Enrollment No.: 2022MSST011

Bivariate Composite Model constructed via Joe copula for modeling claim costs

Project Report

A thesis submitted in fulfillment of the requirements

for the degree

of

Master of Sciences

in

Statistics

Submitted By:

Mr. Saikat Pal

Supervisor:

Dr. Deepesh Bhati

Associate Professor

Department of Statistics

School of Mathematics, Statistics and Computational Sciences

Central University of Rajasthan, Ajmer - 305817

2022-2024

Dr. Deepesh Bhati
Associate Professor
Department of Statistics

Email : deepesh.bhati@curaj.ac.in
Website: http://www.curaj.ac.in

राजस्थान केंद्रीय विश्वविद्यालय
Central University of Rajasthan
(संसद के अधिधियम द्वारा स्थाधित केंद्रीय धिश्वधिद्यालय)
(A Central University by an Act of Parliament)
NH-8, Bandarsindri, 305801
Kishangarh (Ajmer), Rajasthan, INDIA
Phone (Office): +91-1463-238755/260200
Telefax: +91-1463-238722

Certificate

This is to certify that the work embodied in the accompanying project report entitled ***"Bivariate Composite Model constructed via Joe copula for modeling claim costs"*** has been successfully carried by ***Mr. Saikat Pal***, a IV Semester student of M.Sc. Statistics, of the Department of Statistics, Central University of Rajasthan, under my guidance and supervision.

He worked for whole IV semester and his work carried out is satisfactory.

Place: Cental University of
Rajasthan, Bandarsindri
Date:

Dr. Deepesh Bhati
Associate Professor
Department of Statistics

TO WHOMSOEVER IT MAY CONCERN

This is to certify that ***Mr. Saikat Pal***, Enroll. No 2022MSST011, M.Sc. Statistics student of Department of Statistics has done project work ***"Bivariate Composite Model constructed via Joe copula for modelling claim costs"*** under the guidance of ***Dr. Deepesh Bhati***, Associate Professor at Department of Statistics, Central University of Rajasthan towards the partial fulfillment of the award of "Master of Science in Statistics" during the period January 2024 to May 2024.

Place: Cental University of
Rajasthan, Bandarsindri
Date:

Dr. Deepesh Bhati
Head
Department of Statistics
Central University of Rajasthan

Declaration

I, Saikat Pal, hereby declare that the project work entitled **"Bivariate Composite Model constructed via Joe copula for modelling claim costs"** submitted to Department of Statistics, Central University of Rajasthan as a partial fulfilment of requirements of IV-Semester examination, is a bona fide record of work under taken by me, under the supervision of **Dr. Deepesh Bhati**, Associate Professor at Department of Statistics, Central University of Rajasthan and it is not formed the basis for the award of any other Degree/Associateship/Fellowship by any University.

Signature of Candidate

Saikat Pal

Enroll. No.: 2022MSST011

Place: Bandarsindri

Date:

Acknowledgement

I would like to express my sincere gratitude to all those who have supported and guided me throughout the journey of completing this thesis.

First and foremost, I am deeply thankful to my supervisor, **Dr. Deepesh Bhati**, for their invaluable guidance, unwavering support, and insightful feedback. Their expertise and dedication played a pivotal role in shaping the direction of my research and enhancing the quality of this work.

I am also grateful to Ph.D. Scholar, **Mr. Girish Aradhye** who provided huge support and encouragement for learning and research.

My heartfelt thanks go to my family and friends for their constant encouragement and understanding during this challenging phase of my academic journey. Their belief in me kept me motivated to overcome obstacles and strive for excellence.

Lastly, I am indebted to the various resources, libraries, and research materials that have contributed to the depth and authenticity of this thesis.

Thank you all for being a part of this fulfilling endeavor.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

There has been a notable wave in actuarial field during the last few times , focused on modelling the costs of different types of claims in non-life insurance, employing a diverse range of claim severity approaches.

This is because the claim severity distribution has special properties that present a number of difficulties. The distribution frequently covers a range of magnitudes, including a few significant claims with low frequency and small and moderate ones with high frequency. Furthermore, claim size statistics are significantly skewed to the right and unimodal (see, for example, Bakar et al., 2015 [3]).

As is evident,choosing a probability distribution that can effectively fit small, moderate, and large claims becomes essential for insurance pricing, reserving, and risk management when the data spans a wide range of magnitudes. Such datasets can be reasonably fitted using the distribution composition technique (see, e.g., Cooray and Ananda, 2005 [2]; Scollnik, 2007 [11]; Ciumara, 2006 [5]; Scollnik and Sun, 2012 [12], and the references therein).

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.

In this paper we propose a composite distribution model in which we choose exponential distribution for the head part of the distribution to a certain point θ and due to upper mentioned characteristics of tempering , we use exponentially tempered pareto distribution for the tail part of the composite distribution. According to Meerschaert, M. M. et.al (2011) [8], the Exponentially tempered Pareto distribution introduces a tempering parameter that adjusts the tail behavior, making it suitable for applications such as financial risk assessment and insurance claims modeling.

It is noteworthy that even though the literature concerning composite models in the univariate setting is plenty, their bivariate extensions have not been investigated so far. However, in non-life insurance, it is common for the actuary to observe the existence of dependence structures between different types of claims and their associated costs, either from the same type of coverage or from multiple types of coverage, such as motor and home insurance bundled into one single policy.

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 [2].

A suitable copula distribution made up of two marginal composite distributions is used to generate the bivariate distributions. A bivariate composite distribution based on copulas is constructed and explained in general. One particular bivariate composite distribution is provided for explanation's sake. This distribution has a few extreme values but primarily consists of tiny and medium values to characterize the behaviour of bivariate data. The copula distribution is used to calculate dependence measures for the proposed bivariate composite distribution. The inference function for margins (IFM) method is used to estimate the distributions' parameters. It involves estimating the univariate parameters by independently maximizing the marginal composite distribution and estimating the dependence parameters from the bivariate likelihoods that are derived using the copula.

Chapter 2

Modelling framework

2.1 The univariate composite distribution

Bakar et al. (2015) [3] proposed various composite distributions using an unrestricted mixing weight (r), the right-truncated and left-truncated densities truncated at thresholds (θ) for the Head and Tail distributions, respectively. The resulting probability density The composite distribution's function (pdf) can be expressed as

$$f(x) = \begin{cases} r f_1^*(x|\Phi_1, \theta) & \text{for} \quad 0 < x \leq \theta, \\ (1-r) f_2^*(x|\Phi_2, \theta) & \text{for} \quad \theta < x < \infty, \end{cases} \tag{2.1}$$

where Φ_1 and Φ_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 θ, respectively.

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

$$r = r(\theta,\Phi_1,\Phi_2) = \frac{f_2(\theta|\Phi_2)F_1(\theta|\Phi_1)}{f_2(\theta|\Phi_2)F_1(\theta|\Phi_1) + f_1(\theta|\Phi_1)(1-F_2(\theta|\Phi_2))}. \tag{2.2}$$

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

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

Composite Exponential - Exponentially tempered Pareto Distribution

Let X be a random variable (r.v) obtained by considering the Exponential distribution for the head and the Exponentially tempered Pareto distribution (see Meerschaert, M. M. et.al (2011) [8]) for the tail part of the composite model. The probability density function (pdf) of the r.v. X can be written as

$$f(x) = \begin{cases} r \cdot \frac{\lambda e^{-\lambda x}}{1-\exp(-\lambda\theta)} & \text{for } 0 < x < \theta \\ (1-r)\cdot \gamma x^{-\alpha-1}e^{-\beta x}(\alpha + x\beta) & \text{for } x \geq \theta \end{cases} \tag{2.3}$$

where $\lambda > 0$, $\alpha > 0$, $\beta > 0$, and the threshold point $\theta > 0$ and satisfies $\gamma = \theta^\alpha e^{\beta\theta}$ and r is the mixing weight of the composite model. The analytical expression for the mixing weight r can be easily obtained using (2.2).

2.2 The Joe copula

A copula $C(u, v)$ is a bivariate cumulative distribution function (cdf) with uniform marginals on the interval $(0, 1)$ (Joe H. 1997 [6], Nelsen 2006 [9]). 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

$$\pi(x_i) = C_\delta \left(F_1 \left(x_i^{(1)} \right), F_2 \left(x_i^{(2)} \right) \right), \tag{2.4}$$

where $\mathbf{x_i} = (x_i^{(1)}, x_i^{(2)})$ are the realizations of the $\mathbf{X} = (x^{(1)}, x^{(2)})$, C_δ 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 $(\mathbf{X^{(1)}}, \mathbf{X^{(2)}})$. 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 (see Joe H. 2005 [7]). The dependency between two types of claims $x_i^{(1)}, x_i^{(2)}$ can be modelled through the Joe copula as

$$C_\delta \left(F_1(x_i^{(1)}), F_2(x_i^{(2)}) \right) = 1 - \left[(1 - F_1(x_i^{(1)}))^\delta + (1 - F_2(x_i^{(2)}))^\delta - (1 - F_1(x_i^{(1)}))(1 - F_2(x_i^{(2)}))^\delta \right]^{\frac{1}{\delta}}, \tag{2.5}$$

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 variable $X^{(j)}$ for $j = 1, 2$.

Modelling dependence via the Joe copula

Copulas helps to understand the dependencies among the random variables. With the help of copula , we can easliy differentiate the marginal distributions and the underlying dependency. Since a copula that measures reliance is known to be invariant under precisely monotone transformations, it stands to reason that a more precise global measure of dependence would also be invariant under these same transformations. Kendall's τ is, in addition to other dependence measures, invariant under strictly increasing transformations and may be expressed in terms of the associated copula, as we shall demonstrate in the following section. :

- **Kendall's tau:** Kendall's tau, denoted by τ, is a bivariate measure of dependence for continuous variables that measures the amount of concordance present in a bivariate distribution. Let F be a continuous bivariate cumulative distribution function (cdf), and let (X_1, X_2) and (X_1', X_2') be independent random pairs with distribution F. Then Kendall's tau is given by:

$$\tau = \Pr\Big((X_1 - X_1')(X_2 - X_2') > 0\Big) - \Pr\Big((X_1 - X_1')(X_2 - X_2') < 0\Big). \quad (2.6)$$

Unlike the Pearson correlation coefficient, Kendall's tau solely depends on the copula function $C_\delta\left(F_1(y_i^{(1)}), F_2(y_i^{(2)})\right)$. through (see Fredricks and Nelsen, 2007, Nelsen, 2007):

$$\tau = 4\iint_{[0,1]^2} C_\delta\left(F_1(x_i^{(1)}), F_2(x_i^{(2)})\right)\, dC_\delta\left(F_1(x_i^{(1)}), F_2(x_i^{(2)})\right) - 1.$$

Kendall's tau for the Joe copula in terms of the copula parameter δ can be written as:

$$\tau = 1 + \frac{4}{\delta^2}\int_0^1 x\log(x)(1-x)^{\frac{2(1-\theta)}{\theta}}\, dx,$$

where δ is the asymmetric Joe copula parameter.

- **Tail dependence property:** Bivariate tail dependence is the degree of dependence in either the upper or lower quadrant tail of a bivariate distribution. Dependency on extreme values is relevant to the idea of tail dependency. The tail dependency parameter is represented by the symbol λ.If a bivariate copula $C_\delta\left(F_1(x_i^{(1)}), F_1(x_i^{(1)})\right)$ exists such that:

$$\lim_{F_1(x_i^{(1)})\to 1} \frac{C_\delta\left(F_1(x_i^{(1)}), F_1(x_i^{(1)})\right)}{1 - F_1(x_i^{(1)})} = \lambda_U,$$

then $C_\delta\left(F_1(x_i^{(1)}), F_1(x_i^{(1)})\right)$ has upper-tail dependence if $\lambda_U \in (0,1]$ and no upper-tail dependence if $\lambda_U = 0$. The expression for the upper tail dependence parameter λ_U for the Gumbel copula is given by:

$$\lambda_U = 2 - 2^{\frac{1}{\delta}}.$$

2.3 The bivariate composite Exponential-Exponentially tempered Pareto 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 Exponentially tempered Pareto distribution distribution for the tail. The r.v. $X_i^{(j)}$'s probability density function (pdf) can be expressed as

$$f_j(x_i^{(j)}) = \begin{cases} r_{E,ETP}^{(j)} \frac{\lambda^{(j)} \exp\left\{-\lambda^{(j)} x_i^{(j)}\right\}}{1-\exp\left\{-\lambda^{(j)}\theta^{(j)}\right\}}, & \text{for} \quad 0 < x_i^{(j)} \leq \theta^{(j)} \\ \\ (1 - r_{E,ETP}^{(j)})\gamma^{(j)} x_i^{(j)\left(-\alpha^{(j)}-1\right)} \exp\left(-\beta^{(j)} x_i^{(j)}\right)\left(\alpha^{(j)} + x_i^{(j)}\beta^{(j)}\right), & \text{for} \quad \theta^{(j)} < x_i^{(j)} < \infty. \end{cases} \tag{2.7}$$

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. The analytical expression for the mixing weight $r_{E,ETP}^{(j)}$ can be easily obtained using (2.2) .

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

$$F_j(x_i^{(j)}) = \begin{cases} r_{E,ETP}^{(j)} \frac{1-\exp\left\{-\lambda^{(j)} x_i^{(j)}\right\}}{1-\exp\left\{-\lambda^{(j)}\theta^{(j)}\right\}}, & \text{for} \quad 0 < x_i^{(j)} \leq \theta^{(j)} \\ r_{E,ETP}^{(j)} + (1 - r_{E,ETP}^{(j)})\left(1 - \gamma^{(j)} x_i^{(j)\left(-\alpha^{(j)}\right)} \exp\left(-\beta^{(j)} x_i^{(j)}\right)\right)\Big), & \text{for} \quad \theta^{(j)} < x_i^{(j)} < \infty. \end{cases} \tag{2.8}$$

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 (2.5)

2.4 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) [6] 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, substituting the first step estimators for the marginal parameters. The likelihood function is given by

$$l(\mathbf{\Theta^{(j)}}, \delta | x_i^{(j)}) = \sum_{i=1}^{n} \ln C_\delta \left(F_1(x_i^{(1)}), F_2(x_i^{(2)}) \right) + \sum_{i=1}^{n} \ln(f_j(x_i^{(j)} | \mathbf{\Theta})) \tag{2.9}$$

In the first step of inference function for margins (IFM), first we will estimates the parameters of the marginals composite Exponential - Exponentially tempered Pareto distribution described in 2.7 by maximizing

$$l(\mathbf{\Theta^{(j)}} | x_i^{(j)}) = \sum_{i=1}^{n} \ln(f_j(x_i^{(j)} | \mathbf{\Theta})) \tag{2.10}$$

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 δ by maximizing

$$l(\delta | x_i^{(j)}) = \sum_{i=1}^{n} \ln C_\delta \left(F_1(x_i^{(1)}), F_2(x_i^{(2)}) \right) \tag{2.11}$$

Chapter 3

Numerical Illustration

Data set : `Danish` fire insurance data set

The data set presented details fire insurance claims made in Denmark between 1980 and 1990, based on information obtained from the Copenhagen Reinsurance Company. The following website has the data set: www.ma.hw.ac.uk/ mcneil/. `loss to buildings`, `loss to contents`, and `loss to profit` are its three primary parts. But in this instance, we are particularly interested in modelling the dependence between the first two elements. There are 1502 observations in the data set overall. Our study focuses on instances in which both losses have non-null values. A detailed description of the descriptive data for the two different claim categories—`loss to buildings` and `loss to contents`—is provided in Table 3.1. Graphical depictions of these claim kinds are shown in Figures 3.1, 3.2, and 3.3. Remarkably, `loss to buildings` and `loss to contents` have various noteworthy characteristics that are frequently noticed in insurance data, such as a distribution that is positively skewed, the existence of a single mode (unimodality), and large tails that indicate the existence of extreme values. The scatter plot of the `Danish` fire loss data set using natural logarithmic scales shown in Figure 3.4 offers important information about how these assertions relate to one another. It becomes clear that there is a non linear connection between these different claim categories. Rather, the evidence indicates the presence of high value dependence and a non-linear connection. Put differently, the stronger the connection between them, the higher the claim costs. Furthermore, a deeper look at Figure 3.4 reveals an interesting pattern: for smaller values of both `loss to buildings` and `loss to contents`,their dependence is essentially insignificant , almost reaching zero. Nonetheless, the image unmistakably shows the emergence of a positive reliance as losses increase. The nature of the joint distribution between these two types of losses has changed significantly and is worth mentioning. It went from nearly zero reliance for smaller losses to a positive association for greater losses.

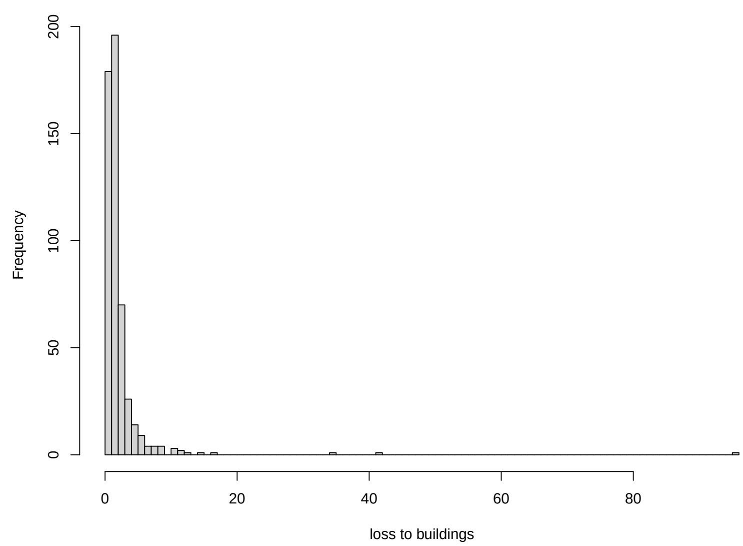

Figure 3.1: Histogram of `loss to buildings` for the `Danish` data set

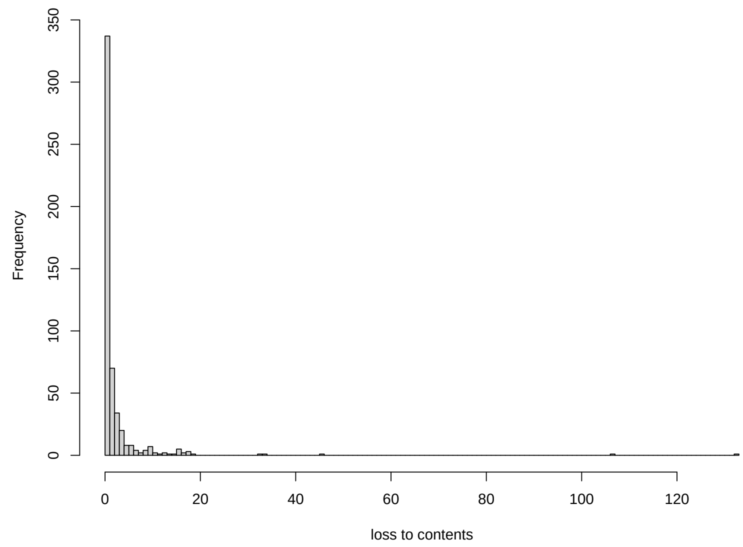

Figure 3.2: Histogram of `loss to contents` for the `Danish` data set

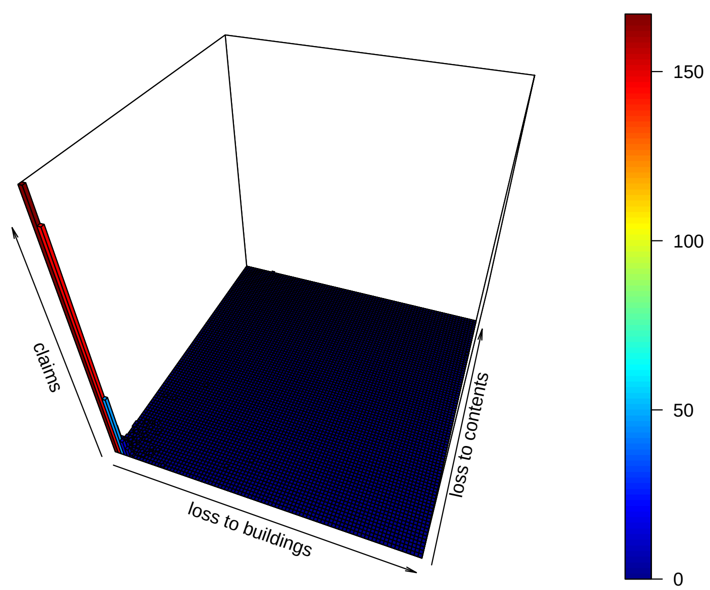

Figure 3.3: Histogram of `loss to buildings` and `loss to contents` for the `Danish` data set

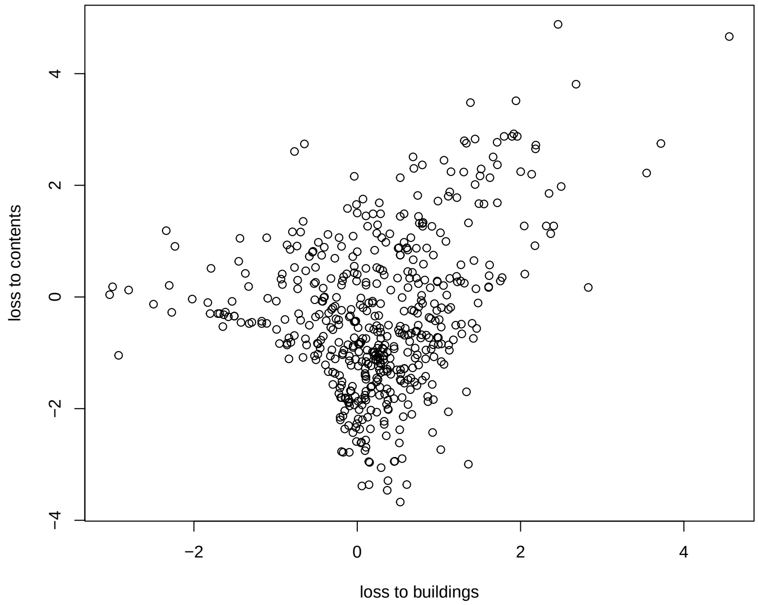

Figure 3.4: Scatter Plot of `loss to buildings` vs. `loss to contents` for the `Danish` data set

Table 3.1: Summary of `loss to buildings` and `loss to contents` for the `Danish` data set.

Variable	Minimum	Maximum	`Q1`	Median	`Q3`	Mean	Skewness	Kurtosis
`loss to buildings`	0.048	95.168	0.838	1.281	2.111	2.159	13.589	231.524
`loss to contents`	0.025	132.013	0.291	0.603	1.524	2.371	11.381	157.842

Chapter 4

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 selection criteria , known as AIC and BIC. The formula of the model selection 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.

4.1 Model Results

Table 4.1 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 4.1: 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

Table 4.2 presents the parameter estimates (marginal model parameters and copula parameter) of the bivariate composite distribution as well as existing bivariate distributions for the `Danish` fire loss data set.

Table 4.2: Parameter estimates of competing distributions for the `Danish` dataset

Parameter	E-ETP	BL	BMPI	BB
$\lambda^{(1)}$	0.6139159	-	-	-
$\alpha^{(1)}$	1.733798699	-	-	-
$\beta^{(1)}$	0.005574184	-	-	-
$\theta^{(1)}$	7.975004616	-	-	-
$\lambda^{(1)}$	1.576737	-	-	-
$\alpha^{(2)}$	0.95672468	-	-	-
$\beta^{(2)}$	0.02348697	-	-	-
$\theta^{(2)}$	1.22266390	-	-	-
δ	1.4	-	-	-
a_1	-	3.086	0.2579	0.9649
a_2	-	0.2274	0.9874	1.3341
p	-	0.4048	1.9874	1.3687

4.2 Likelihood Ratio Test

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

$$\Lambda = \frac{\sup_{\theta \in \Theta_0} L(\theta)}{\sup_{\theta \in \Theta} L(\theta)} \tag{4.1}$$

Now , we know that $-2 \log \Lambda \sim \chi^2$ with degrees of freedom equal to the difference in the number of parameters.

In our study , 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. 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 :

$$H_0 : \delta = 1 \text{ vs } H_1 : \delta > 1 \tag{4.2}$$

After calculating the likelihood values under the null and alternative hypotheses, we will compute the test statistic $-2 \log \Lambda$ according to the (4.1). The critical value for the chi square distribution will be determined using the appropriate degrees of freedom, which is equal to the difference in the number of parameters between the null and alternative models , that is the critical value of chi square with 1 degree of freedom. 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.

After seeing the Table 4.2 of parameter estimates, the estimated value copula parameter δ is 1.4. So, we can say that the dependency between two type of claims encountered by our proposed model.

4.3 Analysis of Dependence

To determine the goodness of fit of the fitted distribution in terms of dependence modelling, we give the Kendall's tau (τ), the upper tail dependence λ_U values for the proposed bivariate composite distributions obtained using Joe copula. The empirical value of Kendall's tau for the `Danish` fire loss data set is 0.085 and the empirical value of the upper tail dependence λ_U for the `Danish` fire loss data set is 0.198.

Table 4.3: Dependence measures of the bivariate composite distribution for the `Danish` fire loss data set

Model	τ	λ_U
Empirical	0.085	0.198
E-ETP	0.184	0.359

After seeing Table 4.3 for the `Danish` fire loss data set, Kendall's tau (τ) for the bivariate composite E-ETP distribution is 0.184 and λ_U for the bivariate composite E-ETP distribution is 0.359 , which is close to empirical value but our proposed bivariate composite model slightly over estimates the dependence between two variables in the upper tail.

Chapter 5

Conclusion

Insurance dataset tends to have extreme values which resulted a heavy tail. The frequency of extreme values maybe very less but it create nuisance in data analysis and which results wrong predictive analysis. To encounter this problem in this study , we have used a new approach called "tail tempering". Tempered Pareto distributions address the issue of excessive probability mass in the extreme tail by introducing a parameter that modifies the tail behaviour. This parameter essentially "temper" or adjust the tail of the distribution, reducing the probability assigned to extreme values. This makes tempered Pareto distributions more suitable for modelling data where the pure power law distribution might not be appropriate due to its overemphasis on extreme events. Now to model two types of claims simultaneously , we introduced a Joe copula based bivariate composite Exponential-Exponentially tempered Pareto distribution. To obtain the upper mentioned advantages of tempered Pareto , we have took the Exponentially tempered Pareto distribution for the tail part of our distribution.

To model two types of claims simultaneously which has some upper tail dependencies , Joe copula is the best choice. We fit our proposed Joe copula based bivariate E-ETP composite model on `Danish` fire loss data set. The numerical results shows that our model significantly outperforms the traditional existing bivariate models. The fruitful line of further research could be to utilise the proposed distribution for the response variable to employ regression framework by utilising several covariates available along with claim amount

Appendix

R codes

```
rm(list = ls())


library(fitdistrplus)
dd=data(danishmulti)
dd
str(danishmulti)
x_data=data.frame(danishmulti$Building,danishmulti$Contents)
x_data[,1]=danishmulti$Building
cor(x_data$danishmulti.Building,x_data$danishmulti.Contents)
dim(danishmulti$Building)
x_data[,1]=as.numeric(x_data[,1])
summary(x_data[,1])
x_data$danishmulti.Building
class(danishmulti$Building)


write.csv(danishmulti,"saikatdanish")
get.wd()
getwd()
x_data[1]=as.numeric(x_data[1])
x_data[2]=as.numeric(x_data[2])

######################################################################
data12=read.csv(file.choose(),header = T)
```

```
str(data12)
cor(data12$Building,data12$Content)
#########################

x_data=data.frame(danishmulti$Building,danishmulti$Contents)
ind <- which(x_data$danishmulti.Building >0 & x_data$danishmulti.Contents >0  )
summary(ind)
claim.nozero <- x_data[ind,]
summary(claim.nozero)
cor(claim.nozero[,1],claim.nozero[,2],method = "kendall")
cor(claim.nozero[,1],claim.nozero[,2])
length(claim.nozero[,1])
quantile(claim.nozero$danishmulti.Building)
plot(log(claim.nozero[,1]),log(claim.nozero[,2]))
hist(claim.nozero[,1], breaks = 130, main = '',
     xlab = 'Loss to Buildings',xlim = c(0,95))
hist(claim.nozero[,2],,breaks = 130,main = '',xlab = 'Loss to Contents')
library(plot3D)
z <- table(claim.nozero[,1], claim.nozero[,2])
hist3D(z=z, border="black")

library(VGAM);
library(stats4);            #for the function 'mle''
# --------------- defining exponential and
exponentially tempered Pareto distribution ----------- #
```

```
pexponential = function(x, para){
  lmd = para[1];
  return( ifelse( x>0, (lmd*exp(-(lmd*x))), 0 ) );
}
dexponential = function(x, para){
  lmd = para[1];
  return( ifelse( x>0, (1 - exp(-(lmd*x))) , 0 ) );
}
pexptempareto = function(x,para){
  alpha = para[1];  beta = para[2];  theta = para[3];
  return( ifelse( x>0, ((theta^alpha) * exp(beta * theta) *
         x^(-(alpha + 1)) * exp(-(beta * x)) * (alpha + (beta * x))) , 0 ))
}
dexptempareto = function(x,para){
  alpha = para[1];  beta = para[2];  theta = para[3];
  return( ifelse( x>0, (1 - ((theta^alpha)*exp(beta * theta) *
           x^(-(alpha)) * exp(-(beta * x)))) , 0 ))
}
Quietmode = "ON";
#-------------------------------------------------------
#Main computation functions
f1_truncated = function(x, para, splitpt){
  return ( f1(x, para)/Bigf1(splitpt, para) );
}
f2_truncated = function(x, para, splitpt){
  return ( f2(x, para)/(1-Bigf2(splitpt, para)) );
```

```
}
composite_f = function(x, para1, para2, r_theta){
  #The composite has parameters #para1 + #para2 + 2 (r and theta) in total
  #Continuity condition gives expression of r, so we have #para1 + #para2 + 1
  r = r_theta[1];
  splitpt = r_theta[2];
  if (x < splitpt){
    return ( r * f1_truncated(x, para1, splitpt) );
  }else{
    return ( (1-r) * f2_truncated(x, para2, splitpt) );
  }
}
composite_Bigf = function(x, para1, para2, r_theta){
  #The composite has parameters #para1 + #para2 + 2 (r and theta) in total
  #Continuity condition gives expression of r, so we have #para1 + #para2 + 1
  r = r_theta[1];
  splitpt = r_theta[2];
  if (x < splitpt){
    return ( r * Bigf1(x,para1)/Bigf1(splitpt, para1) );
  }else{
    return ( r + (1-r) * (Bigf2(x,para2)-Bigf2(splitpt,para2))
              /(1-Bigf2(splitpt, para2)) );
  }
}
ETP_complete_para = function(para){
  #Remapping
```

```
  alpha = para[1]; beta = para[2]; theta= para[3];
  #Complete the parameter set
  splitpt = theta;
  if(splitpt<=0){
    lmd =NA;
  }else{
    lmd = (1/theta)*(alpha + (alpha + (theta*beta))^2);
  }
  #Remapping
  para1 = c(lmd);    para2 = c(alpha, beta, theta);
  #Compute the mixture parameters, the mixing weight and the splitpt.
  r = f2(splitpt, para2)*Bigf1(splitpt, para1) / (f2(splitpt, para2)*
      Bigf1(splitpt, para1) + f1(splitpt, para1)*(1-Bigf2(splitpt, para2)) );
  r_splitpt = c(r, splitpt);
  #Return list of 3 vectors.
  return( list(para1, para2, r_splitpt) );
}
ETP_para_condition = function(para1, para2, r_splitpt){
  #Overall NA check.
  if (is.all_not_na(c(para1,para2,r_splitpt))==FALSE){return(FALSE);}
  #Now value check.
  #condition for parameters in the first distribution
  lmd = para1[1];
  #condition for parameters in the second distribution
  alpha = para2[1];  beta = para2[2];  theta = para2[3];
  #condition for r and splitpt
```

```
  r = r_splitpt[1]; splitpt = r_splitpt[2];
}
name_list = c("pexptempareto","dexptempareto", "pexponential", "dexponential",
              "ETP_para_condition","ETP_complete_para");
#Assignment of functions
f1=getFunction(name_list[3]);      Bigf1=getFunction(name_list[4]);
f2=getFunction(name_list[1]);      Bigf2=getFunction(name_list[2]);
para_condition = getFunction(name_list[5]);
complete_para = getFunction(name_list[6]);
estimation=function(x_vec)
{
  NLL = function(para, x_vec){
    cp = complete_para(para);
    para1 = as.vector( cp[[1]] );
    para2 = as.vector( cp[[2]] );
    r_splitpt = as.vector( cp[[3]] );
    #  if( !para_condition(para1, para2, r_splitpt) ){ return(NA) };
    x_vec = sort(x_vec);
    L = 0;
    for (x in x_vec){
      L = L + log( composite_f(x, para1, para2, r_splitpt));
    }
    return(-L);
  }
  xx1=nlm(NLL,x_vec=claim.nozero[,1],c(1.733798699,0.005574184,7.975004616)
         ,hessian = TRUE)
```

```
  xx2=nlm(NLL,x_vec=claim.nozero[,2],c(0.95672468 ,0.02348697,1.22266390)
          ,hessian = TRUE)
  parameter1= ETP_complete_para(para=c(xx1$estimate[1],xx1$estimate[2]
                ,xx1$estimate[3],xx1$estimate[4]))
  parameter2= ETP_complete_para(para=c(xx2$estimate[1],xx2$estimate[2],
                xx2$estimate[3],xx2$estimate[4]))
  para11=as.vector(parameter1[[1]])
  para12=as.vector(parameter1[[2]])
  r_theta1=as.vector(parameter1[[3]])
  para21=as.vector(parameter2[[1]])
  para22=as.vector(parameter2[[2]])
  r_theta2=as.vector(parameter2[[3]])
  par <- list(para11 = para11,para12 = para12, r_theta1 = r_theta1,
         para21=para21,para22=para22,r_theta2=r_theta2)
  n=length(claim.nozero[,1])
  # The fitted CDF
  #u1 <- sapply(1:n,function(k) composite_Bigf(claim.nozero[,1][k],
        para11, para12, r_theta1)) #MTPL
  #u2 <- sapply(1:n,function(k) composite_Bigf(claim.nozero[,2][k],
        para21, para22, r_theta2)) #MTPL
  u1 <- sapply(1:n,function(k) composite_Bigf(ss[,1][k],
       para11, para12, r_theta1))  ##DANISH
  u2 <- sapply(1:n,function(k) composite_Bigf(ss[,2][k],
       para21, para22, r_theta2))  ##DANISH
  return(c(par,list(u_fit =cbind(u1,u2),value = c(xx1$minimum,xx2$minimum))))
}
```

```
ss=cbind(claim.nozero[,1],claim.nozero[,2])
est=estimation(ss)
est
library(rvinecopulib)
library(cramer)
estimation_cp <- function(u_fit){
  fit_cop <- bicop(u_fit, family_set = 'joe',var_types = c("c", "c"))
  return(list(par = fit_cop$parameters,value = c(fit_cop$loglik)))
}
copula_parameter=estimation_cp(est$u_fit)
fit_cop <- bicop(est$u_fit, family_set = 'joe',var_types = c("c", "c"))
summary(fit_cop)
contour(fit_cop,col="red")
summary_model <- function(mar_fit,cop_fit,nobs){
  npar <- 7
  loglike <- sum(mar_fit$value) - cop_fit$value
  AIC <- 2*loglike + 2*npar
  BIC <- 2*loglike + log(nobs)*npar
  return(c(loglike,AIC,BIC))
}
summary_model(est,copula_parameter,nrow(claim.nozero))
sum(est$value)
sum(est$value) - copula_parameter$value
library(VineCopula)
BiCopPar2Tau(family = 6 , par = 1.4)
BiCopPar2TailDep(family = 6 , par = 1.4)
```

Bibliography

[1] Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19(6), 716–723.

[2] Aradhye, G., Tzougas, G., & Bhati, D. (2024). A Copula-Based Bivariate Composite Model for Modelling Claim Costs. *Mathematics*, 12(2), 350

[3] 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.

[4] Cooray, K. & Ananda,M.M. (2005). Modelling actuarial data with a composite lognormal-Pareto model. *Scandinavian Actuarial Journal*, 2005(5), 321–334.

[5] Ciumara, R. (2006). An actuarial model based on the composite Weibull-Pareto distribution. *Math. Rep.* , 8, 401–414

[6] Joe, H. (1997). Multivariate models and dependence concepts (1st ed.). *CRC Press: New York, NY, USA.*

[7] Joe, H. (2005). Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis*, 94, 401-419.

[8] Meerschaert, M. M., Roy, P., & Shaoz, Q. (2011). Parameter estimation for exponentially tempered power law distributions. *Communications in Statistics - Theory and Methods*, 40(11), 2007-2033.

[9] Nelsen, R. B. (2006). An introduction to copulas. *Springer: Berlin/Heidelberg, Germany.*

[10] P. G. Sankaran & Debasis Kundu. (2014). On A Bivariate Pareto Model. *Statistics: A Journal of Theoretical and Applied Statistics*, 48(2), 241-255.

[11] Scollnik, D. (2007). On composite lognormal-Pareto models. *Scandinavian Actuarial Journal*, 2007, 20–33.

[12] Scollnik, D. P. & Sun, C. (2012).modelling with Weibull-Pareto models. *North American Actuarial Journal*, 16, 260–272.