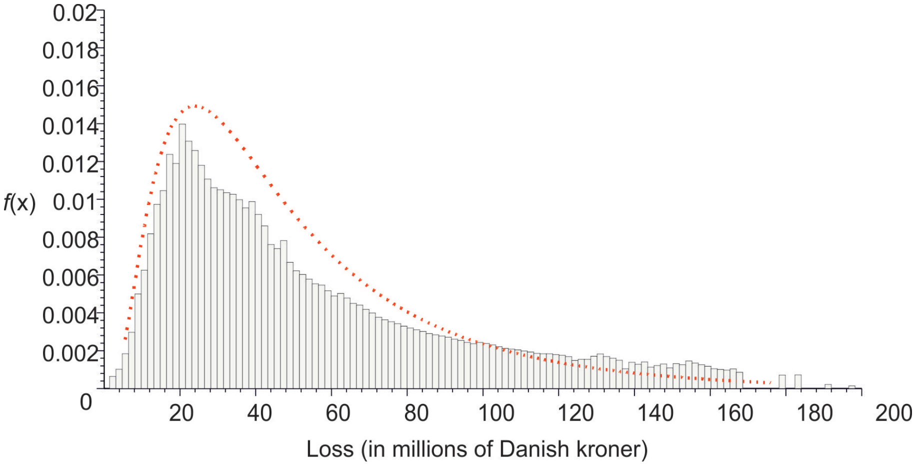
# Bivariate Composite Model constructed via Joe copula for modeling claim costs

Enrollment No. 2022MSST011
Under Guidance of
Dr. Deepesh Bhati

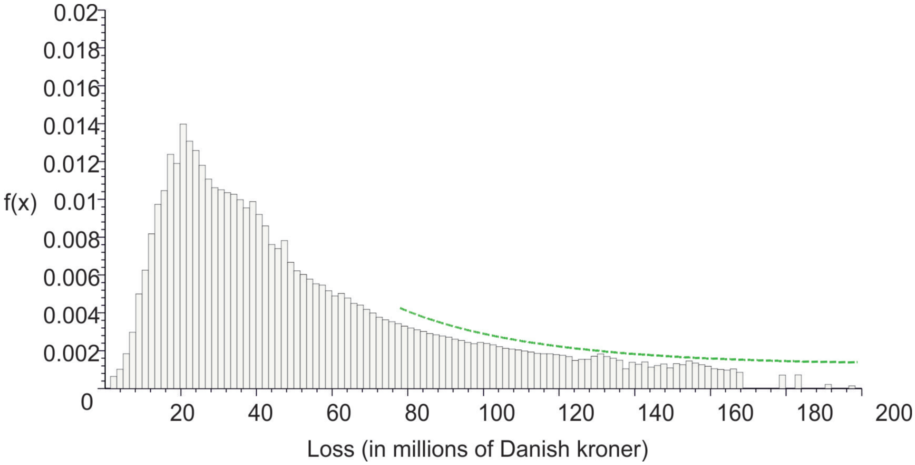
Central University of Rajasthan
Department of Statistics

# Contents

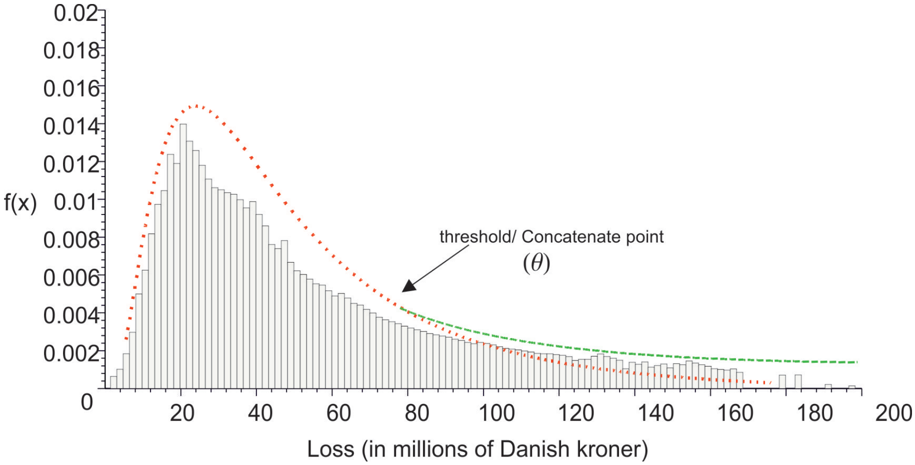
# A Graphical View of Lognormal fit for Danish data



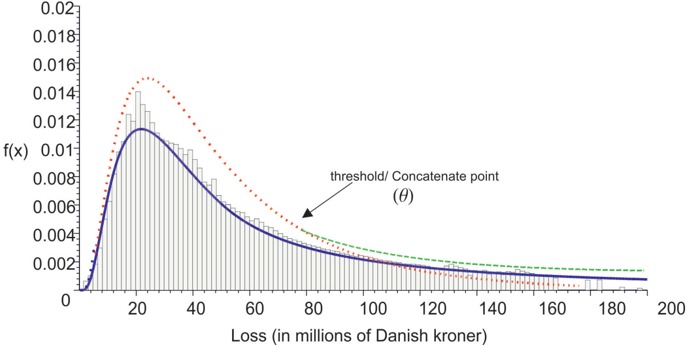
# Graphical View of Pareto fit for Danish data



# Graphical View of Composite Model



# Graphical View of Composite Model




- We call this technique of generating composite models as "Classical Composite (CC) Technique"

# Generation of Composite Model using CC-Technique

Bakar et al. (2015) proposed various composite distributions using an unrestricted mixing weight ($r$), the right-truncated and left-truncated densities truncated at thresholds ($\theta$) for the Head and Tail distributions, respectively. The resulting probability density The composite distribution's function (pdf) can be expressed as

$$f(x) = \begin{cases} rf_1^*(x|\Phi_1, \theta) & \text{for} \quad 0 < x \leq \theta, \\ (1-r)f_2^*(x|\Phi_2, \theta) & \text{for} \quad \theta < x < \infty, \end{cases} \tag{1}$$

where $\Phi_1$ and $\Phi_2$ are the parameter spaces for the head and tail parts of the composite distribution, with $\theta > 0$ and the mixing weight $r \in [0, 1]$. The functions $f_1^*(x|\Phi_1, \theta) = \frac{f_1(x|\Phi_1)}{F_1(\theta|\Phi_1)}$ and $f_2^*(x|\Phi_2, \theta) = \frac{f_2(x|\Phi_2)}{1-F_2(\theta|\Phi_2)}$ are the adequate truncations of the pdfs $f_1$ and $f_2$ up to and after an unknown threshold value $\theta$, respectively.

# Generation of Composite Model using CC-Technique

- The value of weight parameter $r$ is obtained by continuity condition imposed at threshold $\theta$ i.e. $rf_1^*(\theta|\Phi_1,\theta) = (1-r)f_2^*(\theta|\Phi_2,\theta)$. Hence, we get

$$r = r(\theta,\Phi_1,\Phi_2) = \frac{f_2(\theta|\Phi_2).F_1(\theta|\Phi_1)}{f_2(\theta|\Phi_2).F_1(\theta|\Phi_1) + f_1(\theta|\Phi_1).(1-F_2(\theta|\Phi_2))} \quad (2)$$

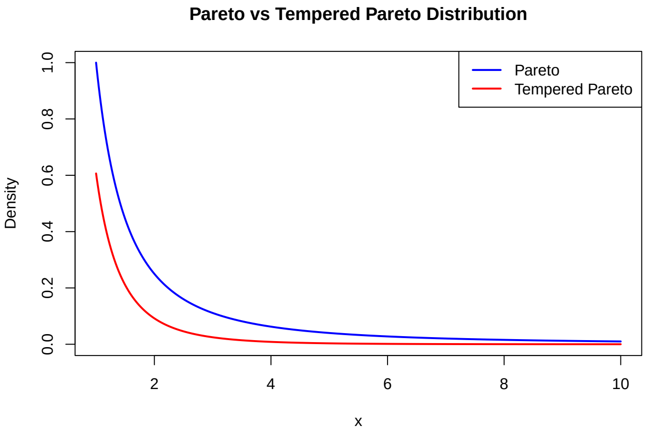
- Further, imposing the differentiability condition at threshold value $\theta$ $rf_1^{'}(\theta|\Phi_1,\theta) = (1-r)f_2^{'}(\theta|\Phi_2,\theta_2)$, makes the density smooth.

These above conditions reduces the number of parameters and makes the resulting density continuous and differentiable. We henceforth refer this technique as Classical Composition technique (CC).

# Tail Tempering

Heavy tailed distributions are very common in actuarial field, that is the distributions consist less extreme value events but the events can have a big impact. Tail tempering is a technique used to modify the distribution of data to reduce the impact of extreme values or "tails". Tail tempering involves adjusting the tails of the distribution to make the tail less "heavy" to get get more stable and manageable risk assessment , pricing models, and forecasting methods. Tail tempering is particularly valuable in fields such as finance and insurance, where accurately predicting rare, high impact events are crucial.

# Graphical view of Tail tempering




- The Pareto distribution has a heavy tail, meaning it's more likely to have extreme values. On the other hand, the Exponentially Pareto distribution also has a long tail but drops off more quickly, so extreme values are less likely.

# Bivariate Extension

Now, in the present study, we introduce a copula-based bivariate distributions having composite marginal distributions to model `Danish` dataset. The motivation to extend our model to a bivariate setup comes from data itself to address the bivariate nature of the data and to model the two types of claims simultaneously. Here we are extending this setup to the multivariate counterpart based on the idea of Aradhaye et al., 2024.

# Copula

A copula $C(u, v)$ is a bivariate cumulative distribution function (cdf) with uniform marginals on the interval $(0, 1)$ (Joe 1997, Nelsen 2006). If $F_j(x_j)$ is the cdf of a univariate r.v. $X_j$, then $C(F_1(x_i^{(1)}), F_2(x_i^{(2)}))$ is a bivariate distribution for $\mathbf{X} = (\mathbf{X^{(1)}}, \mathbf{X^{(2)}})$ with marginal distributions $F_j$, where $i = 1, 2 \cdots, n$ and $j = 1, 2$.

The dependence of $X_i$ among the claim types are modelled using copula, with the joint distribution of $X_i$ given by

$$\pi(x_i) = C_\delta \left( F_1 \left( x_i^{(1)} \right), F_2 \left( x_i^{(2)} \right) \right), \tag{3}$$

where $\mathbf{x_i} = (x_i^{(1)}, x_i^{(2)})$ are the realizations of the $\mathbf{X} = (x^{(1)}, x^{(2)})$, $C_\delta$ is a copula function and $\delta := \{\delta^{(j,j')}\}_{j,j'=1,2}$ is the copula parameter that explains the association between the two random variables say $(\mathbf{X^{(1)}}, \mathbf{X^{(2)}})$.

# Joe Copula

Numerous insurance data sets show strong correlations at high claim amount values but a weaker correlation at low claim amount values. Henceforth, the Joe copula will be a suitable choice to model such a data set in order to jointly model the two types of claims having strong tail dependence. The dependency between two types of claims $x_i^{(1)}, x_i^{(2)}$ can be modelled through the Joe copula as

$$C_\delta\left(F_1(x_i^{(1)}), F_2(x_i^{(2)})\right) = 1 - \Big[(1 - F_1(x_i^{(1)}))^\delta + (1 - F_2(x_i^{(2)}))^\delta - (1 - F_1(x_i^{(1)}))(1 - F_2(x_i^{(2)}))^\delta\Big]^{\frac{1}{\delta}}, \quad (4)$$

where $\delta := \{\delta^{(j,j')}\}_{j,j'=1,2} \in [1, \infty)$ be the asymmetric copula parameter influencing the correlations among $(\mathbf{X}^{(1)}, \mathbf{X}^{(2)})$. $F_j(x_j)$ is the cdf of a univariate random varibale $X_j$ for $j = 1, 2$.

# Bivariate Distribution

Let $X_i^{(j)}$ be a random variable (r.v.) derived from the marginal composite model by taking into account the Exponential distribution for the head and the Inverse Weibull distribution for the tail. The r.v. $X_i^{(j)}$'s probability density function (pdf) can be expressed as

$$f_j(x_i^{(j)}) = \begin{cases} r_{E,ETP}^{(j)} \frac{\lambda^{(j)} \exp\left\{-\lambda^{(j)} x_i^{(j)}\right\}}{1-\exp\left\{-\lambda^{(j)}\theta^{(j)}\right\}}, \text{for} \quad 0 < x_i^{(j)} \leq \theta^{(j)} \\ \\ (1 - r_{E,ETP}^{(j)})\gamma^{(j)} x_i^{(j)\left(-\alpha^{(j)}-1\right)} \exp\left(-\beta^{(j)} x_i^{(j)}\right)\left(\alpha^{(j)} + x_i^{(j)}\beta^{(j)}\right), \\ \text{for} \quad \theta^{(j)} < x_i^{(j)} < \infty. \end{cases} \tag{5}$$

where $i = 1, 2, \cdots, n$ and $j = 1, 2$, where $\lambda^{(j)} > 0$, $\alpha^{(j)} > 0$, $\beta^{(j)} > 0$ is the tempered parameter, threshold point $\theta^{(j)} > 0$, $\gamma^{(j)} = \theta^{(j)\alpha^{(j)}} \exp\{\beta^{(j)}\theta^{(j)}\}$ and $r_{E,ETP}^{(j)} \in [0, 1]$ is the mixing weight of the composite model.

# Bivariate Distribution

The cumulative distribution function (cdf) of the composite E-ETP distribution is

$$F_j(x_i^{(j)}) = \begin{cases} r_{E,ETP}^{(j)} \frac{1-\exp\left\{-\lambda^{(j)} x_i^{(j)}\right\}}{1-\exp\left\{-\lambda^{(j)}\theta^{(j)}\right\}}, \text{for} \quad 0 < x_i^{(j)} \leq \theta^{(j)} \\ r_{E,ETP}^{(j)} + (1 - r_{E,ETP}^{(j)}) \left(1 - \gamma^{(j)} x_i^{(j)\left(-\alpha^{(j)}\right)} \exp\left(-\beta^{(j)} x_i^{(j)}\right))\right), \\ \text{for} \quad \theta^{(j)} < x_i^{(j)} < \infty. \end{cases} \tag{6}$$

Two marginal E-ETP cdfs can be coupled using the Joe copula to generate the joint cdf of the bivariate composite E-ETP distribution, as demonstrated in (4)

# Estimation of parameters using IFM technique

This sections aims to describe how to estimate the marginals and the copula parameters using maximum likelihood (ML) approach. Joe (1997) described a marginally dependent technique for parameter estimation of the copula density called inference function for margins(IFM). Let $X_1^{(j)}, X_2^{(j)}, \ldots, X_n^{(j)}$ for $j = 1, 2$ be a random sample of two types of claims. In this approach at first , the parameters of the are obtained by maximizing the copula's likelihood function, susbtituing the first step estimators for the marginal parameters. The likelihood function is given by

$$\mathcal{L}(\mathbf{\Theta^{(j)}}, \delta | x_i^{(j)}) = \sum_{i=1}^{n} \ln C_\delta \left( F_1(x_i^{(1)}), F_2(x_i^{(2)}) \right) + \sum_{i=1}^{n} \ln(f_j(x_i^{(j)} | \mathbf{\Theta})) \quad (7)$$

# Estimation of parameters using IFM technique

In the first step of inference function for margins (IFM), first we will estimates the parameters of the maginals composite Exponential - Expoentially tempered Pareto distribution described in 5 by maximizing

$$\mathcal{L}(\mathbf{\Theta}^{(\mathbf{j})}|x_i^{(j)}) = \sum_{i=1}^{n} \ln(f_j(x_i^{(j)}|\mathbf{\Theta})) \tag{8}$$

where, $\mathbf{\Theta}$ is the parameter vector for the two marginal composite Exponential-Exponentially tempered Pareto distribution. After getting the estimates of the parameters then we will substitute the estimate values in the following likelihood function and obtain the estimate of $\delta$ by maximizing

$$\mathcal{L}(\delta|x_i^{(j)}) = \sum_{i=1}^{n} \ln C_\delta \left( F_1(x_i^{(1)}), F_2(x_i^{(2)}) \right) \tag{9}$$

# Dataset

- The dataset contains fire insurance claims data from Denmark, spanning the years 1980 to 1990. The source of the data is the Copenhagen Reinsurance Company. Data can be accessed from the website `http://www.ma.hw.ac.uk/~mcneil/`.
- The dataset is composed of three main components: `loss to buildings`, `loss to contents`, and `loss to profit`. This study focuses specifically on modeling the dependence between `loss to buildings` and `loss to contents`.
- There are a total of 1502 observations in the dataset.
- The analysis includes only those cases where both `loss to buildings` and `loss to contents` have non-null values.
- Notable characteristics of the data include:
  - A positive skew in the distribution.
  - Unimodality.
  - Large tails, indicating the presence of extreme values.

# Model Fitting

We fit our bivariate composite distribution, namely Exponential-Exponentially tempered Pareto distribution on the `Danish` fire loss data set and compare our model with existing Bivariate Model. The criterion by which we judge the models based on model sellection criteria , known as AIC and BIC. The formula of the model slection criterion is given by ,

$$AIC = -2\mathcal{L}(\hat{\mathbf{\Theta}}) + 2 \times \mathrm{k},$$

where $\mathcal{L}(\hat{\mathbf{\Theta}})$ is the maximum log-likelihood and $\hat{\mathbf{\Theta}}$ is the vector of the estimated model parameters.

$$BIC = -2\mathcal{L}(\hat{\mathbf{\Theta}}) + \log(n) \times \mathrm{k},$$

where $n$ is the sample size of the dataset and k is the number of fitted parameters of the distribution.

# Model Results

Table 1 show that for the `Danish` fire loss data set,our proposed bivariate composite model , that is Exponential-Exponentially tempered Pareto distribution performs better than the other existing bivariate model.

Table: Values of AIC and BIC for the `Danish` dataset for competing distributions

| Model | Parameters | AIC | BIC |
|---|---|---|---|
| **E – ETP** | **7** | **7912.38** | **7949.59** |
| BL | 3 | 8245.08 | 8261.02 |
| BMPI | 3 | 8043.02 | 8058.96 |
| BB | 3 | 8573.66 | 8589.61 |

# Likelihood Ratio Test (LRT)

Let $X_1, X_2, \ldots, X_n \sim f_\theta(), \theta \in \Theta$
We consider the following testing problem $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_1$ We define the following ratio

$$\Lambda = \frac{\sup_{\theta \in \Theta_0} L(\theta)}{\sup_{\theta \in \Theta} L(\theta)}$$

Now , we know that $-2 \log \Lambda \sim \chi^2$ with degrees of freedom equal to the difference in the number of parameters. We want to test the dependency between two types of claims that can encountered by our copula based bivariate model. If the Joe Copula parameter $\delta = 1$, then we can say that the two types of claims are independent.

# LRT

So we want to test the hypothesis whether `loss to buildings` and `loss to contents` have some dependency or not , can be performed as follows :

$$H_0 : \delta = 1 \text{ vs } H_1 : \delta > 1$$

After calculating the likelihood values under the null and alternative hypotheses, we will compute the test statistic $-2 \log \Lambda$. We will compare the computed test statistic with the critical value and reject the null hypothesis if the test statistic exceeds the critical value, indicating evidence of dependency between the two types of claims.
So, the test statistic $-2 \log \Lambda$ value is 214.74. At 0.05 significance level , we reject the null hypothesis which implies that the two types of claims are independent. So, after apply LRT we can say there exist a certain level of dependency between two types of claim.

# Conclusion

# References


Mark M. Meerschaerty, Parthanil Roy, & Qin Shao (2010). Parameter estimation for exponentially tempered power law distributions

Bakar, S.A. A., Hamzaha, N. A.,Maghsoudia, M. & Nadarajah, S. (2015). modelling loss data using composite models. *Insurance: Mathematics and Economics*, 61, 146–154.

Girish Aradhye, George Tzougas and Deepesh Bhati (2024). A Copula based bivariate composite model for modelling claim costs.