\theta\left(\frac{1}{\beta_{j}}+\sum_{i=1}^{n} x_{i}\right)}}{\Gamma\left(\alpha_{j}\right) \beta_{j}^{\alpha_{j}}} d \theta \\
& =\sum_{j=1}^{K} \frac{k_{j}}{\Gamma\left(\alpha_{j}\right) \beta_{j}^{\alpha_{j}}} \int_{0}^{\infty} \theta^{n+\alpha_{j}-1} e^{-\theta\left(\frac{1}{\beta_{j}}+\sum_{i=1}^{n} x_{i}\right)} d \theta  \tag{6}\\
& =\sum_{i=j}^{K} \frac{k_{j} \Gamma\left(n+\alpha_{j}\right)\left(\frac{1}{\frac{1}{\beta_{j}}+\sum_{i=1}^{n} x_{i}}\right)^{n+\alpha_{j}}}{\Gamma\left(\alpha_{j}\right) \beta_{j}^{\alpha_{i}}} \tag{7}
\end{align*}
$$

The integrand in RHS of (6) is the kernel of the gamma distribution with parameters $n+\alpha_{j}$ and $\frac{1}{\frac{1}{\beta_{j}}+\sum_{i=1}^{n} x_{i}}$.

Using (5) and (7), the posterior distribution $\pi(\theta \mid x)$ is given by

$$
\begin{equation*}
\pi(\theta \mid x)=\frac{f(x, \theta)}{f_{X}(x)}=\frac{\sum_{j=1}^{K} k_{j} \frac{\theta^{n+\alpha_{j}-1} e^{-\theta\left(\frac{1}{\beta_{j}}+\sum_{i=1}^{n} x_{i}\right)}}{\Gamma\left(\alpha_{j}\right) \beta_{j}^{\alpha_{j}}}}{\sum_{j=1}^{K} \frac{k_{j} \Gamma\left(n+\alpha_{j}\right)\left(\frac{1}{\frac{1}{\beta_{j}}+\sum_{i=1}^{n} x_{i}}\right)^{n+\alpha_{j}}}{\Gamma\left(\alpha_{j}\right) \beta_{j}^{\alpha_{j}}}} . \tag{8}
\end{equation*}
$$

And after some algebraic manipulations, (8) reduces to

$$
\begin{gather*}
\pi(\theta \mid x)=\sum_{j=1}^{K} \frac{\frac{k_{j} \Gamma\left(n+\alpha_{j}\right)\left(\frac{1}{\frac{1}{\beta_{j}}+\sum_{i=1}^{n} x_{i}}\right)^{n+\alpha_{j}}}{\Gamma\left(\alpha_{j}\right) \beta_{j}^{\alpha_{i}}}}{\sum_{j=1}^{K} \frac{k_{j} \Gamma\left(n+\alpha_{j}\right)\left(\frac{1}{\frac{1}{\beta_{j}}+\sum_{i=1}^{n} x_{i}}\right)^{n+\alpha_{j}}}{\Gamma\left(\alpha_{j}\right) \beta_{j}^{\alpha_{j}}}} \frac{\theta^{n+\alpha_{j}-1} e^{-\theta\left(\frac{1}{\beta_{j}}+\sum_{i=1}^{n} x_{i}\right)}}{\Gamma\left(n+\alpha_{j}\right)\left(\frac{1}{\frac{1}{\beta_{j}}+\sum_{i=1}^{n} x_{i}}\right)^{n+\alpha_{j}}} \\
=\sum_{j=1}^{K} \frac{k_{j} f_{X}^{j}(x)}{\sum_{j=1}^{K} k_{j} f_{X}^{j}(x)} \pi^{j}(\theta \mid x)=\sum_{j=1}^{K} \beta_{j} \pi^{j}(\theta \mid x) . \tag{9}
\end{gather*}
$$

As expected, the RHS of (9) confirms that the posterior distribution is the mixture distribution of the individual posterior distributions.

Figure 1a provides graphs for five individual gamma distributions with the shape parameter $\alpha$ and the scale parameter $\beta$. Figure 1b provides graphs for equally weighted mixtures for gamma distributions. Figure 1c provides graphs of mixture gamma distributions with equal weights for two gamma distributions with different shaped parameters and equal scale parameters. Figure 1d provides graphs for the mixture of gamma distributions with unequal weights for gamma distributions with the same parameter shape but different scale parameters. Figure 1b-d confirm that the general shape of the pdf for the mixture distribution is significantly different from that of the pdf for individual gamma distributions in Figure 1a.


Figure 1. Comparison of non-mixture gamma with mixture gamma distributions. (a) Gamma distributions with different parameters. (b) Equally weighted mixture gamma distributions. (c) Unequally weighted mixture of two gamma distributions with different shapes and the same scale. (d) Unequally weighted mixture of two gamma distributions with the same shape and different scales.

## 3. Bayesian Approach to Composite Models based on the Mixture Prior Distribution

### 3.1. Bayesian Inference for Composite Exponential-Pareto Based on the Mixture Prior Distribution

 Teodorescu and Vernic (2006) considered the exponential-Pareto composite model.Suppose a random variable $X$ has the pdf defined as a piecewise function,

$$
f_{X}(x)= \begin{cases}c f_{1}(x) & 0<x \leq \theta \\ c f_{2}(x) & \theta \leq x<\infty\end{cases}
$$

where

$$
f_{1}(x)=\lambda e^{-\lambda x}, x>0, \lambda>0,
$$

and

$$
f_{2}(x)=\frac{\alpha \theta^{\alpha}}{x^{\alpha+1}}, \quad x \geq \theta
$$

The pdf of the exponential distribution with parameter $\lambda$ is denoted by $f_{1}(x)$, and the pdf of the Pareto distribution with parameters $\theta$ and $\alpha$ is denoted by $f_{2}(x)$.

Since the pdf of a composite distribution should be a smooth function, the continuity and differentiability conditions on $f_{X}(x)$ at $\theta$ are necessary. Hence,

$$
f_{1}(\theta)=f_{2}(\theta), f_{1}^{\prime}(\theta)=f_{2}^{\prime}(\theta)
$$

As explained in Teodorescu and Vernic (2006), the above equations reduce to

$$
\lambda e^{-\lambda \theta}=\frac{\alpha}{\theta} \quad \lambda^{2} e^{-\lambda \theta}=\frac{\alpha(1+\alpha)}{\theta^{2}}
$$

which lead to

$$
\alpha=\lambda \theta-1 \quad \lambda \theta\left(e^{-\lambda \theta}-1\right)+1=0 .
$$

Numerical methods via Mathematica for the second equation above, lead to

$$
\lambda \theta=1.35 \quad \alpha=0.35
$$

Since $\int_{0}^{\infty} f(x) d x=1$, the normalizing constant $c$ is computed as $c=\frac{1}{2-e^{-\lambda \theta}}=0.574$. Therefore, the initial three parameters reduce to only one parameter $\theta$, and the pdf of the exponential-Pareto distribution is

$$
f_{X}(x \mid \theta)= \begin{cases}\frac{0.775}{\theta} e^{\frac{-1.35 x}{\theta}} & 0<x \leq \theta  \tag{10}\\ \frac{0.2 \theta^{1.35}}{x^{1.35}} & \theta \leq x<\infty\end{cases}
$$

For $0<x \leq \theta$, the cdf is

$$
F_{x}(x \mid \theta)=P(X \leq x)=\int_{0}^{x} \frac{0.775}{\theta} e^{\frac{-1.35 x}{\theta}} d x=\frac{0.775}{1.35}\left(1-e^{\frac{-1.35 x}{\theta}}\right)
$$

When $x>\theta$,

$$
\begin{aligned}
F_{x}(x \mid \theta) & =P(X \leq x)=\int_{0}^{\theta} \frac{0.775}{\theta} e^{\frac{-1.35 x}{\theta}} d x+\int_{\theta}^{x} \frac{0.2 \theta^{0.35}}{x^{1.35}} d x \\
& =\frac{0.775}{1.35}\left(1-e^{\frac{-1.35 x}{\theta}}\right)+\frac{0.2}{0.35}\left(1-\left(\frac{\theta}{x}\right)^{0.35}\right)
\end{aligned}
$$

Therefore, we have the CDF as a piecewise function

$$
F_{X}(x \mid \theta)= \begin{cases}\frac{0.775}{1.35}\left(1-e^{\frac{-1.35 x}{\theta}}\right) & 0<x \leq \theta \\ \frac{0.775}{1.35}\left(1-e^{\frac{-1.35 x}{\theta}}\right)+\frac{0.2}{0.35}\left(1-\left(\frac{\theta}{x}\right)^{0.35}\right) & \theta \leq x<\infty\end{cases}
$$

To find the quantile $x_{p}$ through $F_{x}(x \mid \theta)=P$, for $0<P<1$, we consider two cases. Since $F_{x}(\theta \mid \theta)=\frac{0.775}{1.35}\left(1 e^{\frac{-1.35 \theta}{\theta}}\right)=0.425251$, first consider the case $0<P<0.425251$. Solving
$\frac{0.775}{1.35}\left(1-e^{\frac{-1.35 x_{p}}{\theta}}\right)=P$ for $x_{p}$, gives $x_{p}=\frac{-\theta}{1.35} \ln \left(1-\frac{1.35 P}{0.775}\right)$. Note that for $P=0.25$, the first quartile is $x_{0.25}=0.423545 \theta$.

For the case $0.425251<P<1$, solving

$$
\frac{0.775}{1.35}\left(1-e^{\frac{-1.35 x_{p}}{\theta}}\right)+\frac{0.2}{0.35}\left(1-\left(\frac{\theta}{x_{p}}\right)^{0.35}\right)=P
$$

gives

$$
x_{p}=\frac{\theta}{\left(1-\frac{0.35}{0.2}(P-0.425251)\right)^{\frac{1}{0.35}}}
$$

For the special case $P=0.99$, we have $x_{P}=331,596 \theta$. In light of the above findings, the quantile function for the exponential-Pareto is

$$
x_{p}= \begin{cases}\frac{-\theta}{1.35} \ln \left(1-\frac{1.35 P}{0.775}\right) & 0<P<0.425251 \\ \frac{\left(1-\frac{0.35}{0.2}(P-0.425251)\right)^{0.35}}{0.1} & 0.425251<P<1\end{cases}
$$

For a random sample $x_{1}, \ldots, x_{n}$ from the composite pdf in (10), without loss of generality, assume $x_{1}<x_{2}<\ldots<x_{n}$. The likelihood function can be formulated as

$$
\begin{equation*}
L(\underline{\mathrm{x}} \mid \theta)=c \theta^{0.35 n-1.35 m} e^{-1.35 \sum_{i=1}^{m} x_{i} / \theta} \tag{11}
\end{equation*}
$$

where $c=\frac{0.2^{n-m}(0.775)^{m}}{\prod_{i=m+1}^{n} x_{i}^{1.35}}$. To formulate the likelihood function, we assume that, without loss of generality, there is an $m(m=1,2, \ldots, n-1)$ so that in the ordered sample $x_{m} \leq \theta \leq$ $x_{m+1}$.

The solution to $\frac{\partial \ln (L(\mathbf{X} \mid \theta))}{\partial \theta}=\frac{0.35 n-1.35 m}{\theta}+\frac{1.35 \sum_{i=1}^{m} x_{i}}{\theta^{2}}=0$ is the MLE of $\theta$,

$$
\hat{\theta}_{M L E}=\frac{1.35 \sum_{i=1}^{m} x_{i}}{1.35 m-0.35 n} .
$$

Note that the Fisher information is

$$
I(\theta)=-E\left[\frac{\partial^{2} \ln (L(\underline{\mathrm{x}} \mid \theta))}{\partial^{2} \theta}\right]=\frac{1.35 m-0.35 n}{\theta^{2}}+\frac{2.7 \sum_{i=1}^{m} E\left[X_{i}\right]}{\theta^{3}}
$$

where,

$$
E[X]=\int_{0}^{\theta} x f_{X}(x \mid \theta) d x=0.0064273 \theta
$$

$1 / \sqrt{I(\theta)}$ provides the standard deviation of the MLE.
We can see that the MLE requires the correct value of $m$ for its computation. By the assumption, $x_{m} \leq \theta \leq x_{m+1}$, the algorithm below goes through the following steps to compute $\hat{\theta}_{M L E}$.

1. Get sorted sample observations $x_{1}<x_{2}<\ldots<x_{n}$.
2. Start with $m=1$, compute $\hat{\theta}_{M L E}$, if $x_{1} \leq \hat{\theta}_{M L E} \leq x_{2}$, then $m=1$, otherwise go to step 3.
3. Let $m=2$, compute $\hat{\theta}_{M L E}$, if $x_{2} \leq \hat{\theta}_{M L E} \leq x_{3}$, then $m=2$, otherwise go to next step.

The above process continues until we identify the correct value for $m$. Using the correct value of $m, \hat{\theta}_{M L E}$ can be computed.

Aminzadeh and Deng (2017) developed Bayesian inference for the exponential-Pareto composite model by considering inverse-gamma as the prior distribution for $\theta$,

$$
\begin{equation*}
\pi(\theta)=\frac{b^{a} \theta^{-a-1} e^{-b / \theta}}{\Gamma(a)}, b>0, a>0 . \tag{12}
\end{equation*}
$$

Using (11) and (12), the posterior $\operatorname{pdf} \pi(\theta \mid \underline{\mathrm{x}})$ is

$$
\begin{equation*}
\pi(\theta \mid \underline{\mathrm{x}})=L(\underline{\mathrm{x}} \mid \theta) * \pi(\theta) \propto e^{-\frac{b+1.35 \sum_{i=1}^{m} x_{i}}{\theta}} \theta^{-(a-0.35 n+1.35 m)-1} . \tag{13}
\end{equation*}
$$

Using the squared-error loss function, the Bayes estimator for $\theta$ is

$$
\begin{equation*}
\hat{\theta}_{\text {Bayes }}=E[\theta \mid \underline{\mathbf{x}}]=\frac{B}{A-1}=\frac{b+1.35 \sum_{i=1}^{m} x_{i}}{a-0.35 n+1.35 m-1} . \tag{14}
\end{equation*}
$$

where $A=(a-0.35 n+1.35 m)$ and $B=\left(b+1.35 \sum_{i=1}^{m} x_{i}\right)$.
It is shown in the article that the Bayes estimator (14) is consistently better than the MLE in regards to accuracy.

Now, consider the mixture prior distribution of inverse-gamma distributions. Let

$$
\begin{equation*}
\pi(\theta)=\sum_{j=1}^{K} k_{j} \frac{b_{j}^{a_{j}} \theta^{-a_{j}-1} e^{-b_{j} / \theta}}{\Gamma\left(a_{j}\right)}, b_{j}>0, a_{j}>0, j=1,2, \ldots, K, \sum_{j=1}^{K} k_{j}=1, \tag{15}
\end{equation*}
$$

where the $j^{\text {th }}$ prior distribution is given by

$$
\pi_{j}(\theta)=\frac{b_{j}^{a_{j}} \theta^{-a_{j}-1} e^{-b_{j} / \theta}}{\Gamma\left(a_{j}\right)}
$$

The pdf of the composite model based on the $j^{t h}$ prior is given by

$$
\begin{gather*}
f_{X}^{j}(x)=\int_{0}^{\infty} L(\underline{\mathrm{x}} \mid \theta) \pi_{j}(\theta) d \theta \\
=\int_{0}^{\infty} c \theta^{0.35 n-1.35 m} e^{-1.35 \sum_{i=1}^{m} x_{i} / \theta} \frac{b_{j}^{a_{j}} \theta^{-a_{j}-1} e^{-b_{j} / \theta}}{\Gamma\left(a_{j}\right)} d \theta \\
=\frac{c b_{j}^{a_{j}}}{\Gamma\left(a_{j}\right)} \int_{0}^{\infty} \theta^{-\left(a_{j}-0.35 n+1.35 m\right)-1} e^{-\frac{b_{j}+1.35 \sum_{i=1}^{m} x_{i}}{\theta}} d \theta . \tag{16}
\end{gather*}
$$

The integrand in the last line of (16) is the kernel of inverse-gamma with parameters $A_{j}$ and $B_{j}$, where $A_{j}=\left(a_{j}-0.35 n+1.35 m\right)$ and $B_{j}=\left(b_{j}+1.35 \sum_{i=1}^{m} x_{i}\right)$. Therefore,

$$
f_{X}^{j}(x)=\frac{c b_{j}^{a_{j}}}{\Gamma\left(a_{j}\right)} \frac{\Gamma\left(A_{j}\right)}{B_{j}^{A_{j}}}
$$

Using the above result, the $j^{\text {th }}$ posterior distribution is

$$
\begin{gathered}
\pi^{j}(\theta \mid x)=\frac{L(\underline{\mathrm{x}} \mid \theta) \pi_{j}(\theta)}{f_{X}^{j}(x)}=\frac{c \theta^{0.35 n-1.35 m} e^{-1.35 \sum_{i=1}^{m} x_{i} / \theta} \frac{{\frac{b}{j} j_{j} \theta^{-a_{j}-1} e^{-b_{j} / \theta}}_{\Gamma\left(a_{j}\right)}}{\frac{c b_{j}^{a_{j}}}{\Gamma\left(a_{j}\right)} \frac{\Gamma\left(A_{j}\right)}{A_{j}}}}{=\frac{\frac{c b_{j}^{a_{j}}}{\Gamma\left(a_{j}\right)} \theta^{-\left(a_{j}-0.35 n+1.35 m\right)-1} e^{-\frac{b_{j}+1.35 \sum_{i=1}^{m} x_{i}}{\theta}}}{\frac{c b_{j}^{a_{j}}}{\Gamma\left(a_{j}\right)} \frac{\Gamma\left(A_{j}\right)}{B_{j}^{A_{j}}}}},
\end{gathered}
$$

which reduces to

$$
\pi^{j}(\theta \mid x)=\frac{B_{j}^{A_{j}} \theta^{-A_{j}-1} e^{-B_{j} / \theta}}{\Gamma\left(A_{j}\right)}
$$

Furthermore, we have

$$
\begin{equation*}
f_{X}(x)=\sum_{j=1}^{K} k_{j} f_{X}^{j}(x)=\sum_{j=1}^{K} k_{j} \frac{c b_{j}^{a_{j}}}{\Gamma\left(a_{j}\right)} \frac{\Gamma\left(A_{j}\right)}{B_{j}^{A_{j}}} \tag{17}
\end{equation*}
$$

hence,

$$
\frac{f_{X}^{j}(x)}{f_{X}(x)}=\frac{\frac{c b_{j}^{a_{j}}}{\Gamma\left(a_{j}\right)} \frac{\Gamma\left(A_{j}\right)}{B_{j}^{A_{j}}}}{\sum_{i=1}^{K} k_{j} \frac{c b_{j}^{a_{j}}}{\Gamma\left(a_{i}\right)} \frac{\Gamma\left(A_{j}\right)}{B_{j}^{A_{j}}}}=\frac{\frac{b_{j}^{a_{j}}}{\Gamma\left(a_{j}\right)} \frac{\Gamma\left(A_{j}\right)}{B_{j}^{A_{j}}}}{\sum_{j=1}^{K} k_{j} \frac{b_{j}^{a_{j}}}{\Gamma\left(a_{j}\right)} \frac{\Gamma\left(A_{j}\right)}{B_{j}^{A_{j}}}}
$$

Using the above results, the posterior distribution, based on the mixture prior distribution, is

$$
\begin{gathered}
\pi(\theta \mid \underline{\mathbf{x}})=\sum_{j=1}^{K} \frac{k_{j} f_{X}^{j}(x)}{f_{X}(x)} \pi^{j}(\theta \mid \underline{\mathbf{x}}) \\
\sum_{j=1}^{K} \frac{k_{j} \frac{b_{j}^{a_{j}}}{\Gamma\left(a_{j}\right)} \frac{\Gamma\left(A_{j}\right)}{B_{j}^{A_{j}}}}{\sum_{j=1}^{K} k_{j} \frac{b_{j}^{a_{j}}}{\Gamma\left(a_{j}\right)} \frac{\Gamma\left(A_{j}\right)}{B_{j}^{A_{j}}}} \frac{B_{j}^{A_{j}} \theta^{-A_{j}-1} e^{-B_{j} / \theta}}{\Gamma\left(A_{j}\right)} .
\end{gathered}
$$

Hence, under the squared-error loss function, the Bayes estimator for $\theta$ is

$$
\begin{gather*}
\hat{\theta}_{\text {Bayes }}=E[\theta \mid \underline{\mathbf{x}}]=\sum_{j=1}^{K} \frac{k_{j} \frac{b_{j}^{a_{j}}}{\Gamma\left(a_{j}\right)} \frac{\Gamma\left(A_{j}\right)}{B_{j}^{A_{j}}}}{\sum_{j=1}^{K} k_{j} \frac{b_{j}^{a_{j}}}{\Gamma\left(a_{j}\right)} \frac{\Gamma\left(A_{j}\right)}{B_{j}}} E^{j}[\theta \mid \underline{\mathbf{x}}]=\sum_{j=1}^{K} \frac{k_{j} \frac{b_{j}^{a_{j}}}{\Gamma\left(a_{j}\right)} \frac{\Gamma\left(A_{j}\right)}{B_{j}^{A_{j}}}}{\sum_{j=1}^{K} k_{j} \frac{b_{j}^{a_{j}}}{\Gamma\left(a_{j}\right)} \frac{\Gamma\left(A_{j}\right)}{B_{j}^{A_{j}}}} \frac{B_{j}}{A_{j}-1} \\
=\sum_{j=1}^{K} \frac{k_{j} \frac{b_{j}^{a_{j}}}{\Gamma\left(a_{j}\right)} \frac{\Gamma\left(A_{j}-1\right)}{B_{j}^{A_{j}-1}}}{\sum_{j=1}^{K} k_{j} \frac{b_{j}^{a_{j}}}{\Gamma\left(a_{j}\right)} \frac{\Gamma\left(A_{j}\right)}{B_{j}^{A_{j}}}} . \tag{18}
\end{gather*}
$$

### 3.2. Bayesian Inference for the Composite IG-Pareto Based on the Mixture Prior Distribution

Aminzadeh and Deng (2019) developed the composite inverse-gamma-Pareto model as follows:

Suppose $X$ is a random variable with the pdf $f(x)$, where $f_{1}(x)$ and $f_{2}(x)$, respectively, are the pdfs of inverse-gamma and Pareto distributions.

$$
f_{X}(x)=\left\{\begin{array}{cl}
c f_{1}(x) & 0<x<\theta \\
c f_{2}(x) & \theta \leq x<\infty
\end{array}\right.
$$

where

$$
f_{1}(x)=\frac{\beta^{\alpha} x^{-\alpha-1} e^{-\beta / x}}{\Gamma(\alpha)}, x>0, \alpha>0, \beta>0
$$

and

$$
f_{2}(x)=\frac{a \theta^{a}}{x^{a+1}}, x \geq \theta, a>0, \theta>0
$$

Recall that the composite pdf $f(x)$ should be smooth at $\theta$. Therefore,

$$
f_{1}(\theta)=f_{2}(\theta), f_{1}^{\prime}(\theta)=f_{2}^{\prime}(\theta)
$$

The simultaneous solutions of the above equations, after algebraic manipulations, lead to

$$
\frac{k^{\alpha} e^{-k}}{\Gamma(\alpha)}=\alpha-k
$$

where $k=\frac{\beta}{\theta}$ and $a=\alpha-k>0$, which implies $\alpha>k>0$. The functions on both sides of the above equations are positive and integrable; therefore, the integrals of the functions on a closed interval should be equal. Hence,

$$
\int_{0}^{\alpha} \frac{k^{\alpha} e^{-k}}{\Gamma(\alpha)} d k=\frac{\alpha^{2}}{2}
$$

Using the gamma function, we obtain

$$
\Gamma(\alpha+1)=\int_{0}^{\alpha} t^{(\alpha+1)-1} e^{-t} d t+\Gamma(\alpha+1, \alpha)
$$

where $\Gamma(\alpha+1, \alpha)=\int_{\alpha}^{\infty} t^{(\alpha+1)-1} e^{-t} d t$ denotes the incomplete upper gamma function. In light of this result, the above equation reduces to

$$
\frac{\Gamma(\alpha+1, \alpha)}{\Gamma(\alpha)}+0.5 \alpha^{2}-\alpha=0
$$

Mathematica can solve the above equation numerically. We obtain $\alpha=0.308289$. As a result, we have $k=0.144351, a=\alpha-k=0.163947$. To find the value of $c$, we need (see the definition of the composite pdf above)

$$
c\left[\frac{\int_{0}^{\theta}(k \theta)^{\alpha} x^{-\alpha-1} e^{-k \theta / x} d x}{\Gamma(\alpha)}+\int_{\theta}^{\infty} \frac{a \theta^{a}}{x^{a+1}} d x\right]=1
$$

which leads to

$$
c=\frac{1}{1+G R(\alpha, k)}=0.711384
$$

Note that GR stands for GammaRegularized and $\operatorname{GR}\left(\alpha, \frac{\beta}{x}\right)$ is the cdf of inverse-gamma with parameters $\alpha, \beta$. Therefore, $G R\left(\alpha, \frac{k \theta}{\theta}\right)$, which is the first integral above, reduces to $G R(\alpha, k)$. Mathematica can compute the GR function. The above findings reveal that four initial parameters reduce to only one parameter $\theta$. As a result, the pdf of the IG-Pareto distribution is

$$
f_{X}(x \mid \theta)= \begin{cases}\frac{c(k \theta)^{\alpha} x^{-\alpha-1} e^{\frac{-k \theta}{x}}}{\Gamma(\alpha)}, & 0<x \leq \theta  \tag{19}\\ \frac{c(\alpha-k) \theta^{\alpha-k}}{x^{\alpha-k+1}}, & \theta \leq x<\infty\end{cases}
$$

and its cdf is given by

$$
F_{X}(x \mid \theta)= \begin{cases}c G R\left(\alpha, \frac{k \theta}{x}\right), & 0<x \leq \theta \\ 1-c\left(\frac{\theta}{x}\right)^{\alpha-k}, & \theta \leq x<\infty\end{cases}
$$

The quantile function can be derived similarly to the exponential-Pareto composite distribution. Using the cdf above, we have $F_{X}(\theta \mid \theta)=c G R(\alpha, k)=1-c$.

Case 1: $0<P \leq 1-c \Longrightarrow c G R\left(\alpha, \frac{k \theta}{x_{P}}\right)=P \Longrightarrow$

$$
x_{p}=\frac{k \theta}{\operatorname{InverseGammaRegularized}\left(\alpha, \frac{p}{c}\right)}
$$

where InverseGammaRegularized $\left(\alpha, \frac{P}{c}\right)$ can be computed via Mathematica.
Case 2: $1-c<P<1 \Longrightarrow 1-c\left(\frac{\theta}{x_{P}}\right)^{\alpha-k}=P \Longrightarrow$

$$
x_{p}=\theta\left(\frac{1-P}{c}\right)^{\left(\frac{1}{k-\alpha}\right)} .
$$

For the special cases $P=0.25$ and $P=0.99$, using the constant values $k=0.144351$, $\alpha=0.308289, c=0.711384$, and Mathematica, we obtain

$$
x_{0.25}=0.723798 \theta \text { and } x_{0.99}=1.98138 \times 10^{11} \theta
$$

Suppose $x_{1}, \ldots, x_{n}$ is a random sample from the IG-Pareto distribution, without loss of generality, we assume $x_{1}<x_{2}<\ldots<x_{n}$. The likelihood function is

$$
\begin{equation*}
L(\underline{\mathrm{x}} \mid \theta)=Q e^{-k \theta \sum_{i=1}^{m} \frac{1}{x_{i}}} \theta^{a(n-m)+\alpha m} \tag{20}
\end{equation*}
$$

where $Q=\frac{c^{n} k^{\alpha m} \prod_{i=1}^{m} x_{i}^{-\alpha-1} a^{n-m}}{\prod_{i=m+1}^{n} x_{i}^{a+1}}$. For the formulation of the likelihood function, we assume an $m(m=1,2, \ldots, n-1)$ exists, such that in the sorted sample $x_{m} \leq \theta \leq x_{m+1}$. The solution to $\frac{\partial L(\mathbf{x} \mid \theta)}{\partial \theta}=0$ is the MLE for $\theta$, which is

$$
\hat{\theta}_{M L E}=\frac{m \alpha+(\alpha-k) m}{k S}, \quad S=\sum_{i=1}^{m} x_{i}^{-1} .
$$

Using (20), the Fisher information is

$$
I(\theta)=-E\left[\frac{\partial^{2} \ln (L(\underline{x} \mid \theta))}{\partial^{2} \theta}\right]=\frac{n \alpha}{\theta^{2}}
$$

As a result, the standard error of the MLE is $\frac{1}{\sqrt{I(\theta)}}=\frac{\theta}{\sqrt{n \alpha}}$.
Using the same algorithm in Section 3.1, we identify the correct value, $m$, and compute $\hat{\theta}_{M L E}$.

Aminzadeh and Deng (2019), as a prior distribution for $\theta$, used gamma $(\gamma, \delta)$ with the pdf

$$
\pi(\theta)=\frac{\theta^{\gamma-1} e^{-\theta / \delta}}{\Gamma(\gamma) \delta^{\gamma}}, \gamma>0, \delta>0
$$

then, the posterior pdf is

$$
\begin{equation*}
f(\theta \mid \underline{\mathbf{x}})=\frac{L(\underline{\mathbf{x}} \mid \theta) \times \pi(\theta)}{\int L(\underline{\mathrm{x}} \mid \theta) \times \pi(\theta) d \theta} \propto e^{-\theta\left(k \sum_{i=1}^{m} \frac{1}{x_{i}}+\frac{1}{\delta}\right)} \theta^{n a+m(\alpha-a)+\gamma-1} . \tag{21}
\end{equation*}
$$

The R.H.S. in (21) is the kernel of $\operatorname{gamma}(A, B), A=n a+m(\alpha-a)+\gamma$ and $B=$ $\frac{\delta}{\left(\delta k \sum_{i=1}^{m} \frac{1}{x_{i}}+1\right)}$. As a result, the pdf of the posterior is given by

$$
\pi(\theta \mid \underline{\mathbf{x}})=\frac{\theta^{A-1} e^{-\theta / B}}{\Gamma(A) B^{A}}
$$

as a result, under the squared-error loss function, the Bayes estimator for $\theta$ is

$$
\hat{\theta}_{\text {Bayes }}=E[\theta \mid \underline{\mathbf{x}}]=A B=\frac{\delta(n a+m k+\gamma)}{\left(\delta k \sum_{i=1}^{m} \frac{1}{x_{i}}+1\right)}
$$

Now, we derive the Bayes estimator using the mixture prior distribution based on individual gamma priors,

$$
\pi_{j}(\theta)=\frac{\theta^{\gamma_{j}-1} e^{-\theta / \delta_{j}}}{\Gamma\left(\gamma_{j}\right) \delta_{i}^{\gamma_{j}}},
$$

as a result

$$
\begin{equation*}
\pi(\theta)=\sum_{j=1}^{K} k_{j} \frac{\theta^{\gamma_{i}-1} e^{-\theta / \delta_{j}}}{\Gamma\left(\gamma_{j}\right) \delta_{j}^{\gamma_{j}}}, \gamma_{j}>0, \delta_{j}>0, j=1,2, \ldots, K, \sum_{j=1}^{K} k_{j}=1 . \tag{22}
\end{equation*}
$$

The pdf of the composite model based on $j^{\text {th }}$ prior is given by

$$
\begin{gather*}
f_{X}^{j}(x)=\int_{0}^{\infty} L(\underline{\mathrm{x}} \mid \theta) \pi_{j}(\theta) d \theta \\
=\int_{0}^{\infty} Q e^{-k \theta \sum_{i=1}^{m} \frac{1}{x_{i}}} \theta^{a(n-m)+\alpha m} \frac{\theta^{\gamma_{j}-1} e^{-\theta / \delta_{j}}}{\Gamma\left(\gamma_{j}\right) \delta_{i}^{\gamma_{j}}} d \theta \\
=\frac{Q}{\Gamma\left(\gamma_{i}\right) \delta_{i}^{\gamma_{i}}} \int_{0}^{\infty} \theta^{a(n-m)+\alpha m+\gamma_{j}-1} e^{-\theta\left(k \sum_{i=1}^{m} \frac{1}{x_{i}}+\frac{1}{\delta_{j}}\right)} d \theta \tag{23}
\end{gather*}
$$

The RHS of the last line in (23) is the kernel of $\operatorname{Gamma}\left(A_{j}, B_{j}\right)$, where

$$
A_{j}=a(n-m)+\alpha m+\gamma_{j}, \quad B_{j}=\frac{1}{k \sum_{i=1}^{m} \frac{1}{x_{i}}+\frac{1}{\delta_{j}}} .
$$

Therefore, $f_{X}^{j}(x)=\frac{Q \Gamma\left(A_{j}\right) B_{i}^{A_{j}}}{\Gamma\left(\gamma_{j}\right) \delta_{j}^{\gamma_{j}}}$ and

$$
\begin{gather*}
f_{X}(x)=\sum_{j=1}^{K} k_{j} f_{X}^{j}(x), \\
\frac{f_{X}^{j}(x)}{f_{X}(x)}=\frac{\frac{Q \Gamma\left(A_{j}\right) B_{j}^{A_{j}}}{\Gamma\left(\gamma_{j}\right) \delta_{j}^{\gamma_{j}}}}{\sum_{j=1}^{K} k_{j} \frac{Q \Gamma\left(A_{j}\right) B_{j}^{A_{j}}}{\Gamma\left(\gamma_{i}\right) \delta_{j}^{\gamma_{j}}}}=\frac{\frac{\Gamma\left(A_{j}\right) B_{j}^{A_{j}}}{\Gamma\left(\gamma_{j}\right) \delta_{j}^{\gamma_{j}}}}{\sum_{j=1}^{K} k_{j} \frac{\Gamma\left(A_{j}\right) B_{j}^{A_{j}}}{\Gamma\left(\gamma_{j}\right) \delta_{j}^{\gamma_{j}}}} . \tag{24}
\end{gather*}
$$

Therefore, the pdf of $j^{\text {th }}$ posterior distribution is

$$
\begin{align*}
\pi^{j}(\theta \mid x) & =\frac{L(\underline{x} \mid \theta) \pi_{j}(\theta)}{f_{X}^{j}(x)}=\frac{Q e^{-k \theta \sum_{i=1}^{m} \frac{1}{x_{i}}} \theta^{a(n-m)+\alpha m} \frac{\theta^{\gamma_{j}-1} e^{-\theta / \delta_{j}}}{\Gamma\left(\gamma_{j}\right) \delta_{j}^{\gamma_{j}}}}{\frac{Q \Gamma\left(A_{j}\right) B_{j}^{A_{j}}}{\Gamma\left(\gamma_{j}\right) \delta_{j}^{\gamma_{j}}}} \\
\pi^{j}(\theta \mid x) & =\frac{\theta^{a(n-m)+\alpha m+\gamma_{j}-1} e^{-\theta\left(k \sum_{i=1}^{m} \frac{1}{x_{j}}+\frac{1}{\delta_{j}}\right)}}{\Gamma\left(A_{j}\right) B_{j}^{A_{j}}}=\frac{\theta^{A_{j}-1} e^{\frac{-\theta}{B_{j}}}}{\Gamma\left(A_{j}\right) B_{j}^{A_{j}}} \tag{25}
\end{align*}
$$

From (24) and (25), we conclude that the posterior distribution for IG-Pareto based on the mixture of gamma priors is

$$
\pi(\theta \mid x)=\sum_{j=1}^{K} \frac{k_{j} f_{X}^{j}(x)}{f_{X}(x)} \pi^{j}(\theta \mid x)=\sum_{j=1}^{K} \frac{k_{j} \frac{\Gamma\left(A_{j}\right) B_{j}^{A_{j}}}{\Gamma\left(\gamma_{j}\right) \delta_{j}^{\gamma_{j}}}}{\sum_{j=1}^{K} k_{j} \frac{\Gamma\left(A_{j}\right) B_{i}^{A_{j}}}{\Gamma\left(\gamma_{j}\right) \delta_{j}^{\gamma_{j}}}} \frac{\theta^{A_{j}-1} e^{\frac{-\theta}{B_{j}}}}{\Gamma\left(A_{j}\right) B_{i}^{A_{j}}} .
$$

Hence, under the squared-error loss function, the Bayes estimator for $\theta$ is

$$
\begin{gather*}
\hat{\theta}_{\text {Bayes }}=E[\theta \mid \underline{\mathbf{x}}]=\sum_{j=1}^{K} \frac{k_{j} f_{X}^{j}(x)}{f_{X}(x)} E^{j}[\theta \mid \underline{\mathbf{x}}]=\sum_{j=1}^{K} k_{j} \frac{\frac{\Gamma\left(A_{j}\right) B_{j}^{A_{j}}}{\Gamma\left(\gamma_{j}\right) \delta_{j}^{\gamma_{j}}}}{\sum_{j=1}^{K} k_{j} \frac{\Gamma\left(A_{j}\right) B_{j}^{A_{j}}}{\Gamma\left(\gamma_{i}\right) \delta_{j}^{\gamma_{j}}}} A_{j} B_{j} \\
=\sum_{j=1}^{K} \frac{k_{j} \frac{\Gamma\left(A_{j}+1\right) B_{j}^{A_{j}+1}}{\Gamma\left(\gamma_{j}\right) \delta_{j}^{\gamma_{j}}}}{\sum_{j=1}^{K} k_{j} \frac{\Gamma\left(A_{j}\right) B_{j}^{A_{j}}}{\Gamma\left(\gamma_{j}\right) \delta_{j}^{\gamma_{j}}}} \tag{26}
\end{gather*}
$$

## 4. Simulation

### 4.1. Simulation for Composite Exponential-Pareto

To compare the accuracies of $\hat{\theta}_{M L E}$ and $\hat{\theta}_{\text {Bayes }}$ (with and without a mixture of prior distributions), simulations are conducted using Mathematica. For the same generated sample, the code computes estimators using ( $\left.a_{j}, b_{j}, j=1,2, \ldots K\right)$, weights $k_{1}, k_{2}, \ldots, k_{K},\left(\sum_{j=1}^{K} k_{j}=1\right)$. For each set of input parameters in the simulation, $N=1000$ samples from the composite density (10) are generated.

For a random sample $x_{1}, \ldots, x_{n}$ from the composite pdf (10) and without loss of generality, consider the ordered sample $x_{1}<x_{2}<\ldots<x_{n}$. Recall (18),

$$
\hat{\theta}_{\text {Bayes }}=\sum_{j=1}^{K} \frac{k_{j} \frac{b_{j}^{a_{j}}}{\Gamma\left(a_{j}\right)} \frac{\Gamma\left(A_{j}-1\right)}{B_{j}}}{\sum_{j=1}^{K} k_{j} \frac{b_{j}^{a_{j}}}{\Gamma\left(a_{j}\right)} \frac{\Gamma\left(A_{j}\right)}{B_{j}}}
$$

The following algorithm is used to determine $m$ :

1. Start with $m=1$, check to see if $x_{1} \leq \hat{\theta}_{\text {Bayes }} \leq x_{2}$, if yes, then $m=1$. Otherwise, go to step 2.
2. For $m=2$, if $x_{2} \leq \hat{\theta}_{\text {Bayes }} \leq x_{3}$, then $m=2$, otherwise we consider $m=3$ and continue until we find the correct value for $m$. The idea is to find the value for $m$ so that $x_{m} \leq \hat{\theta}_{\text {Bayes }} \leq x_{m+1}$. The Mathematica code uses the algorithm to find $m$ and compute $\hat{\theta}_{\text {Bayes }}$.
Selecting hyperparameter values could be challenging. Suppose two experts can provide partial prior information about the hyperparameter values. See, Rufo et al. (2010). The idea with the mixture prior distribution is to incorporate both experts' opinions to find the Bayes estimate of $\theta$. In this article, we use the same weights ( $k_{1}=k_{2}=0.5$ ) for each expert's opinion and consider two cases when $K=2$ :
$b_{1} \neq b_{2}, \quad a_{1} \neq a_{2} \quad$ values of $b_{1}$ and $b_{2}$ are provided by the experts.
$a_{1} \neq a_{2}, \quad b_{1} \neq b_{2} \quad$ values of $a_{1}$ and $a_{2}$ are provided by the experts.

It is noted that $B_{1}=b_{1}+1.35 \sum_{i=1}^{m} x_{i}$ and $B_{2}=b_{2}+1.35 \sum_{i=1}^{m} x_{i}$, which implies that $B_{1} \neq B_{2}$. It is also noted $A_{1}=a_{1}-0.35 n+1.35 m$ and $A_{2}=a_{2}-0.35 n+1.35 m$, and $A_{1} \neq A_{2}$. From (18), we have

$$
\hat{\theta}_{\text {Bayes }}=\frac{\frac{b_{1}^{a_{1}} \Gamma\left(A_{1}-1\right)}{\Gamma\left(a_{1}\right) B_{1}^{A_{1}-1}}+\frac{b_{2}^{a_{2}} \Gamma\left(A_{2}-1\right)}{\Gamma\left(a_{2}\right) B_{2}^{A_{2}-1}}}{\frac{b_{1}^{a_{1}} \Gamma\left(A_{1}\right)}{\Gamma\left(a_{1}\right) B_{1}^{A_{1}}}+\frac{b_{2}^{a_{2}} \Gamma\left(A_{2}\right)}{\Gamma\left(a_{2}\right) B_{2}^{A_{2}}}} .
$$

Case 1: In this case, experts are quite sure about the values for $b_{1}$ and $b_{2}$; therefore, there are only two hyperparameter values that should be selected. We would like the optimal values for $a_{1}, a_{2}$. Given values of $\theta, b_{1}, b_{2}$, Mathematica provides optimal values of $a_{1}, a_{2}$ via a numerical optimization and the constraints $E[\theta]=0.5\left(\frac{b_{1}}{a_{1}-1}+\frac{b_{2}}{a_{2}-1}\right)=\theta$ and, $a_{1}>2, a_{2}>2$.

$$
\text { NMinimize } \left.\left[\{\operatorname{Var}(\theta)], E[\theta]==\theta, a_{2}>2, a_{1}>2\right\},\left\{a_{1}, a_{2}\right\}\right] .
$$

For example, for $\theta=5, b_{1}=25, b_{2}=22$, we obtain $a_{1}=5.90681, a_{2}=5.48518$. Note that unlike in simulation studies, for a real dataset, a selected value for $\theta$ is not available. Hence, we propose a data-driven approach be used to compute $a_{1}, a_{2}$. Meaning, the equation $0.5\left(\frac{b_{1}}{a_{1}-1}+\frac{b_{2}}{a_{2}-1}\right)=\hat{\theta}_{M L E}$ along with the above optimization command provide values for $a_{1}$ and $a_{2}$.

Table 1a reveals that by selecting the hyperparameters, as described above, the mixture prior approach gives a more accurate Bayes estimate, as the average squared-error $=$ ASE (Bayes) $=\xi\left(\hat{\theta}_{\text {Bayes }}\right)$ is smaller than its counterpart that does not use a mixture prior. For example, for $b_{1}=260, b_{2}=235$, we obtain $a_{1}=52.921, a_{2}=48.0728$. We can see that the smallest ASE values $=0.45027$ and 0.41351 , corresponding to the optimal set of hyperparameter values of $n=30,100$, respectively. Also, comparing Table 1a with Table 1b, it is clear that both Bayes estimators (with and without the mixture prior) outperform MLE with regard to their accuracies, as $\xi\left(\hat{\theta}_{\text {Bayes }}\right)$ is much smaller than $\xi\left(\hat{\theta}_{M L E}\right)$. Boldface numbers in tables indicate the optimal values.

Table 1. Improved Bayes estimation with hyperparameter selection for mixture priors.

| (a): Comparison of the Bayes estimator without the mixture prior $(\mathrm{K}=1)$ and with the mixture prior $(\mathrm{K}=2)$ when $b_{1}, b_{2}$ are given |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\theta=5$ |  |  |  |  |  |  |  |
| K | $n$ | $b_{1}$ | $b_{2}$ | $a_{1}$ | $a_{2}$ | $\overline{\hat{\theta}}_{\text {Bayes }}$ | $\xi\left(\hat{\theta}_{\text {Bayes }}\right)$ |
| 1 | 30 | 245 | $N A$ | $a_{1}=\frac{b_{1}}{\theta}+1$ | NA | 5.11732 | 0.42551 |
| 2 | 30 | 260 | 235 | 52.921 | 48.027 | 5.11662 | 0.42337 |
| 2 | 30 | 260 | 235 | 55 | 46.32143 | 5.12472 | 0.45379 |
| 1 | 100 | 245 | NA | $a_{1}=\frac{b_{1}}{\theta}+1$ | NA | 5.06908 | 0.41555 |
| 2 | 100 | 260 | 235 | 52.921 | 48.027 | 5.06875 | 0.41351 |
| 2 | 100 | 260 | 235 | 55 | 46.32143 | 5.09362 | 0.54073 |
| (b): Mean and $\sqrt{A S E}$ of MLE of $\theta$ |  |  |  |  |  |  |  |
| $\theta=5$ |  |  |  |  |  |  |  |
|  |  |  | $n$ | $\overline{\hat{\theta}}_{M L E}$ | $\xi\left(\hat{\theta}_{M L E}\right)$ |  |  |
|  |  |  | 30 | 7.39707 | 4.88656 |  |  |
|  |  |  | 100 | 6.09238 | 1.74385 |  |  |

Case 2: In this case, experts are pretty sure about the values for $a_{1}$ and $a_{2}$, and they would like the optimal values for $b_{1}, b_{2}$. Given values of $\theta, a_{1}, a_{2}$, Mathematica pro-
vides optimal values of $b_{1}, b_{2}$ via a numerical optimization and the constraint $E[\theta]=.5$ $\left(\frac{b_{1}}{a_{1}-1}+\frac{b_{2}}{a_{2}-1}\right)=\theta$.

$$
\text { NMinimize } \left.[\{\operatorname{Var}(\theta)], E[\theta]==\theta\},\left\{b_{1}, b_{2}\right\}\right], b_{1}>0, b_{2}>0 .
$$

For example, for $\theta=5, a_{1}=4, a_{2}=6$, we obtain $b_{1}=13.63, b_{2}=27.27$. Similar to case 2 , to compute $b_{1}, b_{2}$, we use $0.5\left(\frac{b_{1}}{a_{1}-1}+\frac{b_{2}}{a_{2}-1}\right)=\hat{\theta}_{M L E}$ in the above optimization command to find $b_{1}, b_{2}$.

Like the previous case, Table 2a confirms that by selecting the hyperparameters, as described above, the mixture prior approach provides a more accurate Bayes estimate, as ASE (Bayes) $=\xi\left(\hat{\theta}_{\text {Bayes }}\right)$ is smaller than its counterpart that does not use a mixture prior. For example, for $a_{1}=110, a_{2}=98$, we obtain $b_{1}=545.312, b_{2}=484.727$. Again, the smallest ASE values $=0.26801$ and 0.24191 , corresponding to the optimal hyperparameter values of $\mathrm{n}=30$ and 100, respectively. Also, comparing Table 2a with Table 2b, in this case, both Bayes estimators (with and without the mixture prior) outperform MLE with regard to their accuracies.

Table 2. Improved Bayes estimation with hyperparameter selection for mixture priors.

| (a): Comparison of the Bayes estimator without the mixture prior $(\mathrm{K}=1)$ and with the mixture prior $(\mathrm{K}=2)$ for given $a_{1}, a_{2}$ |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\theta=5$ |  |  |  |  |  |  |  |
| K | $n$ | $a_{1}$ | $a_{2}$ | $b_{1}$ | $b_{2}$ | $\overline{\hat{\theta}}_{\text {Bayes }}$ | $\xi\left(\hat{\theta}_{\text {Bayes }}\right)$ |
| 1 | 30 | 100 | NA | $b_{1}=\theta\left(a_{1}-1\right)$ | NA | 5.07243 | 0.27438 |
| 2 | 30 | 110 | 98 | 545.312 | 484.727 | 5.07036 | 0.26801 |
| 2 | 30 | 110 | 98 | 560 | 471.651 | 5.07327 | 0.27393 |
| 1 | 100 | 100 | NA | $b_{1}=\theta\left(a_{1}-1\right)$ | NA | 5.03404 | 0.24642 |
| 2 | 100 | 110 | 98 | 545.312 | 484.727 | 5.03312 | 0.23956 |
| 2 | 100 | 110 | 98 | 560 | 471.651 | 5.03479 | 0.25197 |

(b): Mean and $\sqrt{A S E}$ of MLE of $\theta$

| $\theta=5$ |  |  |
| :---: | :---: | :---: |
| $n$ | $\overline{\hat{\theta}}_{M L E}$ | $\xi\left(\hat{\theta}_{M L E}\right)$ |
|  | 30 | 7.44702 |
| 5.34616 |  |  |
|  | 100 | 6.05374 |

### 4.2. Simulation for Composite Inverse-Gamma-Pareto

Simulations compare similarly to the composite exponential-Pareto model to compare the accuracy of $\hat{\theta}_{\text {Bayes }}$ based on with and without mixture prior distributions. For selected values of $n$ and $\theta$, the hyperparameters ( $\gamma_{j}, \delta_{j}, j=1,2, \ldots K$ ) for gamma prior distributions, and weights $k_{1}, k_{2}, \ldots, k_{K},\left(\sum_{j=1}^{K} k_{j}=1\right)$. The simulation study generates $N=1000$ samples from the composite density (19).

Given a random sample $x_{1}, \ldots, x_{n}$ from the composite pdf in (19), without loss of generality, consider the ordered sample $x_{1}<\ldots<x_{n}$. Recall the Bayes estimator (26), which uses the mixture gamma prior distributions,

$$
\hat{\theta}_{\text {Bayes }}=\sum_{j=1}^{K} \frac{k_{j} \frac{\Gamma\left(A_{j}+1\right) B_{j}^{A_{j}+1}}{\Gamma\left(\gamma_{j}\right) \delta_{j}^{\gamma_{j}}}}{\sum_{j=1}^{K} k_{j} \frac{\Gamma\left(A_{j}\right) B_{j}^{A_{j}}}{\Gamma\left(\gamma_{j}\right) \delta_{j}^{\gamma_{j}}}} .
$$

The algorithm described in Section 4.1 determines the value of $m$. Like the exp-Pareto composite distribution case, we must select the prior distributions' hyperparameters. We consider the mixture prior distribution with equal weights and $K=2$. For the prior distribution $\pi_{1}(\theta)=\operatorname{gamma}\left(\gamma_{1}, \delta_{1}\right)$, and $\pi_{2}(\theta)=\operatorname{gamma}\left(\gamma_{2}, \delta_{2}\right)$, consider two cases:
$\gamma_{1} \neq \gamma_{2}, \quad \delta_{1} \neq \delta_{2}$ values of $\gamma_{1}$ and $\gamma_{2}$ are provided by the experts.
$\delta_{1} \neq \delta_{2}, \quad \gamma_{1} \neq \gamma_{2} \quad$ values of $\delta_{1}$ and $\delta_{2}$ are provided by the experts.
Here, $\gamma_{1} \neq \gamma_{2}, \quad \delta_{1} \neq \delta_{2}$, under the assumption $k_{1}=k_{2}=0.5$, it can be shown that

$$
\begin{gathered}
E(\theta)=0.5\left(\gamma_{1} \delta_{1}+\gamma_{2} \delta_{2}\right) \\
\operatorname{Var}(\theta)=0.25\left(\gamma_{1} \delta_{1}-\gamma_{2} \delta_{2}\right)^{2}+0.5 \gamma_{1} \delta_{1}^{2}+0.5 \gamma_{2} \delta_{2}^{2}
\end{gathered}
$$

Tables 3a and 4a provide simulation results for Cases 1 and 2. Since $\delta_{1} \neq \delta_{2}$, and $\gamma_{1} \neq \gamma_{2}$, we have

$$
\begin{gathered}
A_{1}=a(n-m)+\alpha m+\gamma_{1} \neq A_{2}=a(n-m)+\alpha m+\gamma_{2} \\
B_{1}=\frac{1}{k \sum_{i=1}^{m} \frac{1}{x_{i}}+\frac{1}{\delta_{1}}} \neq B_{2}=\frac{1}{k \sum_{i=1}^{m} \frac{1}{x_{i}}+\frac{1}{\delta_{2}}} .
\end{gathered}
$$

From (26), we have

$$
\hat{\theta}_{\text {Bayes }}=\frac{\frac{\Gamma\left(A_{1}+1\right) B_{1}^{A_{1}+1}}{\Gamma\left(\gamma_{1}\right) \delta_{1}^{\gamma_{1}}}+\frac{\Gamma\left(A_{2}+1\right) B_{2}^{A_{2}+1}}{\Gamma\left(\gamma_{2}\right) \delta_{2}^{\gamma_{2}}}}{\frac{\Gamma\left(A_{1}\right) B_{1}^{A_{1}}}{\Gamma\left(\gamma_{1}\right) \delta_{1}^{\gamma_{1}}}+\frac{\Gamma\left(A_{2}\right) B_{2}^{A_{2}}}{\Gamma\left(\gamma_{2}\right) \delta_{2}^{\gamma_{2}^{2}}}} .
$$

Case 1: $\gamma_{1} \neq \gamma_{2}, \quad \delta_{1} \neq \delta_{2}$ where values of $\gamma_{1}$ and $\gamma_{2}$ are provided by the experts. For given values of $\theta, \gamma_{1}$, and $\gamma_{2}$, Mathematica provides optimal values of $\delta_{1}, \delta_{2}$ via a numerical minimization for $\operatorname{Var}(\theta)$, and the constraint $E(\theta)=0.5\left(\gamma_{1} \delta_{1}+\gamma_{2} \delta_{2}\right)=\theta$. Again, when we have a real dataset, $0.5\left(\gamma_{1} \delta_{1}+\gamma_{2} \delta_{2}\right)=\hat{\theta}_{M L E}$ is used in the optimization command below in Mathematica to compute hyperparameters $\delta_{1}, \delta_{2}$.

$$
\text { NMinimize }\left[\left\{\operatorname{Var}(\theta), E[\theta]==\theta, \delta_{1}>0, \delta_{2}>0\right\},\left\{\delta_{1}, \delta_{2}\right\}\right] \text {. }
$$

For example, when $\theta=5, \gamma_{1}=2$, and $\gamma_{2}=2.5$, the optimal solutions are $\delta_{1}=2.41379$, $\delta_{2}=2.06897$. When $\theta=5, \gamma_{1}=5$, and $\gamma_{2}=5.5$, the optimal solutions $\delta_{1}=0.99236$, $\delta_{2}=0.91603$.

Table 3a reveals that by selecting the hyperparameters, as described above, the mixture prior approach gives a more accurate Bayes estimate, as ASE (Bayes) $=\xi\left(\hat{\theta}_{\text {Bayes }}\right)$ is smaller than its counterpart that does not use a mixture prior. For example, for a sample size of $n=30$, the smallest ASE values = 1.72061 and 1.16326, corresponding to the optimal two sets of hyperparameter values. Also, for a sample size of $n=100$, the smallest ASE values 1.06654 and 0.91892 , corresponding to the optimal two sets of hyperparameter values. In this case, Table 3a,b suggest that both Bayes estimators (with and without mixture prior distributions) outperform MLE.

Case 2: $\delta_{1} \neq \delta_{2}, \quad \gamma_{1} \neq \gamma_{2}$ where values of $\delta_{1}$ and $\delta_{2}$ are provided by the experts.
Given the values of $\theta, \delta_{1}$, and $\delta_{2}$, Mathematica provides optimal values of $\gamma_{1}, \gamma_{2}$ via a numerical minimization for $\operatorname{Var}(\theta)$ and the constraint $E(\theta)=0.5\left(\gamma_{1} \delta_{1}+\gamma_{2} \delta_{2}\right)=\theta$.

$$
\text { NMinimize }\left[\left\{\operatorname{Var}(\theta), E[\theta]==\theta, \gamma_{1}>0, \gamma_{2}>0\right\},\left\{\gamma_{1}, \gamma_{2}\right\}\right] .
$$

For example, for $\theta=5, \delta_{1}=2.4$, and $\delta_{2}=2.6$, the optimal solutions are $\gamma_{1}=2.10417$, $\gamma_{2}=1.90384$. For $\theta=5, \delta_{1}=1$, and $\delta_{2}=1.1$, the optimal solutions are $\gamma_{1}=5.025$,
$\gamma_{2}=4.52273$. Similar to other cases, for real data, $0.5\left(\gamma_{1} \delta_{1}+\gamma_{2} \delta_{2}\right)=\hat{\theta}_{M L E}$ along with the above command is used to find hyperparameters $\gamma_{1}, \gamma_{2}$.

Table 4a reveals that the mixture prior approach gives a more accurate Bayes estimate, as ASE (Bayes) $=\xi\left(\hat{\theta}_{\text {Bayes }}\right)$ is smaller than its counterpart that does not use a mixture prior. For example, for a sample size of $n=30$, the smallest ASE values $=1.85072$ and 1.20807, corresponding to the optimal two sets of hyperparameter values. Also, for a sample size of $n=100$, the smallest ASE values $=1.08946$ and 0.96434 , corresponding to the optimal two sets of hyperparameter values. Table 4a,b reveal that both Bayes estimators (with and without mixture prior distributions) outperform MLE.

Table 3. Bayes estimators (with and without mixture prior distributions) outperform MLE.

| (a): Comparison of the Bayes estimator without the mixture prior $(\mathrm{K}=1)$ <br> and with the mixture prior $(\mathrm{K}=2)$ <br> when $\gamma_{1}, \gamma_{2}$ are given |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $K$ | $n$ | $\gamma_{1}$ | $\gamma_{2}$ | $\delta_{1}$ | $\delta_{2}$ | $\overline{\hat{\theta}}_{\text {Bayes }}$ | $\xi\left(\hat{\theta}_{\text {Bayes }}\right)$ |
| 1 | 30 | 2 | $N A$ | $\delta_{1}=\frac{\theta}{\gamma_{1}}$ | NA | 5.44943 | 1.70629 |
| 2 | 30 | 2 | 2.5 | $\mathbf{2 . 4 1 3 7 9}$ | $\mathbf{2 . 0 6 8 9 7}$ | 5.42456 | $\mathbf{1 . 6 3 4 8 1}$ |
| 2 | 30 | 2 | 2.5 | 3 | 1.6 | 5.39921 | 1.70996 |
| 1 | 30 | 5 | $N A$ | $\delta_{1}=\frac{\theta}{\gamma_{1}}$ | NA | 5.30638 | 1.19520 |
| 2 | 30 | 5 | 5.5 | $\mathbf{0 . 9 9 2 3}$ | $\mathbf{0 . 9 1 6 0 3}$ | 5.2944 | $\mathbf{1 . 1 6 3 2 6}$ |
| 2 | 30 | 5 | 5.5 | 1.1 | 0.81818 | 5.30043 | 1.21116 |
| 1 | 100 | 2 | $N A$ | $\delta_{1}=\frac{\theta}{\gamma_{1}}$ | NA | 5.16417 | 1.08021 |
| 2 | 100 | 2 | 2.5 | $\mathbf{2 . 4 1 3 7 9}$ | $\mathbf{2 . 0 6 8 9 7}$ | 5.16189 | $\mathbf{1 . 0 6 6 5 4}$ |
| 2 | 100 | 2 | 2.5 | 3 | 1.6 | 5.13296 | 1.06718 |
| 1 | 100 | 5 | $N A$ | $\delta_{1}=\frac{\theta}{\gamma_{1}}$ | NA | 5.15904 | 0.92941 |
| 2 | 100 | 5 | 5.5 | $\mathbf{0 . 9 9 2 3}$ | $\mathbf{0 . 9 1 6 0 3}$ | 5.15621 | $\mathbf{0 . 9 1 8 9 2}$ |
| 2 | 100 | 5 | 5.5 | 1.1 | 0.81818 | 5.14846 | 0.92938 |

(b): Mean and $\sqrt{A S E}$ of MLE of $\theta$

|  | $\theta=5$ |  |  |
| :---: | :---: | :---: | :---: |
| $n$ | $\overline{\hat{\theta}}_{\text {MLE }}$ | $\xi\left(\hat{\theta}_{M L E}\right)$ |  |
| 30 | 5.90731 | 2.8898 |  |
|  | 100 | 5.16417 |  |

Table 4. Improved Bayes estimation with hyperparameter selection for mixture priors.
(a): Comparison of the Bayes estimator without the mixture prior ( $\mathrm{K}=1$ ) and with the mixture prior $(\mathrm{K}=2)$ when $\delta_{1}, \delta_{2}$ are given

| $\theta=5$ |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $K$ | $n$ | $\delta_{1}$ | $\delta_{2}$ | $\gamma_{1}$ | $\gamma_{2}$ | $\overline{\hat{\theta}}_{\text {Bayes }}$ | $\xi\left(\hat{\theta}_{\text {Bayes }}\right)$ |
| 1 | 30 | 2.5 | $N A$ | $\gamma_{1}=\frac{\theta}{\delta_{1}}$ | NA | 5.58552 | 1.85126 |
| 2 | 30 | 2.4 | 2.6 | $\mathbf{2 . 1 0 4 1 7}$ | $\mathbf{1 . 9 0 3 8 4}$ | 5.58569 | $\mathbf{1 . 8 5 0 7 2}$ |
| 2 | 30 | 2.4 | 2.6 | 3 | 1.07692 | 5.79744 | 2.10257 |
| 1 | 30 | 1 | $N A$ | $\gamma_{1}=\frac{\theta}{\delta_{1}}$ | NA | 5.75102 | 2.27942 |
| 2 | 30 | 1 | 1.1 | 5.025 | $\mathbf{4 . 5 2 2 7 3}$ | 5.30832 | $\mathbf{1 . 2 0 8 0 7}$ |
| 2 | 30 | 1 | 1.1 | 5.5 | 4.09091 | 5.32796 | 1.23492 |
| 1 | 100 | 2.5 | $N A$ | $\gamma_{1}=\frac{\theta}{\delta_{1}}$ | NA | 5.21317 | 1.08963 |
| 2 | 100 | 2.4 | 2.6 | $\mathbf{2 . 1 0 4 1 7}$ | $\mathbf{1 . 9 0 3 8 4}$ | 5.21335 | $\mathbf{1 . 0 8 9 4 6}$ |
| 2 | 100 | 2.4 | 2.6 | 3 | 1.07692 | 5.28317 | 1.14863 |
| 1 | 100 | 1 | $N A$ | $\gamma_{1}=\frac{\theta}{\delta_{1}}$ | NA | 5.22784 | 1.18276 |
| 2 | 100 | 1 | 1.1 | 5.025 | $\mathbf{4 . 5 2 2 7 3}$ | 5.16032 | $\mathbf{0 . 9 6 4 3 4}$ |
| 2 | 100 | 1 | 1.1 | 5.5 | 4.09091 | 5.17063 | 0.97477 |

Table 4. Cont.

| (b): Mean and $\sqrt{A S E}$ of MLE of $\theta$ |  |  |  |
| :---: | :---: | :---: | :---: |
|  | $\theta=5$ |  |  |
|  | $n$ | $\hat{\theta}_{M L E}$ | $\xi\left(\hat{\theta}_{M L E}\right)$ |
| 30 | 6.183 | 3.69041 |  |
| 100 | 5.21317 | 1.08963 |  |

## 5. Numerical Example

### 5.1. Data and Basic Descriptive Statistics

This section considers possible models via methods presented in the article for the dataset. The objectives are to find out if using the mixture prior approach in the Bayesian framework provides better results concerning the Bayes estimate of the parameter $\theta$ and to select the best model that fits the data. The insurance losses from natural events, such as floods, are obtained from EM-DAT, the International Disaster Database. EM-DAT contains all natural events worldwide in raw data on the occurrences and effects from 1900 to the present day. "The database is compiled from various sources, including the United Nations agencies, non-governmental organizations, insurance companies, research institutes, and press agencies". This paper considers flood insurance damage in the USA from 2000 to 2019. EM-DAT also provides the annual average CPI using the base year 2019. To eliminate the effect of inflation, all insurance damage amounts are converted to 2019 dollars.

Figure 2 shows the CPI-adjusted (in 2019 dollars) histogram of insurance damage amounts from 2000 to 2019 in the USA. There were 23 recorded insurance losses due to natural event floods in the USA.

Figure 2 provides the frequentist statistics. The average insurance loss due to a natural event storm is $\bar{x}=$ USD 62.6694 million, the minimum loss is USD 5.3996 million, and the maximum loss is USD 266.302 million. The standard deviation is $s=$ USD 71.0041 million, which indicates that the data are widespread. The skewness is 1.96116 , which indicates that the data are right-skewed. The kurtosis is 5.79031, which tells us the data have a heavy tail. The histogram also shows the high frequency for small amounts of damage and the low frequency for large amounts of insurance losses. The data represent typical insurance data for which composite models are applicable; see Aminzadeh and Deng (2019). The regular distributions, such as normal and exponential, cannot effectively model the losses.


Figure 2. Histogram of insurance damage from 2000 to 2019 in 2019 USD dollars.

### 5.2. Model Selection

Miljkovic and Grün (2016) provide goodness-of-fit measures to determine the appropriateness of the fitted models.

NLL: the negative log-likelihood
NLL is used to compare models with the same number of parameters. NLL corresponds to the model with the minimum value of $-\ln \left(L\left(x_{1}, x_{2}, \ldots, x_{n} \mid \underline{\theta}\right)\right.$, among models
considered, where $L\left(x_{1}, x_{2}, \ldots, x_{n} \mid \underline{\theta}\right)$ is the likelihood function of data, and $\underline{\theta}$ is the parameter that can be multi-dimensional. The model with a smaller NLL value indicates that the model has a better fit for the data than other models under consideration.

AIC: Akaike's information criterion
To compare models with different parameter numbers, we consider AIC (Akaike's information criterion) and BIC (Bayesian information criterion). Both measures penalize the increase in the number of parameters.

Akaike developed AIC,

$$
A I C=-\ln \left(L\left(x_{1}, x_{2}, \ldots, x_{n} \mid \underline{\theta}\right)+2 q\right.
$$

where $q$ is the number of parameters. As $q$ increases, $-\ln \left(\left(L\left(x_{1}, x_{2}, \ldots x_{n} \mid \underline{\theta}\right)\right)\right.$ decreases and $2 q$ becomes larger. It provides the trade-off between the models with different parameters. The model with a smaller AIC value indicates that the model has a better fit for the data than other models under consideration.

BIC: Bayesian information criterion
BIC was proposed by Schwarz and is given by

$$
B I C=-\ln \left(L\left(x_{1}, x_{2}, \ldots, x_{n} \mid \underline{\theta}\right)+q \ln (n)\right.
$$

which also depends on the sample size $n$. BIC not only penalizes the increase in the number of parameters but also the increase in the sample size. The smallest BIC indicates the best-fitted model among the models under consideration.

### 5.2.1. Goodness-of-Fit Measures for Maximum Likelihood Method

Table 5 provides the MLE, NLL, AIC, and BIC values for different models. Standard errors of MLEs (see Sections 4.1 and 4.2) also are listed in the table. For example, when we use the exponential model to fit the insurance flood loss data, there is one unknown parameter $\lambda$. The MLE of $\lambda$ based on the exponential model is $\hat{\lambda}=0.0159$. Using $\hat{\lambda}$, the goodness-of-fit measures-NLL, AIC, and BIC-are computed as 118.171, 238.342, and 239.478. Table 5 reveals that based on NLL, AIC, and BIC, among non-composite models (exponential, inverse-gamma) and composite models (exp-Pareto, IG-Pareto), IG-Pareto has the smallest NLL, which is 105.701 . Therefore, IG-Pareto is the best-fitted model among all four models for insurance losses due to natural event floods from 2000 to 2019.

Table 5. Goodness-of-fit measures and MLEs for non-composite models and composite models.

| Model | MLE and SE(MLE) | NLL | AIC | BIC |
| :---: | :---: | :---: | :---: | :---: |
| Exponential $X \sim \operatorname{Exp}(\lambda)$ | $\begin{gathered} \hat{\lambda}=0.0159, S E(\hat{\lambda})=\frac{\hat{\lambda}}{\sqrt{n}}= \\ 0.00332 \end{gathered}$ | 118.171 | 238.342 | 239.478 |
| Exp-Pareto | $\begin{gathered} \hat{\theta}=61.521, S E(\hat{\theta})=12.4734 \\ m=18 \end{gathered}$ | 126.291 | 254.582 | 255.717 |
| Inverse-gamma $X \sim \operatorname{IG}(\alpha, \beta)$ | $\begin{gathered} \hat{\alpha}=1.2633, S E(\hat{\alpha})= \\ \frac{1}{\sqrt{n \text { Polygamma }(1, \hat{\alpha})}}=0.19196 \\ \hat{\beta}=31.5436, S E(\hat{\beta})= \\ \frac{\hat{\beta}}{\sqrt{n \hat{\alpha}}}=5.85195 \end{gathered}$ | 117.129 | 238.258 | 248.529 |
| IG-Pareto | $\begin{gathered} \hat{\theta}=49.3097, S E(\hat{\theta})=\frac{\hat{\theta}}{\sqrt{n \alpha}}= \\ 18.5178 \\ m=13 \end{gathered}$ | 105.701 | 213.402 | 214.538 |

Note that for the $\operatorname{IG}(\alpha, \beta)$ distribution in Table 5, using the second derivatives of the log-likelihood function, we obtain

$$
\begin{gathered}
\frac{\partial^{2} \ln (L(\underline{x} \mid \alpha, \beta))}{\partial^{2} \alpha}=-n \text { polygamma }(1, \alpha) \\
\frac{\partial^{2} \ln (L(\underline{x} \mid \alpha, \beta))}{\partial^{2} \beta}=-\frac{n \alpha}{\beta^{2}}
\end{gathered}
$$

where polygamma $(1, \alpha)$ is the first derivative of the digamma function, which Mathematica can compute. These derivatives are used in Table 5 to compute the standard error of the MLEs.

### 5.2.2. Bayesian Inference of IG-Pareto

NLL, AIC, and BIC are criteria for evaluating models estimated by maximum likelihood methods. They may not be suitable for Bayesian model selections. Many researchers proposed the Bayesian approach. Ando (2010) introduced the Bayes factor, originally proposed by Kass and Raftery (1995), among other authors. The logic behind using Bayesian inference for the real data is based on the idea that it provides a more accurate estimator than the ML method, as verified by the simulation studies in Tables 1a-4a. The Bayesian estimator based on the mixture prior approach is more accurate than a non-Bayesian method, such as MLE, provided that a data-driven approach is instigated for the hyperparameters.

Bayes factor
The odds of the marginal likelihood of the data, $\underline{x}$, is given by

$$
B_{1,2}(\underline{\mathrm{x}})=\operatorname{Bayes} \operatorname{factor}\left(M_{1}, M_{2}\right)=\frac{P\left(\underline{\mathrm{x}} \mid M_{1}\right)}{P\left(\underline{\mathrm{x}} \mid M_{2}\right)}
$$

where $P\left(\underline{\mathbf{x}} \mid M_{1}\right), P\left(\underline{\mathbf{x}} \mid M_{2}\right)$ are marginal likelihoods of the dataset corresponding to two models: $M_{1}$ and $M_{2}$. If $B_{1,2}(\underline{x})>1$, it is concluded that $M_{1}$ is a better-fitted model than $M_{2}$. Ando (2010) states, "The Bayes Factor chooses the model with the largest value of marginal likelihood among a set of candidate models". The following is Jeffreys' scale of evidence for interpreting the Bayes factor:

1. If $B_{1,2}(\underline{\mathrm{x}})<1$, negative support for $M_{1}$
2. If $1<B_{1,2}(\underline{\mathbf{x}})<3$, barely worth mentioning $M_{1}$
3. If $3<B_{1,2}(\underline{\mathrm{x}})<10$, substantial evidence for $M_{1}$
4. If $10<B_{1,2}(\underline{x})<30$, strong evidence for $M_{1}$
5. If $30<B_{1,2}(\underline{x})<100$, very strong evidence for $M_{1}$
6. If $B_{1,2}(\underline{x})>100$, decisive evidence for $M_{1}$.

## The Marginal Likelihood

Let $x_{1}, x_{2}, \ldots, x_{n}$ be a random sample with the distribution $f(x \mid \theta)$. Then, the likelihood function is given by

$$
\prod_{i=1}^{n} f\left(x_{i} \mid \theta\right)
$$

Let $\pi(\theta)$ be the prior distribution for the parameter $\theta$, then the marginal likelihood function (PML) is defined by

$$
\operatorname{PML}(\underline{x} \mid \text { model })=\int_{0}^{\infty} \prod_{i=1}^{n} f\left(x_{i} \mid \theta\right) \pi(\theta) d \theta
$$

where $\theta>0$, which is the case for composite models, such as exp-Pareto and IG-Pareto considered in this article.

According to Table 5, as IG-Pareto is the best-fitted composite model, going forward, we will consider the mixture prior approach as discussed in the previous sections and apply it to the IG-Pareto composite model. From (20)

$$
L(\underline{\mathrm{x}} \mid \theta)=Q e^{-k \theta \sum_{i=1}^{m} \frac{1}{x_{i}}} \theta^{a(n-m)+\alpha m}
$$

where $Q=\frac{c^{n} k^{\alpha m} \prod_{i=1}^{m} x_{i}^{-\alpha-1} a^{n-m}}{\prod_{i=m+1}^{n} x_{i}^{a+1}}$ and $m$ is a positive integer, such as $x_{m} \leq \theta \leq x_{m+1}$. Let gamma $(\gamma, \delta)$ be the prior distribution with the pdf

$$
\pi(\theta)=\frac{\theta^{\gamma-1} e^{-\theta / \delta}}{\Gamma(\gamma) \delta^{\gamma}}, \gamma>0, \delta>0, \theta>0
$$

and $A=n a+m(\alpha-a)+\gamma, B=\frac{\delta_{i}}{\delta k \sum_{i=1}^{m}\left(\frac{1}{x_{i}}+1\right)}$. The marginal likelihood function (PML) is

$$
\begin{aligned}
& P M L(\underline{\mathrm{x}} \mid \delta, \gamma)=\int_{0}^{\infty} Q e^{-k \theta \sum_{i=1}^{m} \frac{1}{x_{i}}} \theta^{a(n-m)+\alpha m} \frac{\theta^{\gamma-1} e^{-\theta / \delta}}{\Gamma(\gamma) \delta^{\gamma}} d \theta \\
= & \frac{Q}{\Gamma(\gamma) \delta \gamma} \int_{0}^{\infty} e^{-\theta\left(k \sum_{i=1}^{m} \frac{1}{x_{i}}+\frac{1}{\delta}\right)} \theta^{a(n-m)+\alpha m+\gamma-1} d \theta=\frac{Q \Gamma(A) B^{A}}{\Gamma(\gamma) \delta^{\gamma}},
\end{aligned}
$$

where $\Gamma(A)$ denotes the gamma function evaluated at $A$. Now, consider the mixture distribution $\pi(\theta)$ of gamma priors $\pi\left(\gamma_{1}, \delta_{1}\right)$ and $\pi\left(\gamma_{2}, \delta_{2}\right)$, with equal weights $k_{1}=k_{2}=0.5$.

$$
\pi(\theta)=\frac{1}{2}\left(\frac{\theta^{\gamma_{1}-1} e^{-\theta / \delta_{1}}}{\Gamma\left(\gamma_{1}\right) \delta_{1}^{\gamma_{1}}}+\frac{\theta^{\gamma_{2}-1} e^{-\theta / \delta_{2}}}{\Gamma\left(\gamma_{2}\right) \delta_{2}^{\gamma_{2}}}\right), \gamma_{1}, \gamma_{2}>0, \delta_{1}, \delta_{2}>0, \theta>0
$$

and let $A_{h}=n a+m(\alpha-a)+\gamma_{h}, B_{h}=\frac{\delta_{h}}{\delta_{h} k \sum_{l=1}^{m}\left(\frac{1}{x_{i}}+1\right)}, h=1,2$. We can see that, given the mixture prior, the PML is represented as:

$$
P M L\left(\underline{\mathrm{x}} \mid \delta_{1}, \gamma_{1}, \delta_{2}, \gamma_{2}\right)=\frac{1}{2} Q\left(\frac{\Gamma\left(A_{1}\right) B_{1}^{A_{1}}}{\Gamma\left(\gamma_{1}\right) \delta_{1}^{\gamma_{1}}}+\frac{\Gamma\left(A_{2}\right) B_{2}^{A_{2}}}{\Gamma\left(\gamma_{2}\right) \delta_{2}^{\gamma_{2}}}\right) .
$$

As mentioned, selecting the hyperparameters $\gamma, \delta$ is challenging. The expected value and variance for $\pi(\gamma, \delta)$ are given as follows:

$$
E(\theta)=\gamma \delta \quad \operatorname{Var}(\theta)=\gamma \delta^{2}
$$

To find the optimal values of the hyperparameters $\gamma$ and $\delta$, we propose minimizing the variance $\gamma \delta^{2}$ under the constraint $E(\theta)=\theta$. Substituting $E(\theta)=\gamma \delta$ into the variance formula, we have $\operatorname{Var}(\theta)=\theta \delta$. Therefore, the smaller the $\delta$, the smaller the variance. The variance is an increasing function concerning the parameter $\delta$. Note that the coefficient of variation is $\frac{1}{\sqrt{\gamma}}$. Since we do not know the "real" value of $\theta$, as in the simulation section, we replace $\theta$ with its MLE $\hat{\theta}=49.3097$. Table 6 provides PML values for the selected models $\left(M_{1}-M_{4}\right)$ with $\gamma=10,20,30$, and 50 , and the corresponding $\delta=4.93097,2.46549,1.64366$, and 0.98619. It is clear that the smaller the variance, the better the model. For example, the Bayes factor of $M_{4}$ vs. $M_{1}$ is $\frac{3.90122 \times 10^{-56}}{3.28084 \times 10^{-56}}=1.1891 . M_{4}$ is about $118.91 \%$ times as likely as $M_{1}$.

Table 6 also provides PML values based on the mixture prior distribution when $K=2$ and weights $k_{1}=k_{2}=0.5$. The same Mathematica code as in the simulation section computes the hyperparameter values, assuming the "true" parameter value of theta is $\hat{\theta}=49.3097$.

Table 6. Bayesian estimates and marginal likelihood (PML) of IG-Pareto models with $(\mathrm{K}=2)$ or without ( $K=1$ ) mixture priors to the insurance losses due to floods in the USA.

| $\mathbf{K}$ | Model | Prior Distributions | Bayesian <br> Estimates | $P M L$ | $V^{2} R_{0.95}$ |
| :---: | :---: | :---: | :--- | :---: | :---: |
| 1 | $M_{1}$ | $\theta \sim \operatorname{gamma}(10,4.93097)$ | $\hat{\theta}=49.3097$ <br> $m=13$ | $3.28084 \times 10^{-56}$ | $5.18593 \times 10^{8}$ |
| 1 | $M_{2}$ | $\theta \sim \operatorname{gamma}(20,2.46549)$ | $\hat{\theta}=49.3097$ <br> $m=13$ | $3.63167 \times 10^{-56}$ | $5.24085 \times 10^{8}$ |
| 1 | $M_{3}$ | $\theta \sim \operatorname{gamma}(30,1.64366)$ | $\hat{\theta}=49.3097$ <br> $m=13$ | $3.77457 \times 10^{-56}$ | $5.26504 \times 10^{8}$ |
| 1 | $M_{4}$ | $\theta \sim \operatorname{gamma}(50,0.98619)$ | $\hat{\theta}=49.3097$ <br> $m=13$ | $3.90122 \times 10^{-56}$ | $5.28750 \times 10^{8}$ |
| 2 | $M_{5}$ | $\pi_{1}(\theta) \sim \operatorname{gamma}(25,2.11327)$ <br> $\pi_{2}(\theta) \sim \operatorname{gamma}(5,9.15752)$ | $\hat{\theta}=50.1982$ <br> $m=14$ | $1.64218 \times 10^{-55}$ | $5.28485 \times 10^{8}$ |
| 2 | $M_{6}$ | $\pi_{1}(\theta) \sim \operatorname{gamma}(27.5299,2)$ <br> $\pi_{2}(\theta) \sim \operatorname{gamma}(1.74239,25)$ | $\hat{\theta}=51.8988$ <br> $m=15$ | $7.19874 \times 10^{-55}$ | $5.4426 \times 10^{8}$ |

Recall the expected value and variance for $\theta$ under the assumptions $\gamma_{1} \neq \gamma_{2}, \quad \delta_{1} \neq \delta_{2}$, and $k_{1}=k_{2}=0.5$, are

$$
E(\theta)=0.5\left(\gamma_{1} \delta_{1}+\gamma_{2} \delta_{2}\right) \quad \operatorname{Var}(\theta)=0.25\left(\gamma_{1} \delta_{1}-\gamma_{2} \delta_{2}\right)^{2}+0.5 \gamma_{1} \delta_{1}^{2}+0.5 \gamma_{2} \delta_{2}^{2}
$$

Using the Mathematica code below, we find optimal $\delta_{1}, \delta_{2}$ for given $\gamma_{1}, \gamma_{2}$ (see Case 1 in Section 4.2).

$$
\text { NMinimize }\left[\left\{\operatorname{Var}(\theta), E[\theta]==49.3097, \delta_{1}>0, \delta_{2}>0\right\},\left\{\delta_{1}, \delta_{2}\right\}\right]
$$

Using the Mathematica code below, we find optimal $\gamma_{1}, \gamma_{2}$ for given $\delta_{1}, \delta_{2}$ (see Case 2 in Section 4.2).

$$
\text { NMinimize }\left[\left\{\operatorname{Var}(\theta), E[\theta]==49.3097, \gamma_{1}>0, \gamma_{2}>0\right\},\left\{\gamma_{1}, \gamma_{2}\right\}\right]
$$

Table 6 provides the PML values and the Bayesian estimates of the support parameter $\theta$. We note that the model based on the mixture prior outperforms the models without the mixture prior distribution. The model $M_{6}$-based equal-weight mixture of $\pi(27.5299,2)$ and $\pi(1.74239,25)$ provides the maximum PML value, which is $1.61188 \times 10^{-53}$. Model $M_{5}$, with an equal-weight mixture of $\pi(25,2.11327)$ and $\pi(5,9.15752)$, provides the second largest PML value, which is $7.19919 \times 10^{-55}$. For example, the Bayes factor of $M_{6}$ vs. $M_{4}$ is $\frac{7.19874 \times 10^{-55}}{3.90122 \times 10^{-56}}=18.4525$. Therefore $M_{6}$ is about $1845.25 \%$ times as likely as $M_{4}$. Based on Jeffreys' scale evidence, we have strong evidence for $M_{6}$ (that it fits the data).

Table 7 summarizes the Bayes factors for Models $M_{1}$ to $M_{6}$, which we discussed in Table 6. It can clearly be seen that M6 outperforms all other models due to Bayes factor $\left(M_{6}, M_{h}\right)>1$ for $h=1, \ldots, 5$, and M5 is the second best since Bayes factor $\left(M_{5}, M_{h}\right)>1$ for $h=1, \ldots, 4$. Hence, the model with the mixture prior outperforms the model without the mixture prior. The Bayes factor $\left(M_{k}, M_{j}\right)$ is denoted as $B_{k j}$ in Table 7.

Figure 3 presents the virtual presentation of the gamma priors and the gamma mixture priors to the IG-Pareto composite model in Table 6. Figure 3a shows gamma priors corresponding to $M_{1}$ to $M_{4}$. When the shape parameter $\gamma$ increases, the shapes of the curves become more symmetric and bell-shaped; Figure 3 b corresponds to $M_{5}$; the equal weight mixture of two gamma priors shows a bimodal shape; even one mode around 28 is
not significantly recognized. Figure 3c corresponds to $M_{6}$. The equal-weight mixture of two gamma priors clearly shows a bimodal shape.

Table 7. Bayes factors for paired models.

| Paired <br> Models | $\boldsymbol{B}_{\boldsymbol{k j}}$ | Paired <br> Models | $\boldsymbol{B}_{\boldsymbol{k} \boldsymbol{j}}$ | Paired <br> Models | $\boldsymbol{B}_{\boldsymbol{k} \boldsymbol{j}}$ |
| :---: | :---: | :---: | :---: | :---: | :---: |
| $M_{2}, M_{1}$ | 1.1069 | $M_{3}, M_{2}$ | 1.0393 | $M_{5}, M_{3}$ | 4.3506 |
| $M_{3}, M_{1}$ | 1.1505 | $M_{4}, M_{2}$ | 1.0742 | $M_{6}, M_{3}$ | 19.0717 |
| $M_{4}, M_{1}$ | 1.1891 | $M_{5}, M_{2}$ | 4.5218 | $M_{5}, M_{4}$ | 4.2094 |
| $M_{5}, M_{1}$ | 5.0054 | $M_{6}, M_{2}$ | 19.8221 | $M_{6}, M_{4}$ | 18.4525 |
| $M_{6}, M_{1}$ | 21.9418 | $M_{4}, M_{3}$ | 1.0336 | $M_{6}, M_{5}$ | 4.3837 |


(a)

(c)

Figure 3. The gamma distribution and the mixture gamma distribution of the prior distribution for the IG-Pareto composite models corresponding to Table 6. (a) Gamma distribution of the prior distribution for the IG-Pareto with different shapes and sizes. (b) The mixture gamma distribution of the prior distribution for the IG-Pareto with different shapes and sizes. (c) The mixture gamma distribution of the prior distribution for the IG-Pareto with different shapes and sizes.

Table 8 compares the models with the optimal mixture prior $\left(M_{5}, M_{6}\right)$ and without the optimal mixture prior $\left(M_{5}^{\prime}, M_{6}^{\prime}\right)$. Note that in the selection of hyperparameters for $M_{5}^{\prime}, M_{6}^{\prime}$, we do not minimize the variance. We only ensure the mean equation

$$
E(\theta)=\frac{1}{2}\left(\gamma_{1} \delta_{1}+\gamma_{2} \delta_{2}\right)
$$

is satisfied.
Since the "true" value of $\theta$ is unknown, as mentioned before, we use its MLE $\hat{\theta}=49.3097$. Therefore, for given $\gamma_{1}, \gamma_{2}$, and $\delta_{1}$, we have

$$
\delta_{2}=\frac{2 E(\theta)-\gamma_{1} \delta_{1}}{\gamma_{2}}=\frac{2 \times 49.3097-\gamma_{1} \delta_{1}}{\gamma_{2}}
$$

which leads to $M_{5}^{\prime}$. Also, for the given $\delta_{1}, \delta_{2}$, and $\gamma_{1}$, we have

$$
\gamma_{2}=\frac{2 E(\theta)-\gamma_{1} \delta_{1}}{\delta_{2}}=\frac{2 \times 49.3097-\gamma_{1} \delta_{1}}{\delta_{2}}
$$

which leads to $M_{6}^{\prime}$.
Table 8. Comparison between the optimal mixture prior and the non-optimal mixture prior for the insurance losses due to floods in the USA.

| Model | Prior Distributions | Bayesian <br> Estimates | PML | $\boldsymbol{B}_{M_{k} M_{k}^{\prime}}$ |
| :---: | :--- | :--- | :--- | :---: |
| $M_{5}^{\prime}$ | $\pi_{1}(\theta) \sim \operatorname{gamma}(25,2)$ <br> $\pi_{2}(\theta) \sim \operatorname{gamma}(5,9.27388)$ | $\hat{\theta}=49.4899$ <br> $m=13$ | $3.25458 \times 10^{-56}$ | 5.0458 |
|  | $\pi_{1}(\theta) \sim \operatorname{gamma}(26,2)$ <br> $\pi_{2}(\theta) \sim \operatorname{gamma}(1.86478,25)$ | $\hat{\theta}=50.5036$ <br> $m=14$ | $1.44635 \times 10^{-55}$ | 4.9772 |

Table 8 reveals that the models with the optimal mixture prior are better than those without the optimal mixture prior. For example, the Bayes factor $\left(M_{5}, M_{5}^{\prime}\right)=5.0458$. The model with the optimal mixture prior, $M_{5}$, is about $504.58 \%$ as likely as the model without the optimal mixture prior, $M_{5}^{\prime}$.

The value at risk, Klugman et al. (2012), is an important and standard risk measure in the insurance industry. $\mathrm{VaR}_{p}$ is the capital required at a higher probability $p$ to ensure the company will not go bankrupt.

$$
P\left(X \leq \operatorname{Var}_{p}\right)=p
$$

The $V a R_{p}$, or an upper limit prediction for a future value $y$, can be obtained via the predictive density $f(y \mid \underline{x})$. Aminzadeh and Deng (2019) provide the predictive density for the IG-Pareto composite model based on only one gamma prior distribution, as

$$
\begin{gather*}
f(y \mid \underline{\mathbf{x}})=\int_{0}^{\infty} f(\theta \mid \underline{\mathbf{x}}) f_{Y}(y \mid \theta) d \theta \\
=K_{1}(y)\left(1-H_{1}\left(y \mid \alpha+A, \frac{y B}{k B+y}\right)\right)+K_{2}(y) H_{2}(y \mid \alpha-k+A, B) \tag{27}
\end{gather*}
$$

where

$$
K_{1}(y)=\frac{k^{\alpha} \Gamma(\alpha+A) y^{A-1} B^{\alpha}}{(1+G R(\alpha, k)) \Gamma(\alpha)) \Gamma(A)(k B+y)^{\alpha+A}}
$$

and

$$
K_{2}(y)=\frac{(\alpha-k) B^{\alpha-k} \Gamma(\alpha-k+A)}{(1+G R(\alpha, k)) \Gamma(A) y^{\alpha-k+1}}
$$

$H_{1}$ denotes the cdf of gamma $\left(\alpha+A, \frac{y B}{k B+y}\right)$ and $H_{2}$ denotes the cdf of gamma $(\alpha-k+A, B)$, where

$$
A=n a+m(\alpha-a)+\gamma, \quad B=\frac{\delta}{\left(\delta k \sum_{i=1}^{m} \frac{1}{x_{i}}+1\right)}
$$

The above results can be extended to the mixture of two gamma prior distributions. Recall from Section 2.1,

$$
\begin{gather*}
f(y \mid \underline{x})=\sum_{j=1}^{K} \beta_{j} f^{j}(y \mid x)  \tag{28}\\
\beta_{j}=\frac{k_{j} f_{X}^{j}(x)}{\sum_{j=1}^{K} \int_{-\infty}^{\infty} k_{j} \pi_{X}^{j}(\theta \mid x) f_{X}^{j}(x) d \theta}=\frac{k_{j} f_{X}^{j}(x)}{\sum_{j=1}^{K} k_{j} f_{X}^{j}(x)} .
\end{gather*}
$$

Therefore, for the $K=2$ case, based on (27), we have

$$
f^{j}(y \mid \underline{x})=K_{1}^{j}(y)\left(1-H_{1}\left(y \mid \alpha+A_{j}, \frac{y B_{j}}{k B_{j}+y}\right)\right)+K_{2}^{j}(y) H_{2}\left(y \mid \alpha-k+A_{j}, B_{j}\right), \quad j=1,2
$$

and $K_{1}^{j}(y), K_{2}^{j}(y)$ can be defined via (27) using the corresponding $A_{j}, B_{j}$.
The last column of Table 6 for models $M_{1}$ to $M_{6}$ provides $V_{0.95}$ values, which are found using the predictive density (28),

$$
P\left(Y \leq \operatorname{VaR}_{0.95} \mid \underline{x}\right)=0.95
$$

and Mathematica. The values tell us how much the company should reserve under each model to avoid bankruptcy at the $95 \%$ confidence level. For example, if we use model $M_{6}$, we should reserve USD $5.4426 \times 10^{8}$ million to avoid bankruptcy at the $95 \%$ confidence level.

## 6. Conclusions

This article considers the class of composite models. We are interested in exploring the Bayesian estimates of the threshold parameter $\theta$, which separates the small losses with high frequencies and significant losses with low frequencies. Two composite models, exp-Pareto and IG-Pareto, are considered as examples. The prior distribution for parameter $\theta$ uses a mixture of prior distributions. We verify that, in general, the posterior distribution is a mixture of individual posterior distributions. For each composite model considered in the article, the general formula of the Bayes estimator of $\theta$ is derived based on the mean squared error loss function.

Simulation results compare the accuracies of $\hat{\theta}_{\text {Bayes }}$ using with and without mixture prior distributions. Also, the accuracy of $\hat{\theta}_{M L E}$ is compared to the Bayes estimates. For both exp-Pareto and IG-Pareto models, respectively, methods for choosing the optimal hyperparameter $\left(a_{i}, b_{i}\right),\left(\gamma_{i}, \delta_{i}, i=1,2, \ldots K\right)$ values, are proposed. The proposed method is data-driven, as it uses the MLE of $\theta$ based on real data to compute optimal values for hyperparameters. Simulations reveal that the Bayesian estimator with the mixture prior distribution is more accurate than the Bayesian estimator without the mixture prior distribution. Also, both Bayes estimators are more accurate than MLE.

For an illustration of computations involved in the proposed methods, the insurance losses in the USA from 2000 to 2019 due to natural event floods are considered and downloaded from EM-DAT, the International Disaster Database. In order to eliminate the effect of inflation, all insurance damage amounts are converted to 2019 dollars. Based on NLL, AIC, and BIC measures, the conclusion is that IG-Pareto provides the best fit, which leads us to apply the Bayesian method to the IG-Pareto composite model. We have shown that the IG-Pareto model with the mixture gamma prior distribution most optimally fits the data based on the optimal hyperparameter value. For the comparison of Bayesian models, the Bayes factor is used.

Potential future research would involve extending the mixture prior approach to other composite distributions, such as log-normal-Pareto and Rayleigh-Pareto. Furthermore, the mixture prior approach can be investigated for composite models that involve more than two distributions. For example, consider Pareto for the right tail of data with very large losses, a non-heavy tail distribution for small losses in the data, and another distribution that models moderate losses in the center of data.

Author Contributions: Conceptualization, M.D. and M.S.A.; Methodology, M.D. and M.S.A.; Software, M.D. and M.S.A.; Validation, M.S.A.; Formal analysis, M.D. and M.S.A.; Investigations, M.D. and M.S.A.; Resources, M.D. and M.S.A.; Data curation, M.D.; writing-original draft preparation, M.D. and M.S.A.; writing-review and editing, M.D. and M.S.A.; visualization, M.D. and M.S.A. All authors have read and agreed to the published version of the manuscript.
Funding: This research received no external funding.

Data Availability Statement: Not applicable.
Acknowledgments: The authors are grateful for the invaluable time and suggestions of the editors and reviewers to enhance the presentation of the article.

Conflicts of Interest: The authors declare no conflict of interest.

## References

Abdul Majid, Muhammad Hilmi, and Kamarulzaman Ibrahim. 2021a. Composite Pareto Distributions for Modeling Household Income Distribution in Malaysia. Sains Malaysiana 50: 2047-58. [CrossRef]
Abdul Majid, Muhammad Hilmi, and Kamarulzaman Ibrahim. 2021b. On Bayesian approach to composite Pareto models. PLoS ONE 16: e0257762. [CrossRef] [PubMed]
Aminzadeh, Mostafa S., and Min Deng. 2017. Bayesian Predictive Modeling for Exponential-Pareto Composite Distribution. Variance 12: 59-68.
Aminzadeh, Mostafa S., and Min Deng. 2019. Bayesian Predictive Modeling for Inverse Gamma-Pareto Composite Distribution. Communications In Statistics, Theory, and Methods 48: 1938-54. [CrossRef]
Ando, Tomohiro. 2010. Bayesian Model Selection and Statistical Modeling. Orange: Chapman \& Hall/CRC.
Bakar, S. A. Abu, Nor A. Hamzah, Mastoureh Maghsoudi, and Saralees Nadarajah. 2015. Modeling loss data using composite models. Insurance: Mathematics and Economics 61: 146-54.
Bhati, Deepesh, Enrique Calderín-Ojeda, and Mareeswaran Meenakshi. 2019. A new heavy-tailed class of distributions which includes the Pareto. Risks 7: 99. [CrossRef]
Cooray, Kahadawala, and Chin-I. Cheng. 2013. Bayesian Estimators of the Lognormal-Pareto Composite Distribution. Scandinavian Actuarial Journal 2015: 500-15. [CrossRef]
Deng, Min, and Mostafa S. Aminzadeh. 2019. Bayesian predictive analysis for Weibull-Pareto composite model with an application to insurance data. Communications in Statistics-Simulation and Computation 51: 2683-709. [CrossRef]
Deng, Min, Mostafa S. Aminzadeh, and Min Ji. 2021. Bayesian Predictive Analysis of Natural Disaster Losses. Risks 9: 12. [CrossRef]
Dominicy, Yves, and Corinne Sinner. 2017. Distributions and composite models for size-type data. Advances in Statistical Methodologies and Their Application to Real Problems 159. [CrossRef]
Kass, Robert E., and Adrian E. Raftery. 1995. Bayes factors. Journal of the American Statistical Association 90: 773-95. [CrossRef]
Klugman, Stuart A., Harry H. Panjer, and Gordon E. Willmot. 2012. Loss Models from Data to Decisions, 3rd ed. New York: John Wiley.
Miljkovic, Tatjana, and Bettina Grün. 2016. Modeling loss data using mixtures of distributions. Insurance Mathematics, and Economics 70: 387-96. [CrossRef]
Preda, Vasile, and Roxana Ciumara. 2006. On Composite Models: Weibull-Pareto and Lognormal-Pareto-A comparative Study. Romanian Journal of Economic Forecasting 8: 32-46.
Rufo, María Jesús, Carlos. J. Pérez, and Jacinto Martín. 2010. Merging experts' opinions: A Bayesian hierarchical model with a mixture of prior distributions. European Journal of Operational Research 207: 284-89. [CrossRef]
Saleem, Muhammad. 2010. Bayesian Analysis of Mixture Distributions. Ph.D. thesis, Quaid-i-Azam University Islamabad, Islamabad, Pakistan. Available online: http:/ / prr.hec.gov.pk/jspui/bitstream/123456789/1430/1/824S.pdf (accessed on 1 August 2023).
Scollnik, David P. M., and Chenchen Sun. 2012. Modeling with Weibull-Pareto Models. North American Actuarial Journal 16: 260-72. [CrossRef]
Teodorescu, Sandra, and Raluca Vernic. 2006. A composite Exponential-Pareto distribution. The Annals of the "Ovidius" University of Constanta, Mathematics Series 14: 99-108.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

