Model Construction and Predictive Analytics of Insurance Data

Min Deng and M.S.Aminzadeh

Towson University, Maryland, USA

Outlines:

1. Introduction
2. Composite Models
3. MLE and Bayes estimates of parameters
4. Predictive Models
5. Risk Measures
6. Simulation
7. Model Selection
8. Summary

Introduction

- Modeling loss data is one of the most important topics in actuarial science.

Introduction

- Modeling loss data is one of the most important topics in actuarial science.
- Actuaries must predict the future losses based on models in order to provide marketing opportunity and financial risk management.

Introduction

- Modeling loss data is one of the most important topics in actuarial science.
- Actuaries must predict the future losses based on models in order to provide marketing opportunity and financial risk management.
- The choice of claims distribution is an important and challenging task.

Introduction

- Insurance data is very different from the data in other industry,

Introduction

- Insurance data is very different from the data in other industry,
- In most cases, we have high frequencies of small claims and very few large claims with low frequencies.

Introduction

- Insurance data is very different from the data in other industry,
- In most cases, we have high frequencies of small claims and very few large claims with low frequencies.
- Traditional distributions such as normal, exponential, inverse-gamma, etc. are not able to describe the characteristics of insurance data which are both skewed and fat-tailed.

Introduction

- Insurance data is very different from the data in other industry,
- In most cases, we have high frequencies of small claims and very few large claims with low frequencies.
- Traditional distributions such as normal, exponential, inverse-gamma, etc. are not able to describe the characteristics of insurance data which are both skewed and fat-tailed.
- Central Limit Theorem is not very useful for the insurance industry

Introduction

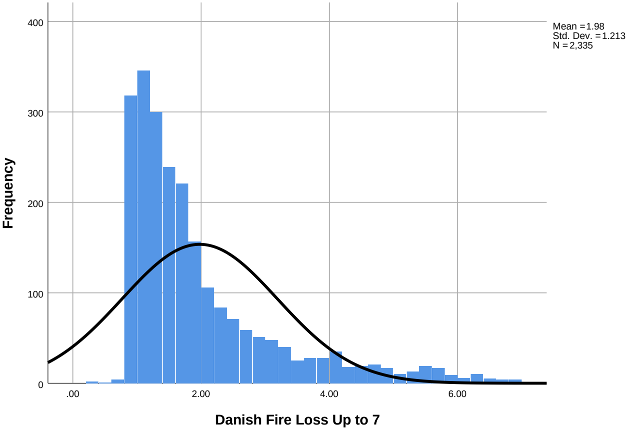

Figure 1: Danish fire insurance loss data has 2492 losses in the period 1980-1990. The data has been adjusted as 1985 values. The adjusted losses range from 0.3134041 to 263.250366 (in millions).

Introduction

- Teodorescu and Vernic (2006) introduced the composite Exponential-Pareto distribution
- Preda and Ciumara (2006) introduced the Weibull-Pareto and Lognormal-Pareto composite models
- Cooray and Cheng (2015) used Bayesian methods for estimating parameters of the lognormal-Pareto
- Scollink and Sun (2012) developed several composite Weibull-Pareto models

Introduction

- Aminzadeh and Deng (2018) developed the composite Inverse Gamma-Pareto model
- Aminzadeh and Deng (2018) used Bayesian methods for estimating parameters of the Exponential -Pareto Composite, Inverse Gamma-Pareto, and Weibull-Pareto models
- Aminzadeh and Deng (2018) also developed the predictive models for the Exponential -Pareto Composite, Gamma-Pareto, and Weibull-Pareto Composite.
- predictive model is the predictive distribution of the next loss based on the past losses.

Composite Models

- Let X be a non-negative real-valued random variable. The general form of a composite pdf involving two distributions with pdfs $f_1(x)$ and $f_2(x)$ is given by the probability density function,

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

where θ is the parameter that represents the boundary of the supports for the two distributions.

Composite Models

- Let X be a non-negative real-valued random variable. The general form of a composite pdf involving two distributions with pdfs $f_1(x)$ and $f_2(x)$ is given by the probability density function,

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

  where θ is the parameter that represents the boundary of the supports for the two distributions.

- In order to make the composite density function smooth, it is assumed that the pdf $f_X(x)$ is continuous and differentiable at θ. That is,

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

Composite Models

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

Composite Models

Exponential-Pareto

- Let

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

and

$$f_2(x) = \frac{\alpha\theta^{\alpha}}{x^{\alpha+1}}, \quad x \geq \theta > 0, \alpha > 0$$

Composite Models

Exponential-Pareto

- Let

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

and

$$f_2(x) = \frac{\alpha\theta^{\alpha}}{x^{\alpha+1}}, \ \ x \geq \theta > 0, \alpha > 0$$

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

Composite Models

Exponential-Pareto

- Let

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

  and

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

Composite Models
Exponential-Pareto

- Let
$$f_1(x) = \lambda e^{-\lambda x}, \quad x > 0, \lambda > 0$$
and
$$f_2(x) = \frac{\alpha\theta^\alpha}{x^{\alpha+1}}, \quad x \geq \theta$$
- $f_1(x)$ is the pdf of Exponential distribution with parameter λ
- $f_2(x)$ is the pdf of Pareto distribution with parameters θ and α.
- This is a three parameters distribution.

- Let

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

  and

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

- $f_1(x)$ is the pdf of Exponential distribution with parameter λ
- $f_2(x)$ is the pdf of Pareto distribution with parameters θ and α.
- This is a three parameters distribution.
- to make the composite density function smooth, we solve the following equations simultaneously,

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

- the solutions are

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

Composite Models

Exponential-Pareto

- The initial three parameters are reduced to only one parameter θ.
- The pdf of composite Exponential-Pareto distribution is given by
- $$f_X(x|\theta) = \begin{cases} \frac{.775}{\theta} e^{\frac{-1.35x}{\theta}} & 0 < x \leq \theta \\ \frac{.2\theta^{.35}}{x^{1.35}} & \theta \leq x < \infty \end{cases}$$

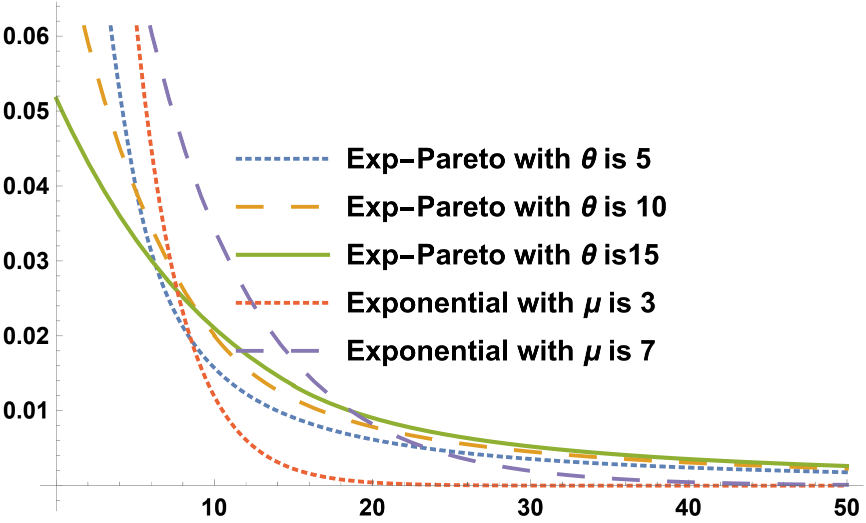

Figure 2:Comparison of Exponential-Pareto Composite Model with different θ (the boundary between the large losses and smaller losses) and Exponential distribution with different mean μ

Composite Models

Inverse Gamma-Pareto

- Let
$$f_1(x) = \frac{\beta^\alpha x^{-\alpha-1} e^{-\beta/x}}{\Gamma(\alpha)}, \quad x > 0, \alpha > 0, \beta > 0$$
- and
$$f_2(x) = \frac{a\theta^a}{x^{a+1}}, \quad x \geq \theta, a > 0, \theta > 0$$
This is initially four parameters model involved in Inverse- Gamma and Pareto distributions which is given by
-
$$f_X(x) = \begin{cases} c\frac{\beta^\alpha x^{-\alpha-1} e^{-\beta/x}}{\Gamma(\alpha)}, \quad \alpha > 0, \beta > 0 \quad 0 < x \leq \theta \\ c\frac{a\theta^a}{x^{a+1}}, \quad a > 0, \theta \leq x < \infty \end{cases}$$

- Applying smoothing conditions, the number of parameters of the composite Inverse Gamma-Pareto distribution is reduced to only one parameter θ. The pdf is given by
- $$f_X(x|\theta) = \begin{cases} \frac{c(k\theta)^{\alpha} x^{-\alpha-1} e^{\frac{-k\theta}{x}}}{\Gamma(\alpha)}, & 0 < x \leq \theta \\ \frac{c(\alpha-k)\theta^{\alpha-k}}{x^{\alpha-k+1}}, & \theta \leq x < \infty \end{cases}$$
- where $\alpha = 0.308289$, $k = 0.144351, a = \alpha - k = 0.163947$, and $c = 0.711384$.

Composite Models

Inverse Gamma-Pareto

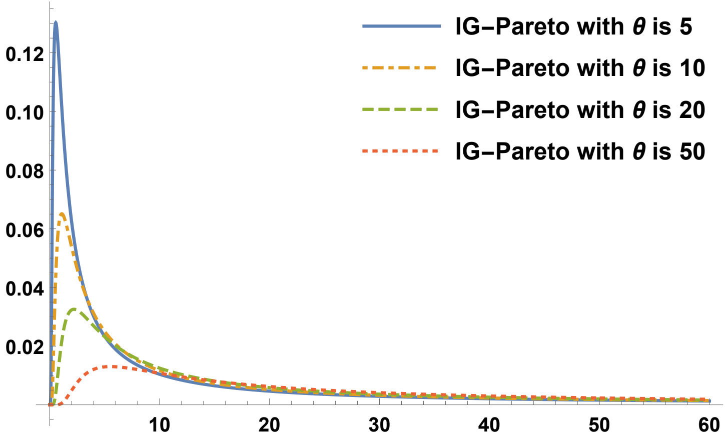

Figure 3: Comparison of Inverse Gamma-Pareto Composite Model with different θ (the boundary between the large losses and smaller losses)

Composite Models

Inverse Gamma-Pareto

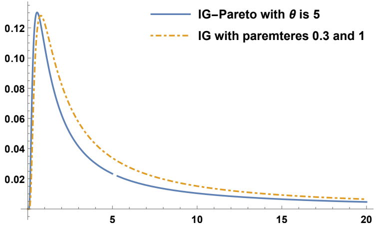

Figure: Models fit Small Losses part

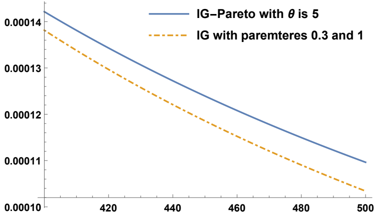

Figure: Models fit Large Losses part

Figure 4 and 5 Compare of Inverse Gamma-Pareto Composite Model with $\theta = 5$ and Inverse Gamma with parameters $\alpha = 0.3$ and $\beta = 1$

Composite Models

Weibull-Pareto

- Let
$$f_1(x) = \frac{\beta}{\gamma^\beta} x^{\beta-1} e^{-(\frac{x}{\gamma})^\beta} \quad x > 0, \gamma > 0, \beta > 0$$
- and
$$f_2(x) = \frac{a\theta^a}{x^{a+1}}, \quad x \geq \theta, a > 0, \theta > 0$$
The weibull-Pareto composite model has initially four parameters. The pdf is given by
-
$$f_X(x) = \begin{cases} c\frac{\beta}{\gamma^\beta} x^{\beta-1} e^{-(\frac{x}{\gamma})^\beta}, \quad \gamma > 0, \beta > 0 & 0 < x \leq \theta \\ c\frac{a\theta^a}{x^{a+1}}, \quad a > 0 & \theta \leq x < \infty \end{cases}$$

Composite Models

Weibull-Pareto

- Applying smoothing conditions, the number of parameters of the composite Weibull-Pareto distribution is reduced to only two parameters α and θ. The pdf is given by
- $$f_X(x|\theta) = \begin{cases} \frac{a\alpha}{x}\left(\frac{x}{\theta}\right)^{\alpha k} e^{-c\left(\frac{x}{\theta}\right)^{\alpha k}}, & 0 < x \leq \theta \\ b\left(\frac{\alpha}{x}\right)\left(\frac{\theta}{x}\right)^{\alpha} & \theta \leq x < \infty \end{cases}$$
- where
$$a = \frac{(k+1)^2}{2k+1}, b = \frac{k+1}{2k+1}, c = \frac{k+1}{k}, k = 2.8573348.$$

MLE and Bayes estimates of parameters

Maximum Likelihood Estimator (MLE)

- Let $x_1, ..., x_n$ be a random sample for the composite pdf and without loss of generality assume that $x_1 < x_2 < ... < x_n$ is an ordered random sample from the composite models
- the MLE of θ for Exponential-Pareto is

$$\hat{\theta}_{MLE} = \frac{1.35 \sum_{i=1}^{m} x_i}{1.35m - .35n}.$$

MLE and Bayes estimates of parameters

Maximum Likelihood Estimator (MLE)

- Let $x_1, ..., x_n$ be a random sample for the composite pdf and without loss of generality assume that $x_1 < x_2 < ... < x_n$ is an ordered random sample from the composite models
- the MLE of θ for Exponential-Pareto is

$$\hat{\theta}_{MLE} = \frac{1.35 \sum_{i=1}^{m} x_i}{1.35m - .35n}.$$

- The MLE of θ for Inverse Gamma-Pareto is

$$\hat{\theta}_{MLE} = \frac{m\alpha + (\alpha - k)m}{k \sum_{i=1}^{m} x_i^{-1}}.$$

- where $\alpha = 0.308289$ and $k = 0.144351$
- A program finds m by iteration such that $x_m \leq \theta \leq x_{m+1}$.

MLE and Bayes estimates of parameters
Maximum Likelihood Estimator (MLE)

- Closed formulas for the MLEs of α and θ for Weibull-Pareto model cannot be found.
- A search algorithm is used in the Mathematica code to find the global maximum for the likelihood function for each value of m and obtain the corresponding optimal values for θ and α.
- the likelihood function of Weibull-Pareto is

$$L(\underline{\mathrm{x}}|\alpha,\theta) = m\ln(a)+m\ln(\alpha)+\alpha k S_1 - m\alpha k\ln(\theta) - cS3 + (n-m)(\ln(b)+\ln(\alpha)+\alpha\ln(\theta)) - S_2(\ln(\alpha)+1)$$

- where

$$S_1 = \sum_{i=1}^{m} \ln(x_i) \quad S_2 = \sum_{i=m+1}^{n} \ln(x_i) \quad S_3 = \sum_{i=1}^{m} (\frac{x_i}{\theta})^{\alpha k}.$$

$$a = \frac{(k+1)^2}{2k+1}, b = \frac{k+1}{2k+1}, c = \frac{k+1}{k}, k = 2.8573348.$$

MLE and Bayes estimates of parameters

Maximum Likelihood Estimator (MLE)

- A search algorithm is as follows,
- For a given sample from the composite model, get sorted sample observations $x_1 \leq x_2 \leq ... \leq x_n$.
- 1. Start with m=1, optimize the likelihood function $L(\underline{x}|\alpha, \theta)$ with respect to θ and α. If $x_1 \leq \hat{\theta}_{MLE} \leq x_2$,

  then $m = 1$, otherwise goto step 2.
- 2. let m=2, optimize the likelihood function $L(\underline{x}|\alpha, \theta)$ with respect to θ and α. If $x_2 \leq \hat{\theta}_{MLE} \leq x_3$, then

  $m = 2$, otherwise goto step 3.
- 3 Repeat the process until the correct value for m is identified. Using the correct value of m the

  corresponding estimates for θ and α are selected that represent the MLEs.

- Let Inverse-Gamma distribution be the conjugate prior for θ with the pdf

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

- Then, the posterior pdf can be written as

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

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

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

- A program finds m by iteration such that $x_m \leq \theta \leq x_{m+1}$. a, b are hyper-parameters of the prior distribution.

- Assume Gamma(γ, δ) as a prior distribution for θ with the pdf

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

- Then, the posterior pdf can be written as

$$f(\theta|\underline{\mathrm{x}}) = \frac{L(\underline{\mathrm{x}}|\theta) * \rho(\theta)}{\int L(\underline{\mathrm{x}}|\theta) * \rho(\theta) d\theta} \propto e^{-\theta(k\sum_{i=1}^{m}\frac{1}{x_i}+\frac{1}{\delta})}\theta^{na+m(\alpha-a)+\gamma-1}.$$

- The expression on the right side is the kernel of Gamma pdf with parameter (A, B), where $A = na + m(\alpha - a) + \gamma$ and $B = \frac{\delta}{(\delta k \sum_{i=1}^{m} \frac{1}{x_i} + 1)}$.
- Therefore under squared error loss function, the informative Bayes estimator for θ is

$$\hat{\theta}_{Bayes_1} = E[\theta|\underline{\mathrm{x}}] = AB = \frac{\delta(na + mk + \gamma)}{(\delta k \sum_{i=1}^{m} \frac{1}{x_i} + 1)}.$$

- A program finds m by iteration such that $x_m \leq \theta \leq x_{m+1}$. Where $k = 0.144351, a = 0.163947$, and γ, δ are hyper-parameters of the prior distribution.

- Let y be a realization of the random variable Y from the composite models. The pdf is denoted by $f_Y(y|\theta)$ for E-P and IG-P models and $f_Y(y|\theta,\alpha)$ for W-P model.
- Based on the observed sample data $\underline{x}$, we are interested in deriving the predictive density $f(y|\underline{x})$.
- In the context of Bayesian framework, predictive density is used to estimate measures such as $E[Y|\underline{x}]$ or Var$[Y|\underline{x}]$ or other risk measures such as VaR, TVaR etc.
- $f(\theta|\underline{x})$ denotes the posterior distribution for E-P and IG-P models and $f(\theta,\alpha|\underline{x})$ for W-P model.
- For E-P and IG-P, the predictive density of $y|x$ is obtained through

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

  and
- For W-P, the predictive density of $y|x$ is obtained through

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

- For E-P model, the model pdf is

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

- The posterior pdf is Inverse-Gamma(A, B), where $A = (a - .35n + 1.35m)$ and $B = (b + 1.35 \sum_{i=1}^{m} x_i)$.
- as a result we get

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

Predictive Models

Predictive Density of E-P Model

- The predictive density $f(y|\underline{x})$ is given by

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

$$= K_2(y) H_2(y|A - .35, B) + K_1(y)(1 - H_1(y|A + 1, B + 1.35y)$$

- where

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

- and H_2 is the cdf of Inverse-gamma distribution with parameters $(A - .35, B)$ and H_1 is the cdf of inverse-gamma distribution with parameters $(A + 1, B + 1.35y)$.
- Similar to the composite density for which $E[X]$ is undefined due to $\alpha = 0.35$, $E[Y|\underline{x}]$ and $TVaR_p$ are also undefined for the predictive pdf.

Predictive Models

Predictive Density of E-P Model

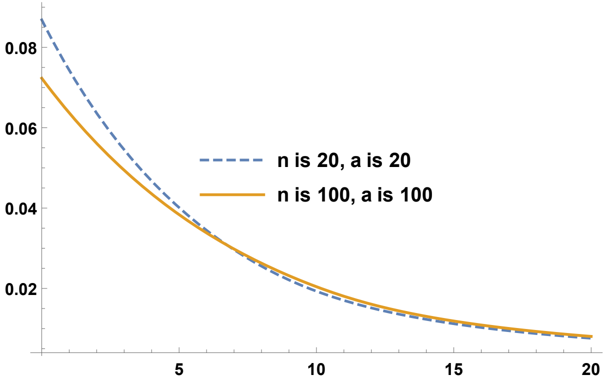

Figure 11: Predictive density for selected values of n (sample size), a (hyper-parameter of prior distribution), and $\theta = 10$. Graphs are based on generated samples from composite E-P pdf.

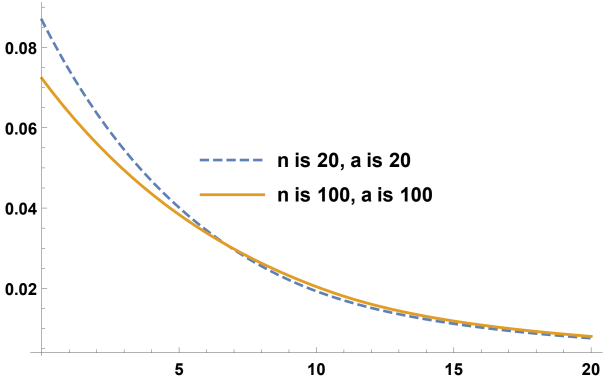

Figure 11: Predictive density for selected values of n (sample size), a (hyper-parameter of prior distribution), and $\theta = 10$. Graphs are based on generated samples from composite E-P pdf.

- Figures 11 reveals 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.

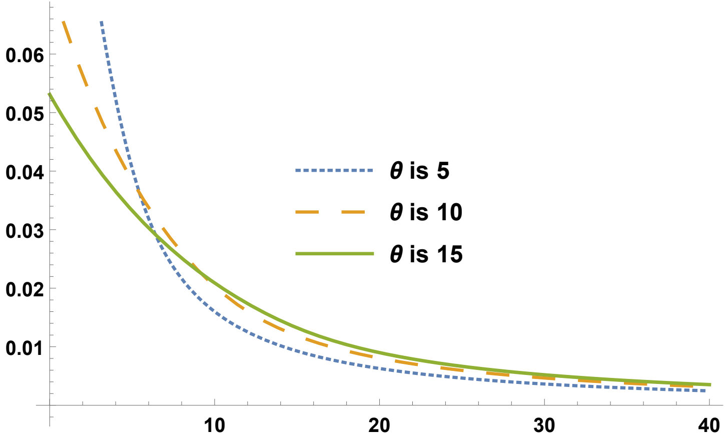

Figure 12: Predictive density with various θ values, n = 50 and a = 50. Graphs are based on generated samples from composite E-P pdf.

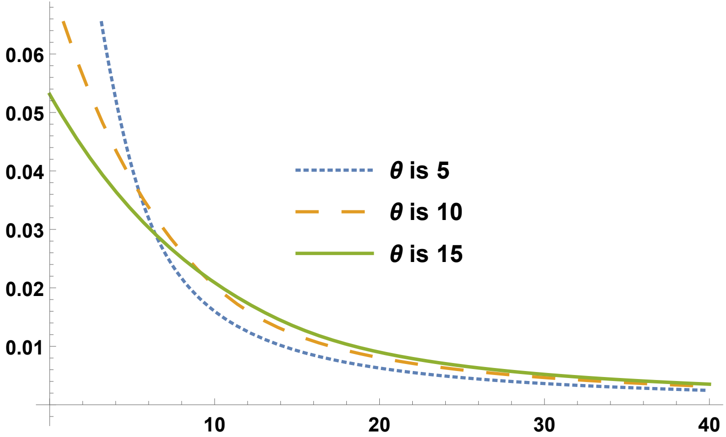

Figure 12: Predictive density with various θ values, n = 50 and a = 50. Graphs are based on generated samples from composite E-P pdf.

- Figure 12 shows, as θ increases, the tail of the predictive density becomes heavier, causing VaR at a specific level, say 0.99, to increase.

Predictive Models

Predictive Density of IG-P Model

- For IG-P model, the model pdf is

$$f_Y(y|\theta) = \begin{cases} \frac{c(k\theta)^{\alpha} y^{-\alpha-1} e^{\frac{-k\theta}{y}}}{\Gamma(\alpha)}, & 0 < y \leq \theta \\ \frac{c(\alpha-k)\theta^{\alpha-k}}{y^{\alpha-k+1}}, & \theta \leq y < \infty \end{cases}$$

- The posterior pdf is Gamma Distribution with parameter (A, B), where $A = na + m(\alpha - a) + \gamma$ and $B = \frac{\delta}{(\delta k \sum_{i=1}^{m} \frac{1}{x_i} + 1)}$.
- as a result we get

$$f(\theta|\underline{\mathrm{x}}) f(y|\theta) = \begin{cases} \frac{c(k\theta)^{\alpha} y^{-\alpha-1} e^{\frac{-k\theta}{y}}}{\Gamma(\alpha)} \frac{\theta^{A-1} e^{-\theta/B}}{\Gamma(A) B^{A}}, & 0 < y \leq \theta \\ \frac{c(\alpha-k)\theta^{\alpha-k}}{y^{\alpha-k+1}} \frac{\theta^{A-1} e^{-\theta/B}}{\Gamma(A) B^{A}}, & \theta \leq y < \infty \end{cases}$$

- The predictive density $f(y|\underline{\mathrm{x}})$ is given by

$$f(y|\underline{\mathrm{x}}) = \int_y^\infty \frac{ck^\alpha y^{-\alpha-1}\theta^{\alpha+A-1}e^{-\theta(\frac{k}{y}+\frac{1}{B})}}{\Gamma(\alpha))\Gamma(A)B^A}d\theta + \int_0^y \frac{c(\alpha-k)\theta^{\alpha-k+A-1}e^{-\frac{\theta}{B}}}{y^{\alpha-k+1}\Gamma(A)B^A}d\theta$$

$$= K_1(y)(1 - H_1(y|\alpha + A, \frac{yB}{kB+y})) + K_2(y)H_2(y|\alpha - k + A, B)$$

- where

$$K_1(y) = \frac{k^\alpha\Gamma(\alpha+A)y^{A-1}B^\alpha}{(1+GR(\alpha,k))\Gamma(\alpha))\Gamma(A)(kB+y)^{\alpha+A}}$$

  and

$$K_2(y) = \frac{(\alpha-k)B^{\alpha-k}\Gamma(\alpha-k+A)}{(1+GR(\alpha,k))\Gamma(A)y^{\alpha-k+1}}$$

- H_1 is the cdf of gamma distribution with parameters $(\alpha + A, \frac{yB}{kB+y})$ and H_2 is the cdf of gamma distribution with parameters $(\alpha - k + A, B)$.
- Similar to the IG-P composite model for which $E[X]$ is undefined, $E[Y|\underline{\mathrm{x}}]$ is also undefined for the predictive pdf.

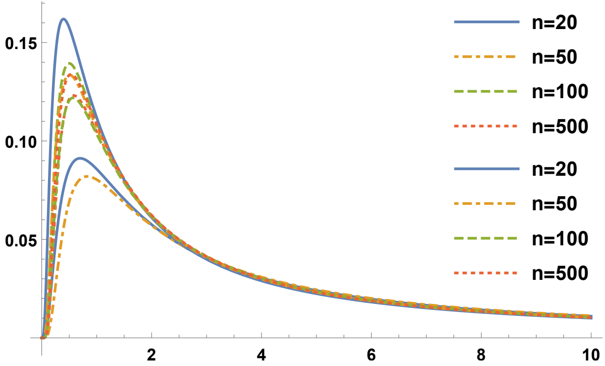

Figure 13: The predictive density curves ($\theta = 5$,$\gamma = 2$,$\delta = 2.5$) based on two generated samples for $n = 500$ (dotted line),$n = 100$ (dashed line),$n = 50$(dot-dashed line), and $n = 20$ (solid line).

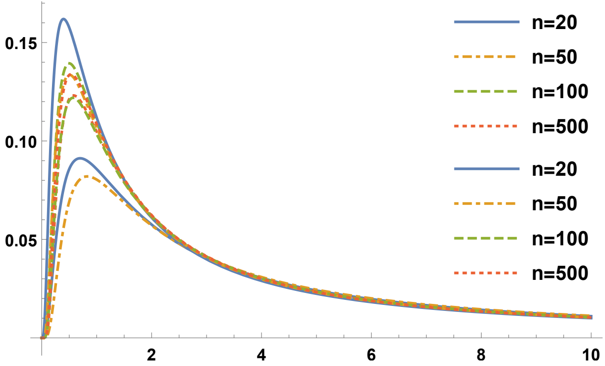

Figure 13: The predictive density curves ($\theta = 5$,$\gamma = 2$,$\delta = 2.5$) based on two generated samples for $n = 500$ (dotted line),$n = 100$ (dashed line),$n = 50$(dot-dashed line), and $n = 20$ (solid line).

- Figure 13 provides graphs of the predictive pdf based on only two samples $\underline{x}$ generated from the composite model IG-P. It is noted that as sample size increases, discrepancies between graphs of predictive pdf even based on two samples decrease.The implication is that for larger n, more accurate estimates for risk measures can be obtained.

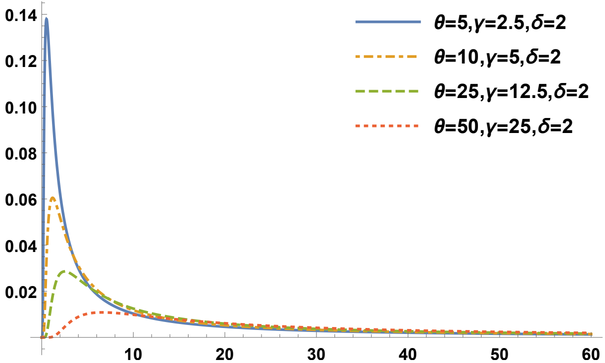

Figure 14: The predictive density curves (n=1000) for $\theta = 50$, $\gamma = 25$, $\delta = 2$ (dotted line),$\theta = 25$, $\gamma = 12.5$, $\delta = 2$ (dashed line), $\theta = 10$, $\gamma = 5$, $\delta = 2$(dot-dashed line), and $\theta = 5$, $\gamma = 2.5$, $\delta = 2$ (solid line).

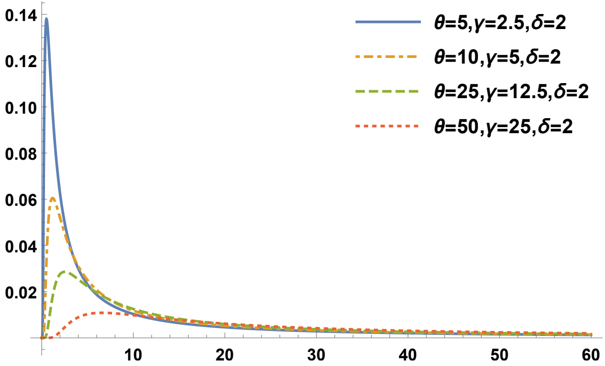

Figure 14: The predictive density curves (n=1000) for $\theta = 50$, $\gamma = 25$, $\delta = 2$ (dotted line),$\theta = 25$, $\gamma = 12.5$, $\delta = 2$ (dashed line), $\theta = 10$, $\gamma = 5$, $\delta = 2$(dot-dashed line), and $\theta = 5$, $\gamma = 2.5$, $\delta = 2$ (solid line).

- Figure 14 provides graphs of the predictive pdf for different values of θ. The hyper-parameter (γ, δ) values are selected in such a way that $\gamma\delta = \theta$. As we can see, for larger θ the tail of the predictive pdf is heavier.

- A popular insurance phenomenon is the maximum benefit. The policy limit reduces the insurance company's future risk and also at the same time it reduces the cost of the insurance policy, because the pure premium is defined as the expectation of the loss.
- Let the policy limit be denoted by b and X be the loss random variable with the pdf $f(x)$. Then limited loss random $X \wedge b$ is defined as

$$X \wedge b = \begin{cases} X & 0 < x \leq b \\ b & b \leq x < \infty \end{cases}$$

- The k^{th} moment of the limited loss is called the k^{th} limited moment (LM) and is defined as

$$E[(X \wedge b)^k] = \int_{-\infty}^{b} x^k f(x)dx + \int_{b}^{\infty} b^k f(x)dx.$$

- The most important cases are $k = 1$ and $k = 2$ which provide respectively, the limited expectation (LE) or pure premium and the limited variance (LV).

$$\text{LE}=E[(X \wedge b)] = \int_{-\infty}^{b} x f(x) dx + \int_{b}^{\infty} b f(x) dx$$

$$\text{LV}=E[(X \wedge b)^2] - (E[(X \wedge b)])^2.$$

- As mentioned earlier, $E[Y|\underline{x}]$ is not defined for the predictive density for both E-P and IG-P models. However in practice one would be interested in estimating k^{th} limited moment through a Bayesian method.

$$E[(Y \wedge b)^k|\underline{x}] = \int_{-\infty}^{b} y^k f(y|\underline{\mathrm{x}}) dy + \int_{b}^{\infty} b^k f(y|\underline{\mathrm{x}}) dy$$

LPE and LPV can be found accordingly.

- Two most important risk measures are Value-at-Risk(VaR) and Tail-Value-at-Risk (TVaR) or Conditional Tail Expectation (CTE). CTE was independently developed and widely used in the insurance industry.
- VaR is the amount of capital the insurance company should have in order to guarantee there is a small probability that the insurance company will bankrupt by one adverse claim over the next period. VaR is defined as
$$P(X \leq VaR_p(X)) = p$$
for a given p and a distribution for X.
- CTE is the expected loss given that the loss is larger than VaR at p level and it is defined as
$$TVaR_p(X) = E[X|X > VaR_p(X)] = \frac{\int_{VaR_p(x)}^{\infty} xf(x)dx}{1-p}.$$

- Limited predictive CTE (LPCTE) can be calculated by using the predictive pdf.

$$\text{LPCTE} = E[(Y \wedge b)|Y > VaR_p(Y)] = \frac{\int_{VaR_p(y)}^{b} y f(y|\underline{x})dy + \int_{b}^{\infty} b f(y|\underline{x})dy}{1-p}.$$

Simulation

- Simulation studies are conducted to assess accuracy of estimates of unknown parameters by different methods, as well as estimates of important risk measures such as PE, PV, VaR, CTE, LPE, LPV, and LPCTE.

Simulation

Exponential-Pareto Model

- To assess accuracy of θ_{Bayes} as well as VaR, simulation studies are conducted.
- For selected values of n, θ and the hyper-parameters (a, b), $N = 300$ samples from the E-P composite density are generated

Simulation

Exponential-Pareto Model

Table 1: Accuracy of Bayes Estimator, MLE, and VaR, θ=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	5.22	0.059	802.25	22.82	7.68	25.17
50	80	5.15	0.021	763.92	39.30	7.47	18.450
100	80	5.07	0.04	27.99	1.36	6.20	3.28

Simulation

Exponential-Pareto Model

Table 2: Accuracy of Bayes Estimator, MLE, and VaR, θ=10

n	a	$\overline{\hat{\theta}_{Bayes}}$	$\xi(\hat{\theta}_{Bayes})$	$\overline{VaR}$	Std(VaR)	$\overline{\hat{\theta}_{MLE}}$	$\xi(\hat{\theta}_{MLE})$
20	50	10.653	.484	1594.48	69.893	13.923	73.893
50	100	10.576	.357	1608.41	39.65	12.503	21.582
100	150	9.998	.091	1810.46	22.52	12.061	13.473

Simulation

Exponential-Pareto Model

- In Tables 1-8, the line on top of each estimator denotes sample mean of simulated estimates. ξ denotes square root of MSE for an estimator.
- Examination of Tables 1 and 2 reveal that for larger n, $\hat{\theta}_{MLE}$ and $\hat{\theta}_{Bayes}$ are more accurate. $\bar{\hat{\theta}}_{Bayes}$, average of Bayes estimates is more closer to the actual value of θ than $\bar{\hat{\theta}}_{MLE}$, average of MLEs.
- It is noted that as θ increases, values of x in a sample get bigger and as a result m increase. Therefore, one would need to choose a larger values for “a”to satisfy the condition $A = a - .35n + 1.35m > 0$.
- For all values of n in Tables 1, 2, we note that MSE(Bayes)=$\xi(\hat{\theta}_{Bayes})$ is significantly smaller than MSE(MLE) =$\xi(\hat{\theta}_{MLE})$.
- Also Tables 1, 2 reveal that as θ increases, the sample mean $\overline{VaR}$ and StDev of VaR increase. This is anticipated, because, large value for θ implies more fatter composite density. However, for a fixed value of θ, as n increases, VaR estimate becomes more stable.

Table 3: Accuracy of $\hat{\theta}_{MLE}$, $\hat{\theta}_S$, and $\hat{\theta}_{Bayes_1}$

$(\theta = 5 \;\; \gamma = 5 \;\; \delta = 1)$

n	$\bar{\hat{\theta}}_{MLE}$	ξ_{MLE}	$\bar{\hat{\theta}}_S$	ξ_S	$\bar{\hat{\theta}}_{Bayes_1}$	ξ_{Bayes_1}
20	6.098	3.525	7.916	10.388	5.235	1.120
50	5.408	1.933	5.928	3.594	5.186	1.094
100	5.242	1.119	5.387	2.354	5.157	0.857
500	5.044	0.522	5.072	0.815	5.040	0.497

Simulation

Inverse Gamma-Pareto Model

Table 4: Accuracy of $\hat{\theta}_{MLE}$, $\hat{\theta}_S$, and $\hat{\theta}_{Bayes_1}$

$(\ \theta = 5 \ \ \gamma = 2 \ \ \delta = 2.5)$

n	$\bar{\hat{\theta}}_{MLE}$	ξ_{MLE}	$\bar{\hat{\theta}}_S$	ξ_S	$\bar{\hat{\theta}}_{Bayes_1}$	ξ_{Bayes_1}
20	5.395	1.680	8.714	12.060	5.395	1.680
50	5.445	1.787	5.782	13.961	5.295	1.365
100	5.361	1.298	5.580	2.522	5.106	1.159
500	5.042	0.522	5.063	0.902	5.040	0.511

Table 5: Accuracy of $\hat{\theta}_{MLE}$, $\hat{\theta}_S$, and $\hat{\theta}_{Bayes_1}$

$(\theta = 10 \quad \gamma = 2 \quad \delta = 5)$

n	$\bar{\hat{\theta}}_{MLE}$	ξ_{MLE}	$\bar{\hat{\theta}}_S$	ξ_S	$\bar{\hat{\theta}}_{Bayes_1}$	ξ_{Bayes_1}
20	11.500	14.657	17.889	23.900	11.500	4.139
50	11.290	4.021	12.518	9.852	10.888	3.019
100	10.36	2.232	10.630	4.664	10.289	1.988
500	10.033	0.989	10.040	1.780	10.029	0.969

Table 6: Accuracy of $\hat{\theta}_{MLE}$, $\hat{\theta}_S$, and $\hat{\theta}_{Bayes_1}$

$(\theta = 10 \;\; \gamma = 10 \;\; \delta = 1)$

n	$\bar{\hat{\theta}}_{MLE}$	ξ_{MLE}	$\bar{\hat{\theta}}_S$	ξ_S	$\bar{\hat{\theta}}_{Bayes_1}$	ξ_{Bayes_1}
20	13.188	11.698	18.162	27.407	10.110	1.380
50	10.694	3.637	11.602	6.580	10.080	1.576
100	10.402	2.350	11.193	5.340	10.159	1.470
500	10.130	1.0276	10.144	1.816	10.110	0.935

Table 7: Accuracy of $\hat{\theta}_{MLE}$, $\hat{\theta}_S$, and $\hat{\theta}_{Bayes_1}$

$(\theta = 6 \;\; \gamma = 1 \;\; \delta = 8)$

n	$\bar{\hat{\theta}}_{MLE}$	ξ_{MLE}	$\bar{\hat{\theta}}_S$	ξ_S	$\bar{\hat{\theta}}_{Bayes_1}$	ξ_{Bayes_1}
20	6.369	3.959	7.073	11.843	6.373	3.120
50	5.581	2.009	6.098	3.719	5.695	1.938
100	5.263	1.228	5.416	3.018	5.339	1.218
500	4.973	0.500	4.884	0.800	4.991	0.498

Table 8: Accuracy of $\hat{\theta}_{MLE}$, $\hat{\theta}_S$, and $\hat{\theta}_{Bayes_1}$

$(\theta = 7 \;\; \gamma = 2 \;\; \delta = 4)$

n	$\bar{\hat{\theta}}_{MLE}$	ξ_{MLE}	$\bar{\hat{\theta}}_S$	ξ_S	$\bar{\hat{\theta}}_{Bayes_1}$	ξ_{Bayes_1}
20	6.480	4.643	7.429	9.410	6.439	2.930
50	5.476	1.884	5.671	3.346	5.709	1.745
100	5.259	1.276	5.435	2.785	5.406	1.267
500	5.019	0.470	5.022	0.966	5.054	0.471

Simulation

Inverse Gamma-Pareto Model

- Tables 3-8 reveals that as sample size increase all estimators become more accurate.
- In Tables 3-4, hyper-parameters γ and δ are selected so that $E[\theta] = \gamma\delta$ (mean of gamma prior), is the same as the true value of θ. However, in Table 3, Var$(\theta) = \gamma\delta^2$ is smaller than in Table 4. And we can see that the Bayes estimator in Table 3 has a smaller MSE than in Table 4, in particular for small sample sizes.
- The same is true when comparing Table 5 with Table 6.
- In Table 7-8, on purpose, values of hyper-parameters are selected in such a way that $\gamma\delta$ is not equal to the true value of θ.
- Comparing Table 3 with Table 7, reveals that Bayes estimator is more accurate, in particular for small sample sizes (n=20,100,150) when "appropriate" values for hyper-parameters are not selected. As sample size increase, the impact of "inappropriate" hyper-parameters values diminishes.

Table 9: Accuracy of LPE and LPV

$(\theta = 5 \;\; \gamma = 2 \;\; \delta = 2.5)$

n	$\overline{LPE}$	$StDev(LPE)$	$\overline{LPV}$	$StDev(LPV)$
20	114306	5828.35	9.10002×10^{10}	3.96769×10^{9}
50	115132	4943.69	9.15723×10^{10}	3.35435×10^{9}
100	115528	4221.32	9.18478×10^{10}	2.86383×10^{9}
200	114792	2814.43	9.13568×10^{10}	1.91621×10^{9}

Table 10: Accuracy of LPE and LPV

$(\theta = 10 \quad \gamma = 2 \quad \delta = 5)$

n	$\overline{LPE}$	$StDev(LPE)$	$\overline{LPV}$	$StDev(LPV)$
20	129489	7862.23	1.1098×10^{11}	5.12453×10^{9}
50	129518	556.64	1.01148×10^{11}	3.60382×10^{9}
100	129459	4633.25	1.0118×10^{11}	3.02295×10^{9}
200	129041	3468.18	1.00855×10^{11}	2.25531×10^{9}

Table 11: Accuracy of VaR and LPCTE

$(\theta = 5 \ \ \gamma = 2 \ \ \delta = 2.5)$

n	$\overline{VaR_{.85}(y)}$	$StDev(VaR_{.85}(y))$	$\overline{LPCTE}$	$StDev(LPCTE)$
20	67662.7	22063.3	750455	36267.2
50	69315.4	19353.9	754681	30568.6
150	68438.4	13400.8	755312.	20734.1

Simulation

Inverse Gamma-Pareto Model

- Tables 9 and 10 reveal that as n increase, LPE and LPV become more accurate and stable as we expected.
- It is noted that LPE and LPVAR do not depend on sample size, they only depend on sampled data $\underline{x}$.
- Also LPE and LPV estimates are larger for $\theta = 10$ than for θ=5, and this is due to the heavier tail of the model with θ=10 than for the model with θ=5. $b = 10^6$ is used as the policy limit.
- Simulation results for VaR and LPCTE based on N=200 random samples with size n are given in the Table 11. It is assumed that the policy limit is $b = 10^6$.
- Table 11 indicates that as sample size increases, precision of VaR and LPCTE increase.

Model Selection

- Danish fire insurance loss data set has been used by many researches for a variety of composite models.
- There are 2492 losses in Danish Krone (DKK) data from the years 1980 to 1990. The data has been adjusted as 1985 values. The adjusted losses range from 0.3134041 to 263.250366 (in millions).
- We apply all three composite models E-P, IG-P, and W-P to the data and compute both MLE and Bayesian estimates. Use Anderson-Darling test to determine the best model among three of them.
- the risk measures for the best model will be found accordingly.

Model Selection

Estimates of parameters of composite models

Table 20: Estimates of parameters of Danish fire data Based on E-P Model

	Hyper-Parameter	Estimate of $\hat{\theta}$
MLE		2.76786
Bayes1	$a = 5, b = \hat{\theta}_{MLE}(a-1)$	2.77217
Bayes2	$a = 80, b = \hat{\theta}_{MLE}(a-1)$	2.92854

Table 21: Estimates of parameters of Danish fire data Based on IG-P Model

	Hyper-Parameter	Estimate of $\hat{\theta}$
MLE		3.32553
Bayes1	$\gamma = 5, \delta = \frac{\gamma}{\hat{\theta}_{MLE}}$	3.3387
Bayes2	$\gamma = 80, \delta = \frac{\gamma}{\hat{\theta}_{MLE}}$	3.70178

Model Selection

Graphs of E-P composite model to fit Danish Fire Data

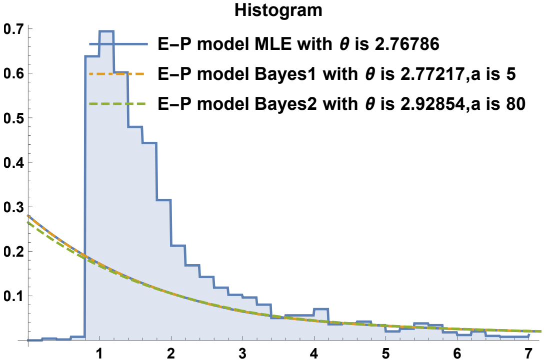

Figure 17: Comparison of the E-P model with empirical distribution of Danish Fire data.

- Figure 17 shows the empirical distribution of Danish Fire data (histogram) and the graphs of the E-P density using the estimated values in Table 20. The graphs of E-P composite density doesn't fit the histogram due to $\alpha < 1$.

Model Selection

Graphs of IG-P composite model to fit Danish Fire Data

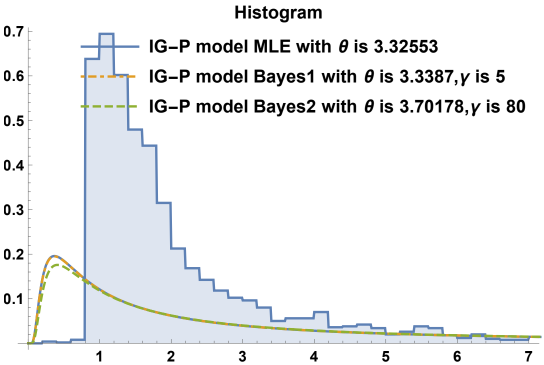

Figure 18: Comparison of the IG-P model with empirical distribution of Danish Fire data.

- Figure 18 shows the empirical distribution of Danish Fire data (histogram) and the graphs of the IG-P density using the estimated values in Table 21. The graphs of IG-P composite density don't fit the histogram due to $\alpha < 1$. But IG-P is better than E-P model.

Summary

- A Bayes estimator via Inverse-gamma prior for the boundary parameter θ, that separates large losses and small losses in insurance data is derived based on the Exponential-Pareto, Inverse Gamma-Pareto, and Weibull-Pareto composite model.
- Simulation studies indicate that the Bayes estimators consistently outperform the maximum likelihood estimators in all three composite models.
- A Bayesian predictive density is derived via the posterior pdf for θ for both E-P and IG-P models and θ and α for W-P model.
- The Bayesian predictive pdf is used to estimate important risk measures that are used in Actuarial Science field, such as VaR, LPE, LPV, and LPCTE. Simulation studies indicate that large sample size would make these estimates more accurate.
- All three composite models are used to fit the insurance data-the Danish fire data. Weibull-Pareto is the best model out of three.

References

Abu Bakar, S.A., Hamzeh, N.A., Maghsoudi, M., Nadarajah, S. ”Modeling loss data using composite models” Insurance: Mathematics and Economics, 2015, 61: pp146-154

Aminzadeh, M.S. , Deng, M. “Bayesian Predictive Modeling for Exponential-Pareto Composite Distribution ” To appear in Variance, 2018.

Aminzadeh, M.S. , Deng, M. “Bayesian Predictive Modeling for Inverse Gamma-Pareto Composite Distribution” To appear in Communications in Statistics, Theory and Methods, 2018.

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

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

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

References

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

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

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

Vekas, P. “An asymptotic test for the conditional value-at-risk with applications in capital adequacy” Presented to the Actuaries Institute ASTIN, AFIR/ERM and IACA Colloquia, 2015. URL: http://actuaries.asn.au/Library/Events/ASTINAFIRERMColloquium/ 2015/VekasAsymptoticPaper.pdf

Thank you