SPANISH JOURNAL OF STATISTICS
Vol. 7 No. 1 2025, Pages 9–38
doi:https://doi.org/10.37830/
SJS.2025.1.02

REGULAR ARTICLE

Copula based inference for certain types of actuarial datasets: A brief survey

Indranil Ghosh[1], Carman Greenway[2], Hannah Powers[3], and Brandon Kelly Kortan[4]

[1]University of North Carolina, Wilmington, ghoshi@uncw.edu
[2]University of North Carolina, Wilmington, cdg8092@uncw.edu
[3]University of North Carolina, Wilmington, hep6882@uncw.edu
[4]University of North Carolina, Wilmington,bkk6389@uncw.edu

Received: May 26, 2025. Returned: -. Revised: -. Accepted: October 6, 2025.

Abstract: Copula is a useful tool for modeling bivariate/multivariate dependency structures among others. In this paper, we aim to study various types of dependence indicated by well-known measures of dependence such as Spearman's ρ, Kendall's τ, and Blomqvist's β etc., for certain types of actuarial datasets, which is obtained from the CAS datasets in R software package. Although our primary focus is on the insurance claim datasets, the adopted copula-based procedure can be mimicked in other types of actuarial datasets and other domains as well. On using the CDA vine package in R, we find the best fitted bivariate copula for a given dataset, and subsequently study various structural properties of the derived best fitted bivariate copula. The adopted strategy can be envisioned in identifying and exploring multivariate dependence via Vine copula strategy which will be discussed in a separate article.

Keywords: Bivariate copula; Measures of Association; Kendall's tau; Copula fitting

MSC: 62H05, 62H20, 91G70

1 Introduction

Copula, Latin for "link, tie, or bond," in the mathematical world, is a function that allows us to combine univariate distributions to obtain a joint distribution with a particular dependency structure, according to the work of Durante and Sempi (2016). While copulas were first formally introduced to the mathematical and statistical world in 1959, the idea was not a foreign concept, as primitive versions of copulas were seen in Wassily Hoeffding's works as early as 1940. Hoeffding established possible bounds for these functions and studied measures of dependence invariant under strictly increasing transformations. Maurice Fréchet's work in 1951 also had ideas similar to the copula present, as his work on bounds for joint distributions with given marginals laid the foundation for copulas.

Copulas, introduced by Abe Sklar in 1959, are a fundamental resource in defining dependency structure between random variables. Sklar's theorem allows for a transformation of dependency structure to a simpler form involving a joint uniform cumulative distribution function in a transformed random variable and marginal probability density functions. The idea is to use the probability integral transform

$$Y = \int_{-\infty}^{x} f_U(u)du = F_X(x),$$

for some random variable X to define a uniformly distributed random variable Y. While the original theorem makes no mention of such a condition, it is often preferable to use continuous marginal cumulative distribution functions $F_{X_n}(X_n) = U_n$, so that each $U_n \sim U(0,1)$. Using Y as dummy variable, one may write

$$F_Y(y) = \mathbb{P}[Y \leq y] = \mathbb{P}[F_X(X) \leq y] = \mathbb{P}[X \leq F_X^{-1}(y)] = F_X(F_X^{-1}(y)) = y.$$

The dependence between two random variables, say X and Y, is completely described by the joint distribution function $F_{X,Y}(x,y))$. The major motivation of separating $F_{X,Y}(x,y))$ in two parts: the one which describes the dependence structure, and the other one which describes the marginal behavior, leads to the concept of copula. To every bivariate distribution function $F_{X,Y}(x,y))$, with continuous marginals $F_X(x)$ and $F_Y(y)$, corresponds to a unique function
$C : [0,1] \times [0,1] \to [0,1]$, such that

$$F_{X,Y}(x,y)) = C(F_X(x), F_Y(y)),$$

for $(x,y) \in (-\infty,\infty) \times (-\infty,\infty)$. Extension of this definition to a higher dimension (say, in d - dimension, where $d \geq 3$) can certainly be envisioned. This could be summarized as follows.

A d-dimensional copula is a function $C : [0,1]^d \to [0,1]$ that satisfies:

(i) $C(u_1, u_2, \ldots, u_{i-1}, 0, u_{i+1}, \ldots, u_d) = 0$ for all $1 \leq i \leq d$, and $0 \leq u_k \leq 1$ for $k = 1, \ldots, d$ with $k \neq i$.
(ii) $C(1, \ldots, 1, u, 1, \ldots, 1) = u$ for all $0 \leq u \leq 1$, in each of the d arguments. That is, the copula equals u when one argument is u and all others are 1.
(iii) For any vectors $\mathbf{s} = (s_1, \ldots, s_d)$ and $\mathbf{w} = (w_1, \ldots, w_d)$ such that $s_i \leq w_i$ for all $i = 1, \ldots, d$, the following inequality holds:

$$\sum_{i_1=1}^{2} \cdots \sum_{i_d=1}^{2} (-1)^{i_1+\cdots+i_d} C(u_1^{(i_1)}, \ldots, u_d^{(i_d)}) \geq 0,$$

where $u_j^{(1)} = s_j$ and $u_j^{(2)} = w_j$ for each $j = 1, \ldots, d$. That is, the copula is non-decreasing in d dimensions.

The copula is then the joint cumulative distribution function $C(\vec{u})$, where $\vec{u} = (u_1, u_2, \ldots, u_d)$ of the transformed random variables u and contains all dependency information of the initial random variables x_d. According to Sklar's (1959) theorem,

$$C(\vec{u}) = C(F_{X_1}(x_1), ..., F_{X_d}(x_d)) = \mathbb{P}[U_1 \leq u_1, ..., U_d \leq u_d],$$

such as elliptical, Archimedean, and Pickands.

Sklar's formal introduction of copulas provided a powerful tool to model and analyze dependency structures between random variables, regardless of marginal distributions. His theorem is not limited to the mathematical world, as fields such as finance, hydrology, and engineering all need an understanding of the joint behavior of random variables through certain copulas, see, Nadarajah (2017) for pertinent details in this context.

The Gaussian copula family of models was first introduced to the financial field by Oldrich Vasicek, see, for pertinent details, see, Trivedi (2007) and the references cited therein. He first introduced the copula model at a corporate loan firm to help reduce concentration of risk in specific geographic regions and industries. From there, the use of copulas in finance increased heavily in 2000, when David X. Li applied them to model default correlations in credit risk. This work skyrocketed the popularity of the Gaussian copula model to price collateralized debt obligations (CDOs) and assess the risk of credit portfolios. However, the Gaussian copula, like all other types of copulas, has limitations that users must be mindful of, particularly in capturing tail dependencies. The overuse of the Gaussian copula has been linked to the 2007-2008 financial crisis because of failure to predict extreme joint movements, which led to significant financial loss. This shows the need to determine the correctness of assumptions underlying the use of specific copulas. Next, we discuss another type of copula which is useful in financial risk modeling, namely, the Gumbel copula.

The Gumbel Copula, named after Emil Julius Gumbel, a German mathematician known for his contributions to the Extreme Value Theorem, is a part of the Archimedean copula tree that specializes in deciphering the dependence structures between random variables, mostly focusing on upper tail dependence, scenarios where extreme high values in one variable are associated with extreme high values in another, see, Tinungki (2023) and the references cited therein for an extensive discussion on this matter. The Gumbel copula has been applied to various fields, such as risk management in the financial field, actuarial science, modeling insurance risks, and hydrology, modeling joint distributions in events such as extreme rainfall and river discharge, helping to assess flood risks among others. The Gumbel Copula also has its limitations. While it is great for modeling the upper tail dependence, it lacks a similar strength in magnitude for modeling the lower tail dependence, which is the tendency of variables to jointly exhibit low values. In addition, there is another family (alias collection of copulas) which are quite useful in actuarial data application, popularly known as Archimedean copula.

The term "Archimedean Copula" was first introduced in statistical literature by Christian Genest and Jock MacKay in their work, see, Genest (1986). The most common Archimedean are the Gumbel Copula, as previously discussed, the Clayton Copula, the Frank Copula, the Ali-Mikhail-Haq (AMH) Copula, the Joe Copula, and the Nelsen Copula etc. All of these are designed to model dependencies between random variables, with each copula model specializing in different fields with different tail dependency modeling.

Although not that much of use (as compared to a Gaussian copula), the t-Copula, also known as the Student's t-Copula, is another copula used to model dependence structures between multiple random variables. The term t-Copula first gained notoriety in 2005, through the work of Stefano Demarta and Alexander J. McNeil. In their work, they analyzed the t-Copula's properties and applications, which contributed significantly to its adoption in statistical modeling. The t-Copula is heavily applied in the risk management and financial fields, especially in portfolio management. The t-copula can also be divided into two special cases, namely, the Skewed t-Copula and the Grouped t-Copula. The Skewed t-Copula extension introduces asymmetry into the dependence structure, which allows for different behaviors in the upper and lower tail structures, which is extremely useful in the financial field. Meanwhile, the Grouped t-Copula model allows for varying degrees of freedom parameters for different groups of variables, which allows us to get a more nuanced modeling of dependencies in heterogeneous data sets, see, Venter(2002). The t-copula, and their extensions have drawbacks, especially in terms of computation. In the next, we discuss the motivation to carry out this project.

This paper leverages the copula models available in the VineCopula package in R to explore about the dependency structure among concomitant variables present in an various types of insurance data. As the misuse of copula models was partly to blame for the financial collapse of 2008, we concern ourselves with tail dependence, or dependence of extreme events. An example of the importance of tail dependence in insurance data is as follows: Suppose that there is a strong upper tail dependence of average charges to a health insurer and the number of stays at the hospital. This would imply that the overall cost is much higher. If the insurer finds that many of its policyholders have frequent hospital visits, this could pose a significant risk to the insurer. The copula fit to the data, as in this example, gives important insights into dependency and tail dependence so that the appropriate decision(s) can be made.

Our work intends to fit bivariate copulas to understand these tail risks, as well as on the overall dependence. The remainder of this paper is organized as follows. In Section 2, we describe various dependence measures and the associated copula version of it. Section 3 briefly outlines various datasets from insurance domains and we provide adequate rationale on the selection of these data sets as well as variable selection. In Section 4, we provide the details of the copula fitting to each of these datasets. Finally, some concluding remarks are made in Section 5.

2 Dependence Measures

In the study of bivariate dependence, there are multiple measures specific to each types of different scenarios that an experimenter/researcher can envision. In this paper, we focus on three distinct types of popular dependence measures, each of which can be obtained via a copula which are:(a) Pearson's correlation coefficient; (b) Kendall's Tau, and (c) Blomqvist's β.

- Pearson's correlation coefficient is given by

$$\rho = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y},$$

  where $\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$, and where $|\rho| \leq 1$.
  In terms of copula, Spearman's ρ_S can be written as

$$\rho_S = 12 \iint_{[0,1]^2} C(u, v)\, dC(u, v) - 3 = 12 \iint_{[0,1]^2} \left[C(u, v) - uv\right]\, du\, dv.$$

  However, Pearson's correlation is highly sensitive to outliers as it involves central moments, as a single extreme value can distort the coefficient and produce a misleading output. Pearson's correlation assumes the relationship between two variables is linear, and does not capture nonlinear associations. Additionally, it assumes that the data is approximately Gaussian which may produce unreliable results if the data is skewed or otherwise non-Gaussian. In cases where assumptions are violated or the relationship is non-linear, different methods should be used, such as Kendall's τ. In the next, we discuss some useful details on Kendall's τ.
- Kendall's τ, developed by Maurice Kendall in 1938, is a non-parametric measure of the strength and direction of the relationship between two variables. Unlike Pearson's correlation, which is designed for linear relationships, Kendall's Tau is useful for ordinal (ranked) data or when the relationship between variables is nonlinear. Kendall's Tau compares the relative ordering of pairs and focuses primarily on counting the number of concordant and discordant pairs. A pair is said to be concordant if:

$$(X_1 - X_2)(Y_1 - Y_2) > 0,$$

where both variables move in the same direction (both increasing or both decreasing). A pair is said to be discordant if:

$$(X_1 - X_2)(Y_1 - Y_2) < 0,$$

where one variable increases while the other decreases, and (X_1, Y_1) and (X_2, Y_2) are two pairs of random variables from a joint distribution function. If a pair holds the same rank, it is considered tied.

There are two variations of Kendall's Tau: Tau-a and Tau-b. Tau-a is used when there are no tied ranks in the data, while Tau-b is the most commonly used version when ties are present. The formula for Kendall's Tau is given by:

$$\tau = \frac{C - D}{\sqrt{(C + D + T_1)(C + D + T_2)}},$$

where C is the number of concordant pairs, D is the number of discordant pairs, and T_1, T_2 are the number of tied pairs. Alternatively, τ can be written as:

$$\tau = \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) > 0] - \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) < 0]$$

Let X and Y be continuous random variables with copula C. Then Kendall's tau is:

$$\tau = 4 \iint_{[0,1]^2} C(u, v)\, dC(u, v) - 1$$

The result ranges from -1 (perfect negative rank correlation) to $+1$ (perfect positive rank correlation), with 0 indicating no correlation. Kendall's Tau is widely used in the social sciences, medical research, and economics. However, it has limitations, such as a computational complexity of $O(n^2)$ and sensitivity to ties in data. Next, we consider the role of Blomqvist's β as a measure of dependence.

- Blomqvist β, introduced by Nils Blomqvist in 1950, is a non-parametric measure of statistical dependence. It measures the strength of association between two variables based on medians rather than considering the full distribution ranks. Unlike Pearson's correlation coefficient, which measures linear relationships, or Kendall's Tau, which assesses monotonic relationships, Blomqvist's β focuses on how observations cluster around their median values. Blomqvist's β is particularly useful when dealing with ordinal data, non-Gaussian distributions, and datasets with outliers. It is defined as:

$$\beta = 4\mathbb{P}\left[X > M_X, Y > M_Y\right] - 1,$$

where M_X and M_Y are the medians of X and Y, respectively, and $P\left(X > M_X, Y > M_Y\right)$ is the probability that both variables fall above their medians. The copula based expression of Blomqvist β is defined as

$$\beta = 4C\left(\tfrac{1}{2}, \tfrac{1}{2}\right) - 1.$$

Blomqvist's β ranges from -1 (strong negative association) to $+1$ (strong positive association), with 0 indicating no association. While less commonly used, Blomqvist's β is applied in finance for skewed financial returns, Denuit (2005); medical research for disease rates, Nooraee (2014), and social sciences for survey analysis, Agresti (2010). However, it has weaknesses, including loss of information due to median-based classification, sensitivity to ties in data, and poor performance for weak dependencies. All considered, it is necessary to use either Blomqvist's β or other correlation measures as demanded by the application.

Looking strictly at the financial field, copulas enable modeling of dependencies between different assets, capturing tail dependencies that traditional correlation measures might miss. This allows for a more accurate estimation of risk measures such as Value-at-Risk (VaR) and Conditional Value-at-Risk (CVar), leading to better-informed investment decisions. The flexibility of copulas allows for the modeling of complex, non-linear dependencies, and tail risks, which are critical in stress testing and scenario analysis. By selecting appropriate copula functions, risk managers can simulate various market conditions and assess potential impacts on portfolios. Some of the most commonly used copulas are the Gumbel copula for extreme distributions, the Gaussian copula for linear correlation, and the Archimedean copula and t-copula for dependence in tails.

Our goal is to model and examine the strength and direction the dependency structure(s) of concomitant random variables that are selected from the CAS datasets for an efficient way of creating guidelines for a better decisions pertaining to insurance pricing or so. This paper will use Pearson's correlation, Kendall's Tau, and Blomqvist's β to analyze the relationship between bivariate distributions in terms of the best fitted copula for each of those illustrative data sets and using copula based definition of those dependency measures. Choosing the most accurate statistical model is essential for extracting relevant information from the data. It is to be noted that the Pearson's correlation measure is best for identifying the strength of linear relationships, Kendall's τ is preferred for monotonic but nonlinear relationships, and the Blomqvist's β is suitable when median-based dependencies matter. Next, we provide some useful details on the real data sets selected from the CAS data sets in `R`.

3 Real Data Application

Swedish motor insurance data is provided by the CAS datasets repository on Github for public use (link: https://cas.uqam.ca/pub/web/CASdatasets-manual.pdf), especially with the Computational Actuarial Sciences textbook. This insurance data was collected in 1977 by the Swedish Committee on the Analysis of Risk premiums. The data of concern for this repository are *claims*, the number of claims by each policyholder; *insured*, the number of years of the policy; and *payment*, the sum of payments made by the policyholder. We will use bivariate copulas to model the bivariate dependency structures of both *claims* and *insured* to *payment*. We chose to model *claims* against *payment* to describe tail risks associated with the number claims made to payment, as this is a concern for financial risk purposes. We chose to model *insured* to *payment* to determine how payments toward the policy decide how long the policy is held, or vice versa.

Next, term life insurance data is also provided by the CAS datasets repository. This is data from the United States in a survey from survey of consumer finances. It is a nationally representative random sample of 500 households. The data of concern for this repository are *income*, the income of the family, and *face*, the amount of the payout in the event of death. We chose to model *income* to *face* to see how the choices of the value of the policy are decided by the income of the family.

Next, we consider Medicare Hospital Costs which is a dataset from the Health Care Financing Administration, Bureau of Data Management and Strategy. It is mainly for use with Regression Modeling with Actuarial and Financial Applications by Frees (2009). We look specifically at the Medical Expenditure Panel Survey (MEPS), which was a nationally representative survey conducted by the U.S. Agency of Health Research and Quality. The data of concern in this dataset are *income*, the median income of families, and *borrowed*, the amount borrowed on the life insurance policy. We chose to model *income* to *borrowed* to determine the relationship between the income of the family and how much it borrows from its life insurance policy.

3.1 Variable Selection

To model the above mentioned example data sets via a bivariate copula, we first revisit the Swedish motor insurance data set and a US life insurance data set, the details on this data is given earlier in Section 2. This data set has several concomitant variables. However, we describe below (in the form of several models) variable selection from this data set and study the associated dependence structure via a best fitted copula.

1. **Model 1:** We consider Insured and Payment as two concomitant variables and we want to explore the strength and direction of the dependency between them. By implementing a copula based approach (via the CDA vine copula package in R), the best fitted bivariate copula in this case is found to be the following: BB6 copula with par1= 1.59, par2 = 2.81, τ = 0.73. Some useful structural details of this copula is provided in Section 4.
2. **Model 2:** Involving components Claims and Payment. Here, the best fitted bivariate copula is the bivariate t copula with par1= 0.98, par2 = 2, τ = 0.87. Some useful structural details of this copula is provided in Section 4.

After analyzing the individual scatter plots, we decided upon further analysis by comparing the contour plots, which can be found below in Figure 3. The left contour plot of Figure 3 shows very tight contour lines, once again suggesting a high dependence relationship between claims and payment. The lack of curvature in the corners of the contour graph suggests low-tail asymmetry. For the right contour plot, we see very tight contour lines in the top right and more loose in the bottom left, a clear sign of upper tail dependence. The sharper corner in the top right of the graph could suggest that the more extreme values of insured have a more closely associated extreme value of payment values, whereas lower values of insured may have more variability for the payment values. When modeling via the copulas, we found that the Insured and Payment variables had a Kendall's $\tau = 0.73$, which indicates a relatively strong correlation between the two variables. The best fitted copula is the BB6 copula, which is a member of the Archimedean family. The BB6 copula is particularly well suited for capturing strong tail dependencies, meaning that extreme values in one variable are likely to correspond to extreme values in the other. This makes it an appropriate choice for modeling insurance related datasets where large payments may be influenced by the number of insured individuals. The BB6 copula is parameterized by θ (par 1) and δ (par 2) which control the dependence structure. The θ parameter governs the overall dependence structure strength, while δ determines tail dependence, with larger values indicating stronger tail dependence. The CDF graph in figure 4 shows the cumulative probability over the unit square $[0, 1]^2$, based on the BB6 copula. The steep climb towards the corner suggests strong positive dependence, particularly in the upper tail. This kind of sharp slope usually indicates that more extreme high values of the variables tend to occur jointly more frequently than under independence, confirming our previous findings. As for the PDF graph, this shows the derivative of the CDF, which is a visualization of where the associated copula density is concentrated. There is a very sharp spike in the top right corner of (1,1), and almost flat everywhere else, confirming a strong upper tail dependence. This graph also captures the extreme co-movements, once again confirming that when one variable is high, the other variable will tend to be high too. For illustrative purposes, we provide the scatterplot between Claims and Payment in Figure 1 in the Appendix. From Figure 1, it appears that there exists a strong positive linear relationship. In Figure 4, we provide the pdf and the cdf for the BB6 copula with the estimated parameter values.

For the second model, the best fitted bivariate copula in this case is a Student's t-Copula, which is a member from the Elliptical family, and is particularly useful for capturing both linear and tail dependence. Using a Student's t-Copula allows for flexible correlation structures that better model financial and insurance datasets where extreme values can occur at the same time. The two key parameters for the Student's t-Copula are ρ (par 1) and ν (par 2). The ρ parameter represents the correlation between two variables, while ν, the degrees of freedom, controls the heaviness of tails, with smaller ν values indicating stronger tail dependence. The Student's t-Copula makes for a suitable choice for modeling extreme claims and payment behavior, as it allows for capturing risk concentration in the tails. In terms of the Claims and Payment copula, we found that the Kendall's τ value to be 0.87, which indicates a significantly strong correlation between the two variables. Figure 1 suggests a very strong, positive correlation, which confirms our high Kendall's τ of 0.87. As Claims increases, Payment tends to increase, with low variation. For lower claims, payment is more spread out, implying more noise or variability with how smaller claims are handled. For illustrative purposes, we provide the scatterplot between the two variables Insured and payment. Figure 2 exhibits a strong positive correlation. As Insured values increase, so do the payment values. These points are clustered very tightly in the top right corner, which is a sign of upper tail dependence, meaning high insured values are closely associated with high payment values. There is some higher vertical spread within the 0.25 to 0.75 range, suggesting some moderate variability within the dependence of the two variables. In Figure 7, we provide the pdf and the cdf of a bivariate t-copula with the estimated parameter values.

The tables below detail the summary of the goodness of fit, estimates of the model parameters etc. of each model. Note that all of these computations were performed in R using the `CDAvine package`.

Table 1: Dependence measures between variables for Swedish Motor data.

Swedish Motor/Model	X_1	X_2	Kendall's τ	Spearman's ρ
Model1	Claims	Payments	0.87	0.962
Model2	Insured	Payments	0.73	0.903

Table 2: Model diagnostics and goodness of fit statistics for the best fitted copula for the Swedish Motor data.

Swedish Motor Model	Best Fitted Copula	Parameter Estimates	AIC	BIC	Log Likelihood
Model1	T-Copula	$(0.98, 2)$	-7129.3	-7117.92	3566.65
Model2	BB6	$(1.59, 2.81)$	-4095.96	-4084.58	2049.98

3.2 Data Set 2

This data set, the US Term Life Insurance data set, was a survey with 500 household participants carried out in 2004 by the Survey of Consumer Finances (SCF) group. This included characteristics such as gender, age, marital status, education, etc. This data set is also publicly available on the

Wisconsin School of Business FreesBook-RMAFA website. Here, we consider two different models based on two sets of concomitant variables selected for this purpose.

1. **Model 3:** Involving components Income and Face. Here, the best fitted copula is given by Tawn type 1 with par1 = 1.84, par2 = 0.49, tau = 0.28.
2. **Model 4:** Involving components Income and Borrow CV Life Policies. Here, the best fitted bivariate copula is the bivariate Frank copula with par1= 1.59, par2 = 0.17, tau = 0.17.

As seen above in the Term Life Insurance copula between the variables Income and Face, we see the Kendall's Tau value is 0.28, which indicates a positive but very weak correlation, which at first glance indicates that the variables chosen have almost no impact on the life insurance policy these households would use/already have purchased. The Tawn type 1 copula, used to model this relationship, is an extension of the Gumbel copula that allows for asymmetric tail dependence. Unlike symmetric copulas such as the Gaussian or Clayton copulas, the Tawn type 1 copula introduces an additional asymmetry parameter, which enables it to better capture imbalances in dependency structure. This copula is characterized by θ parameter, which controls the overall strength of dependence, and is equal to 1.84. The second parameter, δ, which introduces asymmetry (one tail may be stronger than the other), is equal to 0.49. Given that the Kendall's Tau is very low, the copula structure confirms that there is minimal dependency between Income and Face in influencing life insurance policy choices. However, the choice of a Tawn type 1 copula suggests that there may still be some asymmetric tail dependence, meaning that individuals at extreme values of Income or Face might exhibit slightly different dependency patterns than those in the middle of the distribution.

Upon careful observation of the scatter plot and contour plots of Income and Face, which can be found in Figures 5 and 6 respectively, the scatterplot is much more diffuse than the previous ones. There is a very loose upward trend, which means that as the income values increase, the face values tend to increase, but it is not very tight and there is a large vertical spread across most income levels. The horizontal band of points at the bottom could suggest either a minimum face value for each policy given regardless of income, or a mass of ties in the data points when transformed into pseudo-observations. The contour plot also supports our previous summation of the scatter, as Figure 6 shows the contours are fairly elliptical and centered, which is a hallmark of weak symmetrical dependence. There is no strong skew or curvature, suggesting that there is no upper or lower tail dependence. The very slight elongation along the 45 degree line suggests a small positive correlation, meaning that increases in income tend to increase with face value, but very weakly. There are no strong outliers or asymmetry to speak of.

The tables below detail the results of each model.

Table 3: Dependence measures between variables for Term Life Insurance data

Term Life/Model	X_1	X_2	Kendall's τ	Spearman's ρ
Model1	Income	Face	0.28	0.365
Model2	Income	Borrow CV	0.17	0.210

3.3 Data Set 3

The third dataset we examined was Medicare Hospital Costs, where data was obtained from the Health Care Financing Administration, Bureau of Data Management and Strategy. This data set includes impatient hospital charges covered by the Medicare program for the years 1990-1995. This

Table 4: Model diagnostics and goodness of fit statistics for the best fitted copula for Term Life Insurance data

Term Life Model	Best Fitted Copula	Parameter Estimates	AIC	BIC	Log Likelihood
Model1	Tawn Type 1	$(1.84, 0.49)$	-112.75	-104.32	58.37
Model2	Frank	$(1.59, 0.17)$	-20.42	-16.21	11.21

included variables such as state name, total hospital charges, number of hospital stays, number of discharged, etc. This data set is available on the FreesBook-RMAFA website of the Wisconsin School of Business. This dataset involves components Charges and Number of Stays. On using the CDAVinepackage in R, the best fitted bivariate copula happens to be a Survival Gumbel with the following parameter choices: (par1 = 6.12, par2 = 0.84, τ= 0.84)

As indicated above in the Medicare data, the bivariate copula between the two seemingly concomitant variables, name ly the Charges and the Number of stays, we see the Kendall's τ= 0.84, which indicates a positive and strong correlation. This means that the total hospital charges covered by medicare has a strong, positive dependency with the number of hospital stays. The Survival Gumbel copula is a copula function derived from the Gumbel copula, which is known to capture a strong positive dependency in extreme values. Unlike the original Gumbel copula, which emphasizes the dependence on the lower tail, the survival version allows the copula to focus on the upper tail rather than the lower tail.

The tables below detail the results of each model.

Table 5: Dependence measures between variables for Medicare data.

Medicare Costs	X_1	X_2	Kendall's τ	Spearman's ρ
Medicare	Charges	Number of Stays	0.84	0.97

Table 6: Model diagnostics and goodness of fit statistics for the best fitted copula for Medicare data.

Medicare Costs	Best Fitted Copula	Parameter Estimates	AIC	BIC	Log Likelihood
Medicare	Survival Gumbel	$(6.12, 0.84)$	-895.61	-891.83	448.8

In the next, we discuss several useful structural properties related to each of the fitted bivariate copula beginning with the BB6 copula.

4 Structural properties of the fitted Copula

This section presents the analysis of certain structural properties of the copulas. We begin our discussion with the BB6 copula.

4.1 BB6 (Joe-Gumbel) Copula

The BB6 copula is an Archimedean copula, so that for generator $\phi(t)$,

$$C(u,v) = \phi^{-1}(\phi(u) + \phi(v)).$$

In this case, the generator and its inverse ϕ^{-1} are

$$\phi(t) =(-\log[1-(1-t)^{\theta}])^{\delta}, \theta > 0, \delta \geq 1;$$

$$\phi^{-1}(s) = 1-(1-\exp[-s^{1/\delta}])^{1/\theta}.$$

$$u \geq 0, v \leq 1, \theta \geq 1, \delta \geq 1.$$

Therefore, the expression of the BB6 copula is given by

$$C\left(u,v\right) = 1-\left\{1-\exp\left\{-\left[\left(-\log\left(1-\bar{u}^{\theta}\right)\right)^{\delta}+\left(-\log\left(1-\bar{v}^{\theta}\right)\right)^{\delta}\right]^{1/\delta}\right\}\right\}^{1/\theta}, \tag{1}$$

where $\bar{u} = 1-u$, and $\bar{v} = 1-v$.

- The BB6 copula is symmetric in the sense that for this copula $C\left(u,v\right) - C\left(v,u\right) = 0 \quad \forall (u,v) \in [0,1]^2$.
- Next, we determine the value of Blomqvist β for the BB6 copula based on the parameters of the Swedish Automobile insurance model with $\theta = 1.59$ and $\delta = 2.81$. Therefore, on using the formula for Blomqvist's β, the lower tail and upper tail dependence coefficients can be calculated using the same methodology that we used for the Frank copula. For the upper tail dependence coefficient, we obtain the following:

$$\begin{aligned}\lambda_U &= \lim_{u\uparrow 1} \frac{1-2u+1-\left(1-\exp(-[2(-\log(1-\overline{u}^{\theta}))^{\delta}]\right))^{\frac{1}{\theta}}}{1-u} \\ &\overset{H}{=} \lim_{u\uparrow 1} 2-2^{\frac{1}{\delta}}(1-u)^{\theta-1}\exp\left[2^{\frac{1}{\delta}}\log(1-(1-u)^{\theta})\right](1-\exp\left[2^{\frac{1}{\delta}}\log(1-(1-u)^{\delta}\right]))^{\frac{1}{\theta}-1} \\ &= 2-2^{\frac{1}{\theta\delta}}.\end{aligned}$$

Similarly for the lower tail dependence coefficient:

$$\begin{aligned}\lambda_L &= \lim_{u\downarrow 0} \frac{1-\left(1-\exp(-[2(-\log(1-\overline{u}^{\theta}))^{\delta}]\right))^{\frac{1}{\theta}}}{u} \\ &\overset{H}{=} \lim_{u\downarrow} 2^{\frac{1}{\delta}}(1-u)^{\theta-1}\exp\left[2^{\frac{1}{\delta}}\log(1-(1-u)^{\theta})\right](1-\exp\left[2^{\frac{1}{\delta}}\log(1-(1-u)^{\delta}\right]))^{\frac{1}{\theta}-1} \\ &= 0,\end{aligned}$$

which have been independently obtained in Ghosh et al. (2023).
Next, we consider the conditional copula of U given $V = v$ and vice versa.

– The conditional copula of U given $V = v$ will be

$$\begin{aligned}
&C_1 (u|V = v) \\
&= \frac{\partial C(u, v)}{\partial v} \\
&= \left\{1 - (1 - v)^{\theta}\right\}^{-1} \\
&\times \Bigg[(1 - v)^{\theta-1} \left(-\log\left(1 - (1 - v)^{\theta}\right)\right)^{\delta-1} \Bigg(\left(-\log\left(1 - (1 - u)^{\theta}\right)\right)^{\delta} \\
&+ \left(-\log\left(1 - (1 - v)^{\theta}\right)\right)^{\delta}\Bigg)^{\frac{1}{\delta}-1} \\
&\times \left(1 - \exp\left(-\left(\left(-\log\left(1 - (1 - u)^{\theta}\right)\right)^{\delta} + \left(-\log\left(1 - (1 - v)^{\theta}\right)\right)^{\delta}\right)^{\frac{1}{\delta}}\right)\right)^{\frac{1}{\theta}-1} \\
&\times \exp'\left(-\left(\left(-\log\left(1 - (1 - u)^{\theta}\right)\right)^{\delta} + \left(-\log\left(1 - (1 - v)^{\theta}\right)\right)^{\delta}\right)^{\frac{1}{\delta}}\right)\Bigg].
\end{aligned} \tag{2}$$

– Likewise, the The conditional copula of V given $U = u$ will be

$$\begin{aligned}
&C_2 (v|U = u) \\
&= \left[1 - (1 - u)^{\theta}\right]^{-1} \\
&\times (1 - u)^{\theta-1} \left(-\log\left(1 - (1 - u)^{\theta}\right)\right)^{\delta-1} \Bigg(\left(-\log\left(1 - (1 - u)^{\theta}\right)\right)^{\delta} \\
&+ \left(-\log\left(1 - (1 - v)^{\theta}\right)\right)^{\delta}\Bigg)^{\frac{1}{\delta}-1} \\
&\times \left(1 - \exp\left(-\left(\left(-\log\left(1 - (1 - u)^{\theta}\right)\right)^{\delta} + \left(-\log\left(1 - (1 - v)^{\theta}\right)\right)^{\delta}\right)^{\frac{1}{\delta}}\right)\right)^{\frac{1}{\theta}-1} \\
&\times \exp'\left(-\left(\left(-\log\left(1 - (1 - u)^{\theta}\right)\right)^{\delta} + \left(-\log\left(1 - (1 - v)^{\theta}\right)\right)^{\delta}\right)^{\frac{1}{\delta}}\right).
\end{aligned} \tag{3}$$

Observe that, one may use the conditional copula of U given $V = v$, given in Eq. (2) and to simulate from the propose BB6 copula as given in Eq. (1) using the following steps:

– Simulate and v_i and u_i^* from a standard uniform distribution.
– If $v_i \leq 1$, then solve $C_1(u|v_i) = u_i^*$.
– Repeat the previous two steps, say, n times to obtain independence and identically distributed realizations (u_i, v_i), for $i = 1, 2, \cdots, n$ from the BB6 copula as given in Eq. (1).

A similar algorithm can be elaborated to simulate from the BB6 copula based on the conditional copula of V given $U = u$ as given in Eq. (3).

Next, we consider few other useful structural properties of the BB6 copula that has not been discussed as of yet to the best of the knowledge by the authors.

1. **Proposition 1.** The BB6 copula defined in Eq. (1) is decreasing with respect to its' dependence parameter θ, i.e., if $\theta_1 < \theta_2$ then
$$C_{\theta_2}(u,v) \le C_{\theta_1}(u,v),$$
for all $(u,v) \in I^2 = [0,1] \times [0,1]$.
Proof.
Let us consider the partial derivative of Eq. (1) w.r.t. θ. We have
$$\begin{aligned}\frac{\partial C(u,v)}{\partial \theta} &= \left\{ \frac{\frac{\delta(1-u)^\theta \log(1-u)\left(-\log\left(1-(1-u)^\theta\right)\right)^{\delta-1}}{1-(1-u)^\theta} + \frac{\delta(1-v)^\theta \log(1-v)\left(-\log\left(1-(1-v)^\theta\right)\right)^{\delta-1}}{1-(1-v)^\theta}}{\theta\left((-\log(1-(1-u)^\theta))^\delta + (-\log(1-(1-v)^\theta))^\delta\right)} \right. \\ &\quad \left. - \frac{\log\left(\left(-\log\left(1-(1-u)^\theta\right)\right)^\delta + \left(-\log\left(1-(1-v)^\theta\right)\right)^\delta\right)}{\theta^2} \right\} \\ &\quad \times \left\{ \exp'\left(-\left(\left(-\log\left(1-(1-u)^\theta\right)\right)^\delta + \left(-\log\left(1-(1-v)^\theta\right)\right)^\delta\right)^{\frac{1}{\theta}}\right) \right\}. \end{aligned} \tag{4}$$
Next, observe that for $\delta \ge 1$, and for $(u,v) \in [0,1]^2$,
   - it is easy to see that $\left(-\log\left(1-(1-u)^\theta\right)\right)^\delta > 0$, and $\left(-\log\left(1-(1-v)^\theta\right)\right)^\delta > 0$.
   - Also, $\log(1-u) < 0$, and $\log(1-v) < 0$.

   Therefore, the numerator in the first two terms are negative, the second term is also negative. The third term involving $\exp'()$ is positive. Consequently, $\frac{\partial C(u,v)}{\partial \theta} < 0$, which completes the proof.
2. **Proposition 2.** The regression of U given $V = v$ is strictly decreasing in v.
Proof. The proof is left as an exercise to the reader.
3. **Proposition 3.** The BB6 copula as given in Eq. (1) is absolutely continuous. To establish the absolute continuity of the BB6 copula, we need to show that
$$\int_0^u \int_0^v \frac{\partial^2 C(s,t)}{\partial s \partial t} = C(u,v).$$
Proof. Simple and thus excluded.
It must be noted that these properties can also be derived for all the bivariate copulas described/utilized in this paper. However, for the sake of brevity, we have not considered them all.

4.2 Bivariate t Copula

The 2 dimensional unique t copula (see, Embrechts et al. (2001), McNeil, and Straumann (2001) or Fang & Fang (2002)) associated with a bivariate random vector $Y = (Y_1, Y_2)^T$, is given by
$$C_\delta^t(u,v) = \int_{-\infty}^{t_\delta^{-1}(u)} \int_{-\infty}^{t_\delta^{-1}(v)} \frac{\Gamma((\delta+2)/2)}{\Gamma(\delta/2)\sqrt{\{(\pi\delta)^2 |\Sigma|\}}} \left[1 + \frac{y^T \Sigma^{-1} y}{\delta}\right]^{-\frac{\delta+2}{2}} dy_1 dy_2,$$

Table 7: Dependence Structures of the BB6 Copula based on Swedish Motor insurance data.

Generator Function	$\phi(t) = (-\log[1-(1-t)^{\theta}])^{\delta}$
Blomqvist β (General)	$4C(0.5, 0.5) - 1$
Blomqvist β (Swedish Auto)	0.7397
Upper Tail Dependence(General)	$2 - 2^{\frac{1}{\delta\theta}}$
Upper Tail (Swedish Auto)	0.8321
Lower Tail Dependence	0
Kendall's τ	0.73

where $t_{\delta}^{-1}(.)$ denotes the quantile function of a standard univariate $t_{\delta}(.)$ distribution. Furthermore, Σ is the correlation matrix given by

$$\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix},$$

where ρ is the correlation coefficient between Y_1, and Y_2. The determinant of this matrix, denoted by $|\Sigma|$ is given by $|\Sigma| = 1 - \rho^2$. Next, one may verify the following regarding the dependence structure for a bivariate t copula

- This copula is also symmetric.
- Kendall's τ will be

  $$\tau = \frac{2}{\pi} \arcsin \rho,$$

  for the proof, see Fang & Fang (2002).
- Regarding the Spearman's correlation coefficient, there is no analytically tractable expression that are available. However, if ρ is the Spearman's correlation coefficient, and by denoting $\rho_t(\rho, v)$, and $\rho_N(\rho)$, as the Spearman's correlation coefficient for the bivariate t-copula and the bivariate Normal copula, one may show that:
  (i) when $\rho < 0$, $rho_t(\rho, v) > \rho_N(\rho)$; (ii) when $\rho > 0$, $rho_t(\rho, v) < \rho_N(\rho)$.
- The tail dependence coefficient (associated with a bivariate t copula as given earlier) λ is given by

  $$\lambda = 2t_{\delta+1}\left(-\frac{\sqrt{\{\delta+1\}}\sqrt{\{1-\rho\}}}{\sqrt{\{1+\rho\}}}\right),$$

  where $t_{\delta+1}$ is the univariate central student t distribution with $(\delta+1)$ degrees of freedom and ρ is the correlation coefficient. It is important to note that a student t-copula may exhibit both the positive tail-dependence although the "overall" association is negative $\rho < 0$. Furthermore, a student t-Copula with a large value of δ will tend to have a 0 tail-dependence even though the correlation is $.0$ The t-copula can capture the asymptotic dependence even when the variables are negatively (inversely) associated (see, Embrechts et al. 2001). In t-copula formula, as δ increases, the tail dependence weakens, and thus, the probability of occurrence of extreme values reduces

For illustrative purposes, we provide the following picture figure below (generated through the https://*copulatheque.shinyapps.io/copulas/* created by BenGraeler) shows a student t-copula with $\rho =$

0.995 and $\delta = 2$ which gives the value of Kendall's $\tau = 0.87$ and upper and lower tail-dependence of $\lambda = 0.87$.

The estimation of a student t-copula is quite difficult. Noticeably, the marginal tails (for bivariate and/or multivariate data distributions) of financial data are usually heavy tailed and hence this should be fitted by a t-distribution and not by a Gaussian distribution. In addition, the dependence in joint extremes of bivariate and/or multivariate financial data suggests a dependence structure allowing for tail dependence. Consequently, the use of t- copulas have become popular for modeling dependencies in financial data. Some recent applications have been: analysis of nonlinear and asymmetric dependence in the German equity market (Sun et al., 2008); estimation of large portfolio loss probabilities (Chan and Kroese, 2010); risk modeling for future cash flow (Pettere and Kollo, 2011). See also Dakovic and Czado (2011). Figure 7 represents the PDF and CDF of the bivariate t-copula with the estimated model parameters. For the CDF, there is a smooth increase from (0,0) to (1,1). The CDF has a little more curvature at the corners, which usually indicates tail dependence. The PDF confirms the tail dependence, as there is a spike both at (0,0) and (1,1) indicating both lower and upper tail dependence. It should be noted that due to the limitations of our R-Package, we could not directly graph our given parameters, our DF was equal to 2 in the t-copula we derived, and the package we are using must have DF greater than 2. This is as close as we could get it.

4.3 Tawn Type-1 copula

Here are some useful details on the Tawn Type-1 copula.

- The Tawn copula is a nonexchangeable extension of the Gumbel copula with three parameters (also known as the asymmetric logistic copula).
- Tawn copula's definition is based around so-called Pickands dependence functions, see Franc et al. (2011) for pertinent details. Eq. (4) in Franc et al. (2011) presents the way one can compute the density in the probability space using a Pickands function M:

$$C\left(u,v\right) = \left(u,v\right)^{\eta(w)},$$

  with $w = \frac{u}{uv}$.
- The Pickands dependence function derives Franc et al. (2011) for pertinent details. Equation (4) in Franc et al. (2011) presents the way one can compute the density in the probability space using a Pickands function M:

$$C\left(u,v\right) = \left(u,v\right)^{\eta(w)},$$

  with $w = \frac{u}{uv}$.
- The Tawn copula's Pickand function is

$$M(t) = \left(1-\psi_2\right)\left(1-t\right) + \left(1-\psi_1\right)t + \left[\left(\psi_1\left(1-t\right)\right)^{\theta} + \psi_2^{\theta}\right]^{1/\theta}.$$

The CDF in Figure 8 (see, Appendix), appears to show one side rising steeper than the other, which hints at asymmetry, a mainstay feature of Tawn type-1 copulas. Specifically, Tawn type 1 copulas model upper tail asymmetry, which can be noticed in the CDF graph's upper right hand region spike. The PDF shows a spike in the upper right hand corner, and a much softer presence elsewhere, again indicating upper tail dependence. These graphs also do a good job depicting what the Tawn type 1 models, which is the co-movements of high extreme values only.

4.4 Frank copula

The Frank copula is the best-fitted copula for the US Term Life variables income and amount borrowed on the life insurance policy. The cdf in Figure 8 shows nice curvature but no tail bias, as it appears to be symmetric across the diagonal. This indicates a moderate to strong overall dependence between the studied variables, but no tail dependence. The PDF in Figure 8 shows a crater-like structure, where it peaks in the center, meaning it assigns higher density to the mid-ranges of the variables' dependence, but is lighter in the co-extremes in the lower left or upper-right corners.

4.5 Survival Gumbel Copula

First, it would be nice to have the basic preliminaries related to a bivariate survival copula. We begin by stating the Sklar's theorem for survival functions:

Let $S(t_1, t_2)$ be a bivariate distribution with marginal survivor functions defined by $S_1(t_1)$ and $S_2(t_2)$. Then, there exists a bivariate copula $\bar{C}$ such that for all $(t_1, t_2) \in \mathbb{R}^2$

$$S(t_1, t_2) = C(S_1(t_1), S_2(t_2)).$$

The survival copula C couples the joint survival function to its' univariate marginals in a manner completely analogous to the way a copula connects the joint distribution function to its marginals. There exists a link between the survival copula C and the copula C. In the bivariate case, it is

$$\bar{C}(u_1, u_2) = u_1 + u_2 - 1 + C(1 - u_1, 1 - u_2).$$

It should be pointed out that the survival copula is also a copula, i.e. $C(u, v)$ is also a proper distribution function on $[0, 1] \times [0, 1]$. It is well known that many dependence properties of a bivariate distribution are copula properties, and therefore, can be obtained by studying the corresponding copula. These properties, however, do not depend on the marginals.

Table 8: Dependence Structures of the Frank Copula.

Generator Function	$\phi(t) = -\log_\alpha(\frac{\alpha^{-t}-1}{\alpha-1})$
Blomqvist β (General)	$\beta = 4\log_\alpha[1 + \frac{(\alpha^{0.5}-1)^2}{\alpha-1}] - 1$
Blomqvist β (US Term Life)	-0.264
Upper Tail Dependence (General)	0
Lower Tail Dependence (General)	0
Kendall's τ (US Term Life)	0.17

Next, we discuss some useful structural properties of the bivariate Frank copula.

- The Frank copula is an important Archimedean copula. Archimedean copulas are commutative, meaning that the order of the variables does not change the nature of the copula. The Frank copula is asymptotically independent in both directions. This means that it will have upper and lower tail dependence 0 in all cases.
- Its generator function is

$$-\log_\alpha(\frac{\alpha^{-t} - 1}{\alpha - 1}), \text{ for } \alpha > 0, \alpha \neq 1.$$

- The copula is defined as

$$C(u,v)=\log_{\alpha}[1+\frac{(\alpha^u-1)(\alpha^v-1)}{\alpha-1}],\ \text{for}\ \alpha>0.$$

- Its Kendall's τ is

$$\tau=4[\frac{3\log(\alpha)(-2\log(1-\alpha)+\log(\alpha)+2)-6Li_2(\alpha)+\pi^2}{6\log^2(\alpha)}]-1,$$

- The conditional copula of U given $V=v$ will be

$$C_1\left(u|V=v\right)=\frac{\left(\alpha^u-1\right)\alpha^v}{(\alpha-1)\left(\frac{(\alpha^u-1)(\alpha^v-1)}{\alpha-1}+1\right)} \tag{5}$$

Likewise, the conditional copula of V given $U=u$ will be

$$C_2\left(v|U=u\right)=\frac{\alpha^u\left(\alpha^v-1\right)}{(\alpha-1)\left(\frac{(\alpha^u-1)(\alpha^v-1)}{\alpha-1}+1\right)}. \tag{6}$$

- The conditional mean and variance of V given $U=u$ will be

$$\begin{aligned}&E\left[V|U=u\right]\\&=\left(2\log^3(\alpha)\right)^{-1}\\&\times\Big[-2\mathrm{Li}_3\left(\frac{\alpha^v-1}{\alpha^v-\alpha}\right)+2\mathrm{Li}_3\left(\frac{\alpha\left(\alpha^v-1\right)}{\alpha^v-\alpha}\right)-2\log(\alpha)\mathrm{Li}_2\left(\frac{\alpha\left(\alpha^v-1\right)}{\alpha^v-\alpha}\right)+\log^2(\alpha)\\&\times\left(\log\left(\alpha^v\right)-\log\left(-\frac{(\alpha-1)\alpha^v}{\alpha^v-\alpha}\right)\right)\Big].\end{aligned} \tag{7}$$

$$\begin{aligned}&Var\left[V|U=u\right]\\&=\left(3\log^4(\alpha)\right)^{-1}\Big[6\mathrm{Li}_4\left(\frac{\alpha^v-1}{\alpha^v-\alpha}\right)-6\mathrm{Li}_4\left(\frac{\alpha\left(\alpha^v-1\right)}{\alpha^v-\alpha}\right)-3\log^2(\alpha)\mathrm{Li}_2\left(\frac{\alpha\left(\alpha^v-1\right)}{\alpha^v-\alpha}\right)\\&+6\log(\alpha)\mathrm{Li}_3\left(\frac{\alpha\left(\alpha^v-1\right)}{\alpha^v-\alpha}\right)+\log^3(\alpha)\left(\log\left(\alpha^v\right)-\log\left(-\frac{(\alpha-1)\alpha^v}{\alpha^v-\alpha}\right)\right)\Big]\\&-\left(4\log^6(\alpha)\right)^{-1}\Big[\left(-2\mathrm{Li}_3\left(\frac{\alpha^v-1}{\alpha^v-\alpha}\right)+2\mathrm{Li}_3\left(\frac{\alpha\left(\alpha^v-1\right)}{\alpha^v-\alpha}\right)-2\log(\alpha)\mathrm{Li}_2\left(\frac{\alpha\left(\alpha^v-1\right)}{\alpha^v-\alpha}\right)\right.\\&\left.+\log^2(\alpha)\left(\log\left(\alpha^v\right)-\log\left(-\frac{(\alpha-1)\alpha^v}{\alpha^v-\alpha}\right)\right)\right)^2\Big].\end{aligned} \tag{8}$$

Next, we define the following two propositions for the bivariate Frank copula.

Proposition 1. The bivariate Frank copula is subharmonic for $0<\alpha<1$.

Proof. Let us consider the following:

$$\begin{aligned}\nabla^2 C_\alpha(u,v) &= \frac{\partial^2 C_\alpha(u,v)}{\partial u^2}+\frac{\partial^2 C_\alpha(u,v)}{\partial v^2}\\ &= -\frac{\log(\alpha)\,(\alpha^u-1)\,(\alpha^u-\alpha)\,\alpha^v}{(\alpha-\alpha^u+\alpha^{u+v}-\alpha^v)^2}-\frac{\log(\alpha)\alpha^u\,(\alpha^v-1)\,(\alpha^v-\alpha)}{(\alpha-\alpha^u+\alpha^{u+v}-\alpha^v)^2}\\ &= -\frac{\log(\alpha)\left(\alpha^{u+1}-2\alpha^{u+v}-2\alpha^{u+v+1}+\alpha^{2u+v}+\alpha^{u+2v}+\alpha^{v+1}\right)}{(\alpha-\alpha^u+\alpha^{u+v}-\alpha^v)^2}. \end{aligned} \tag{9}$$

From Eq. (9), it appears that for all choices of $(u,v)\in[0,1]\times[0,1]$ and for $0<\alpha<1$, $\nabla^2 C_\alpha(u,v)\geq 0$. This completes the proof.

Proposition 2. The bivariate Frank copula is absolutely continuous.

Proof. Simple and thus excluded.

Proposition 3. The bivariate Frank copula is symmetric.

Proof. It is easy to observe that for the Frank copula, $C(u,v)=C(v,u)$ which implies the result.

Figure 9 represents the pdf and the cdf of a bivariate Frank copula with the estimated parameter value for the Income and BorrowCVLife poi. Figure 9 exhibits nice curvature but no tail bias, as it appears to be symmetric across the diagonal. This indicates a moderate overall dependence between the studied variables, but no tail dependence. The PDF in Figure 9 shows a crater-like structure, where it peaks in the center, meaning it assigns higher density to the mid-ranges of the variables' dependence, but is lighter in the co-extremes in the lower left or upper-right corners. Figure 10, however, highlights the importance of running multiple tests. As we can see, the ranked numerical data makes the scatterplot just be horizontal lines, so more analysis is needed other than the scatter plot, as we can draw no real conclusions from it. Figure 11's contour plot, shows roughly elliptical contours that are centered around the origin, which is a hallmark of weak/moderate dependence. The inner contours represent the highest area of density, which aligns with our scatter plot, Figure 10. The contours are not skewed, which confirms the symmetry in our data. Since it is also rounded and there are no corners formed, this confirms joint co-extremes are not likely, and more moderate values are going to move jointly.

4.6 Survival Gumbel Copula

The best-fitted copula of the Medicare data was the Survival Gumbel copula. The Survival Gumbel copula is the Gumbel copula under the survival transformation. Whereas the Gumbel copula is useful for strong upper tail dependence, this transformation rotates the copula so that lower tail dependence is strong instead. Looking below at the scatter, contour, PDF and CDF plots, which can be found in Figures 12, 13 and 14, respectively. Figure 12's scatterplot has a nearly perfect diagonal pattern, suggesting a very strong and monotonic dependence relationship. The relationship is also very symmetric, and shows very strong co-movement: as CONV_CHG increases, so does TOT_D. In Figure 13, the contours are thin and very narrow along the 45 degree line, which is textbook near perfect dependence. The density is very tightly concentrated along the diagonal, indicating very high correlation. There is no sign of tail dependence and the contours are symmetrically distributed along the diagonal, ruling out asymmetry. In Figure 14, the surface of the CDF of Figure 10 shows a sharp increase towards the lower left corner (0,0). That sharp slope indicates concentration in mass in the lower tail, meaning that both variables tend to be small together. Note that the regular Gumbel CDF tends to show more mass concentration in the upper right, meaning that both variables tend to be larger together. The PDF also reinforces this assertion, as there is a steep spike in the bottom left corner, showing that the co-extremes are more closely associated in the lower tail.

[htp]

Table 9: Dependence Structures of the Survival Gumbel Copula for the Medicare Cost dataset.

Generator Function	$\phi(t) = (-\log t)^\theta$
Blomqvist β (General)	$\$\beta = 4C(0.5, 0.5) - 1$
Blomqvist β (Medicare)	0.79
Upper Tail Dependence (General)	0
Lower Tail Dependence (General)	$2 - 2^{1/\theta}$
Lower Tail Dependence (Medicare)	0.88
Kendall's τ (Medicare)	0.84

- The usual Gumbel copula has the generator function

$$\psi(t) = (-\log t)^\theta, \theta \geq 1.$$

- The Gumbel copula is then

$$C_{\text{Gumbel}}(u, v) = \exp[-((-\log(u))^\theta + (-\log(v))^\theta)^{1/\theta}].$$

- The Survival Gumbel copula is defined as

$$C(u, v) = u + v - 1 + \exp[-((-\log(1-u))^\theta + (-\log(1-v))^\theta)^{1/\theta}]$$

- Its Kendall's τ is then

$$\tau = 1 + 4\int_0^1 \frac{\phi(t)}{\phi'(t)}dt = 1 - \frac{1}{\theta}.$$

- The lower tail dependence of the Survival Gumbel copula is equivalent to the upper tail dependence of the Gumbel copula:

$$\lambda_L = \lambda_U^{\text{Gumbel}} = \lim_{u \uparrow 1} \frac{1 - 2u + C(u, u)}{1 - u}$$

$$= \lim_{u \uparrow 1} \frac{1 - 2u + \exp(-2(-\log u)^\theta)}{1 - u} = 2 - 2^{1/\theta},$$

on applying L'Hôpital's rule.

5 Conclusion

The best fitted bivariate copulas in all the cases that we have studied, the Kendall's τ ranges from 0.17 to 0.87; and the Spearman's ρ ranges from from 0.21 to 0.97. The log-likelihoods, or model fitness, range from 11.21 to 3566.65. One of the possible reason(s) for the varying dependence and fitness is the use of ranked data. Copulas are valid for these types of data but tell a less interesting story when few ranks are available. In the code, we used the pseudo-observation function to approximate the probability integral transform discussed in the introduction. This is most powerful when there are many ranks to analyze: when ranks are dense enough to be approximately continuous. For data with few ranks, copulas have less fit and seem to have less dependence.

In the Swedish Motor Insurance data, we see strong fit and dependence, particularly for *Claims* to *Payments*. *Insured* to *Payments* has high dependence and log-likelihood. This suggests copula modeling has value for this data. Inspecting the data, we have numerical values for *Claims*, *Payments*, and *Insured*. This gives significant evidence for the algorithm to select the best-fitted copula. The resulting copulas then showed high dependence.

The Medicare Cost provided a strong dependence from *Charges* to *Number of Stays*. This is an example of strong dependence that has little dependence at one tail. Specifically, the Survival Gumbel copula was the best-fitted. This copula has no dependence in the upper tail. This indicates that high charges to Medicare and high number of hospital stays do not have a particular dependence besides random chance.

In the US Term Life data, we see weaker dependence and fit, particularly for *Income* to *Borrow CV Life*. Upon inspection, *Income* is numerical, although clearly rounded compared to what you would expect for such data. *Face* is either rounded or companies require round values for the face value of the life insurance. *Borrow CV Life* is ranked from 0-5 by the amount borrowed from the policy. These create issues of fitness. This illustrates the importance of detailed data for copula modeling.

If the US Term Life data had more granularity, we may recover a better fit copula than the Tawn Type-1 and Frank copulas. This asserts that insights from the copulas generated require more care or more data.

We provide useful results of copula modeling for three major insurance institutions. Swedish Motor Insurance data exhibits strong dependence in the variables chosen. Medicare data does as well, with independence of large charges and numbers of stays. US Term Life suffers from constraints of granularity for copula modeling and could be studied further with more granular data. For the Swedish Motor and Medicare data, these considerations could be useful for pricing and risk mitigation.

References

[1] Tawn, J. A. (1990). Modelling multivariate extreme value distributions. Biometrika, 77(2), 245-253.

[2] A. Sklar (1959). Fonctions de répartition à N dimensions et leurs marges. Publications de l'Institut Statistique de l'Université de Paris, 8, 229–231.

[3] Blomqvist, N. (1950). On a measure of dependence between two random variables. The Annals of Mathematical Statistics, 593-600.

[4] Durante, F., & Sempi, C. (2016). Principles of copula theory (Vol. 474). CRC Press.

[5] Joe, H. (1997). Multivariate models and multivariate dependence concepts. CRC press.

[6] Joe, H. (2014). Dependence modeling with copulas. CRC press.

[7] Denuit, M., Dhaene, J., Goovaerts, M., & Kaas, R. (2005). Actuarial Theory for Dependent Risks: Measures, Orders and Models. John Wiley & Sons.

[8] Nadarajah, S., Afuecheta, E., & Chan, S. (2017). A compendium of copulas. Statistica, 77(4), 279-328.

[9] Nooraee, N., et al. (2014). Measuring agreement between raters in medical imaging: Blomqvist's β as a robust alternative. Journal of Biomedical Science and Engineering, 7(8), 601–609.

[10] Agresti, A. (2010). Analysis of Ordinal Categorical Data (2nd ed.). Wiley.

[11] Stigler, S. M. (1986). The History of Statistics: The Measurement of Uncertainty before 1900. Harvard University Press.

[12] Trivedi, P. K., & Zimmer, D. M. (2007). Copula modeling: an introduction for practitioners. Foundations and Trends® in Econometrics, 1(1), 1-111.

[13] Geenens, G. (2024). (Re-) Reading Sklar (1959)—A Personal View on Sklar's Theorem. Mathematics, 12(3), 380.

[14] Dorey, M., & Joubert, P. (2005). Modelling copulas: an overview. The Staple Inn Actuarial Society, 1-27.

[15] MacKenzie, D., & Spears, T. (2014). 'The formula that killed Wall Street': The Gaussian copula and modelling practices in investment banking. Social Studies of Science, 44(3), 393-417.

[16] Patton, A. J. (2009). Copula–based models for financial time series. In Handbook of financial time series (pp. 767-785). Berlin, Heidelberg: Springer Berlin Heidelberg.

[17] Kole, E., Koedijk, K., & Verbeek, M. (2007). Selecting copulas for risk management. Journal of Banking & Finance, 31(8), 2405-2423.

[18] Genest, C., & MacKay, J. (1986). The joy of copulas: Bivariate distributions with uniform marginals. The American Statistician, 40(4), 280-283.

[19] Haugh, M. (2016). An Introduction to Copulas. IEOR E4602: Quantitative Risk Management. Lecture Notes. New York: Columbia University

[20] Tinungki, G. M., Siswanto, S., & Najiha, A. (2023). The Gumbel Copula Method for Estimating Value at Risk: Evidence from Telecommunication Stocks in Indonesia during the COVID-19 Pandemic. Journal of Risk and Financial Management, 16(10), 424.

[21] Venter, G. G. (2002, March). Tails of copulas. In Proceedings of the Casualty Actuarial Society (Vol. 89, No. 171, pp. 68-113).

[22] Nelsen, R. B. (2003, March). Properties and applications of copulas: A brief survey. In Proceedings of the first brazilian conference on statistical modeling in insurance and finance (pp. 10-28). Sao Paulo: University Press USP.

[23] Nelsen, R. B. (2005). Dependence modeling with archimedean copulas.

[24] Deng, L., Smith, M. S., & Maneesoonthorn, W. (2024). Large skew-t copula models and asymmetric dependence in intraday equity returns. Journal of Business & Economic Statistics, 1-17.

[25] Hintz, E., Hofert, M., & Lemieux, C. (2022). Computational Challenges of t and Related Copulas. Journal of Data Science, 20(1).

[26] Frees, E. W. (2009). Regression modeling with actuarial and financial applications. Cambridge University Press.

[27] Tawn, J. A. (1988). Bivariate extreme value theory: models and estimation. Biometrika, 75(3), 397-415.

Appendix

Appendix A

.On the R Package: Vine Copula

Here, we provide a generic R-code based on the Vine Copula package which is used in the main body of the text for selecting the best possible bivariate copula for the four different insurance datasets:

```
install.packages("copula")
library("copula")
m<-pobs(a)
n<-pobs(b)
 install.packages("VineCopula")
library("VineCopula")
selectedCopula<-BiCopSelect (m,n,familyset=NA)
summary(selectedCopula)
```

- **Remark 1** In the above code, a and b are the transformed (on a log (to the base e) scale) variable values corresponding to two components of the associated bivariate data.
- **Remark 2** The best-fitted bivariate copulas mentioned here do not possess a closed form of expression in terms of their density function (i.e., the p.d.f.). However, in order to obtain the p.d.f. of each of these copulas, one may use R. Next, we provide an example as to how one can simulate from the p.d.f. of a Survival BB1 copula with specific parameter choices in R.

Simulate from a bivariate BB6 copula:

```
 install.packages("VineCopula")
library("VineCopula")
BB6<-BiCop( family = 8 , par =0.25 , par2 = 0.75)
sim<-BiCopSim( 10000, BB6).
```

Appendix B

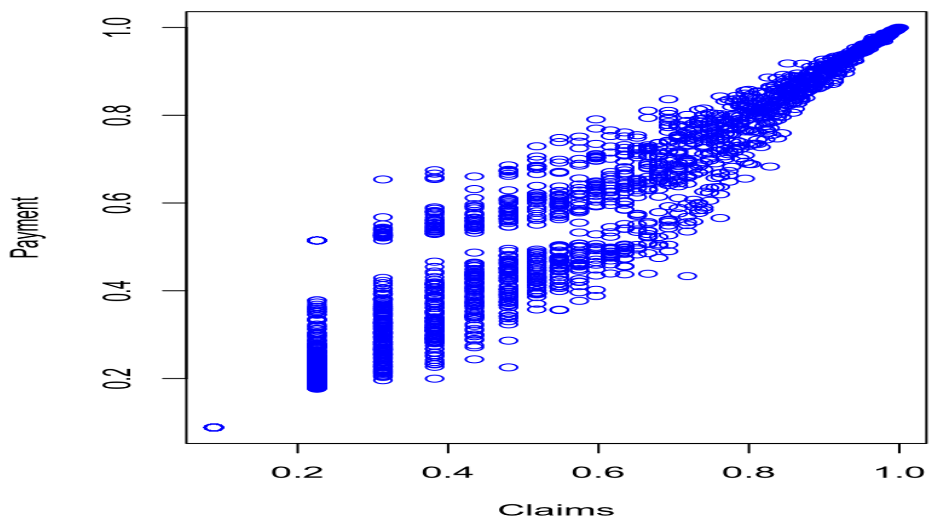

Figure 1: Scatter plot of Claims & Payment.

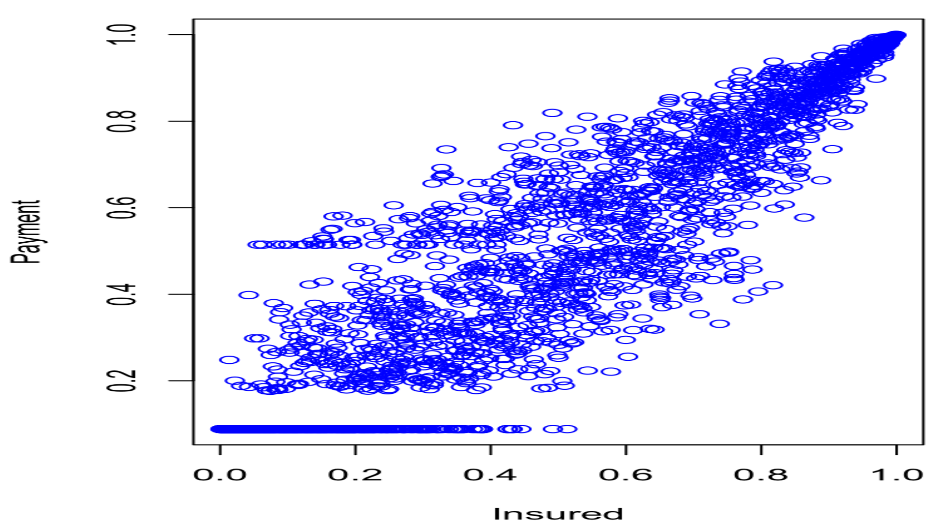

Figure 2: Scatter plot of Insured and Payment.

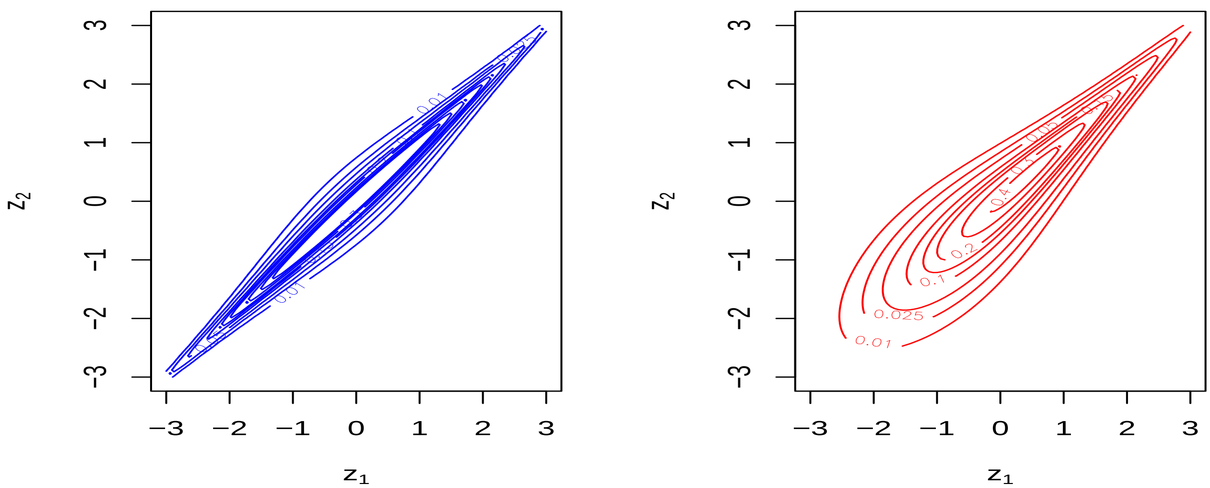

Figure 3: Contour plots of Claims & Payment (right panel) and Insured and Payment (left panel).

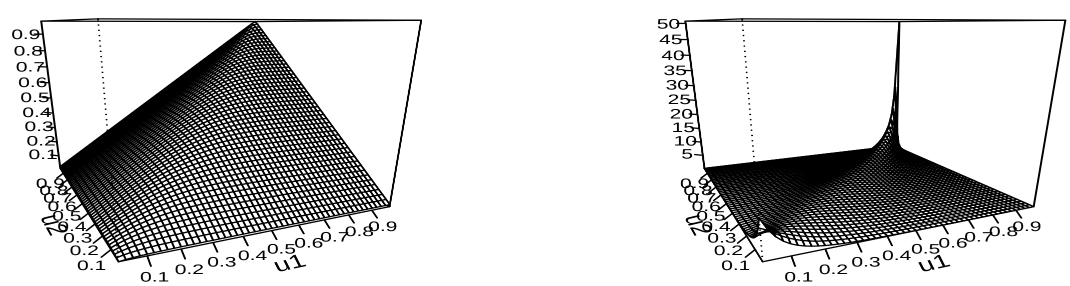

Figure 4: BB6 copula cdf and pdf with parameter values (1.59, 2.81) for Insured and Payment.

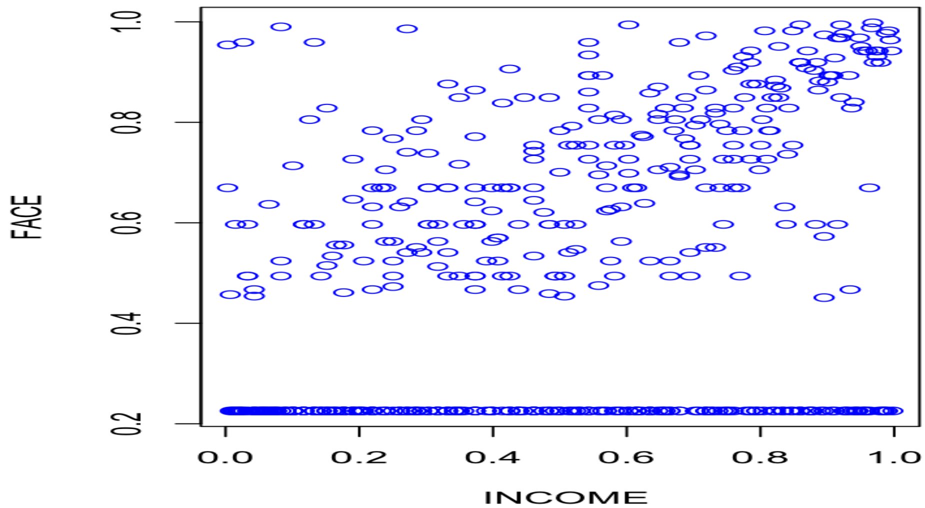

Figure 5: Scatter plot of Income & Face.

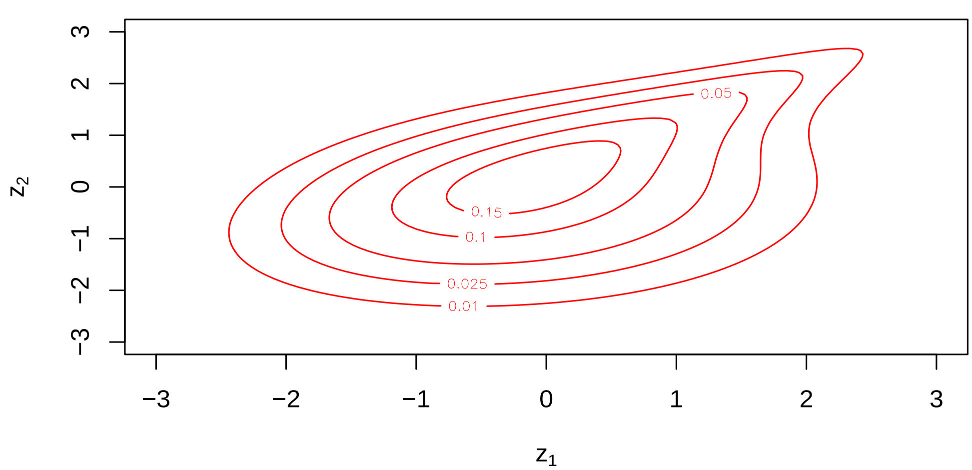

Figure 6: Contour plot of Income & Face.

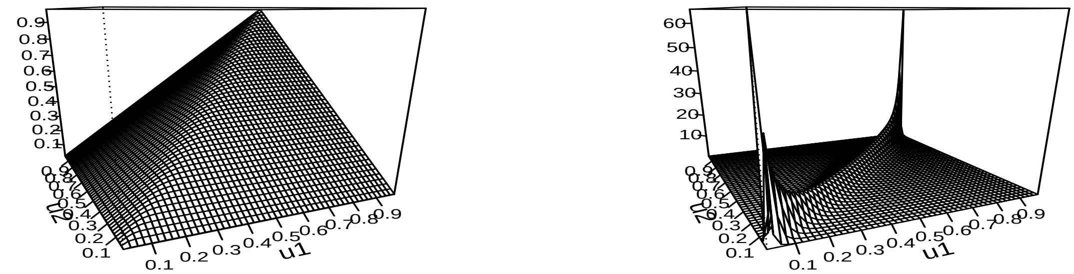

Figure 7: Student t-copula pdf and cdf with parameter values (0.98, 2) for Claims and Payment.

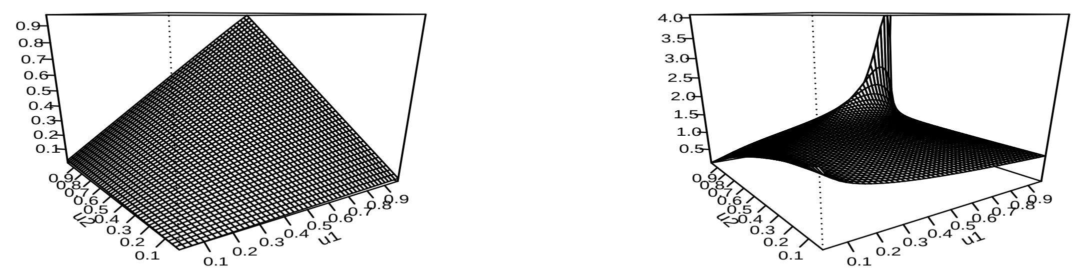

Figure 8: Tawn Type 1 Copula PDF and CDF with parameter values (1.84) and (0.49) for Income and Face.

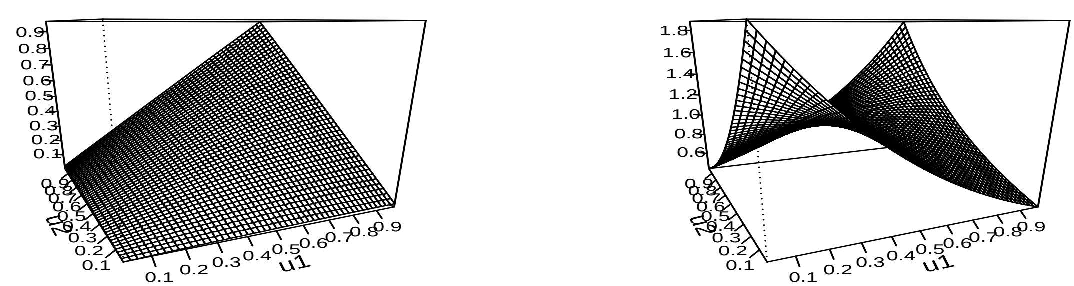

Figure 9: Frank Copula PDF and CDF with parameter value (1.59) for Income and BorrowCVLifePol.

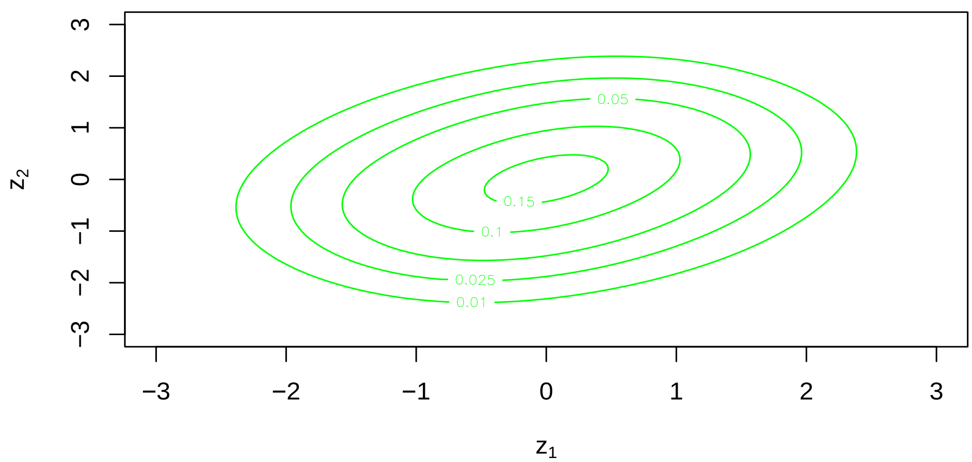

Figure 10: Frank Copula Scatter plot with parameter value (1.59) for Income and BorrowCVLifePol.

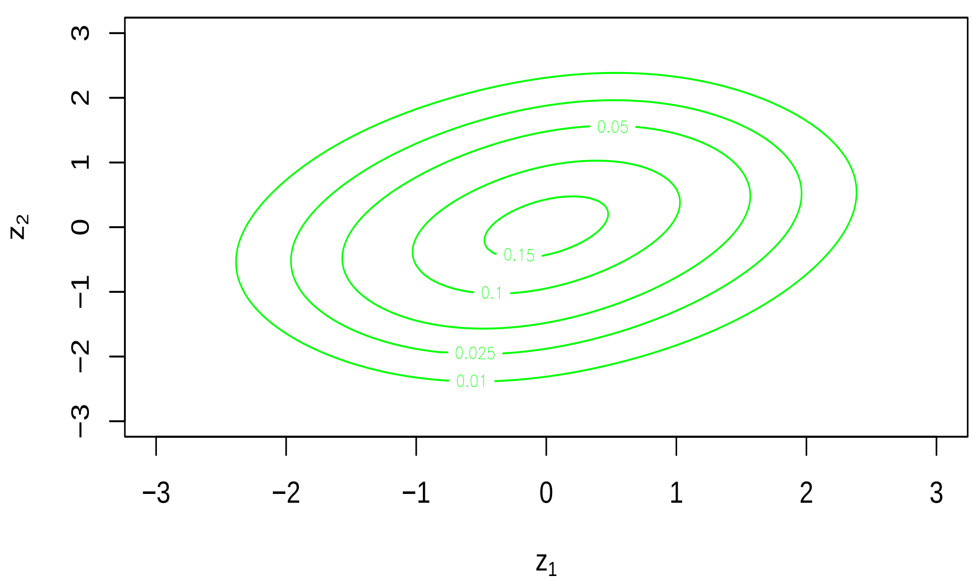

Figure 11: Frank Copula Contour plot with parameter value (1.59) for Income and BorrowCVLifePol.

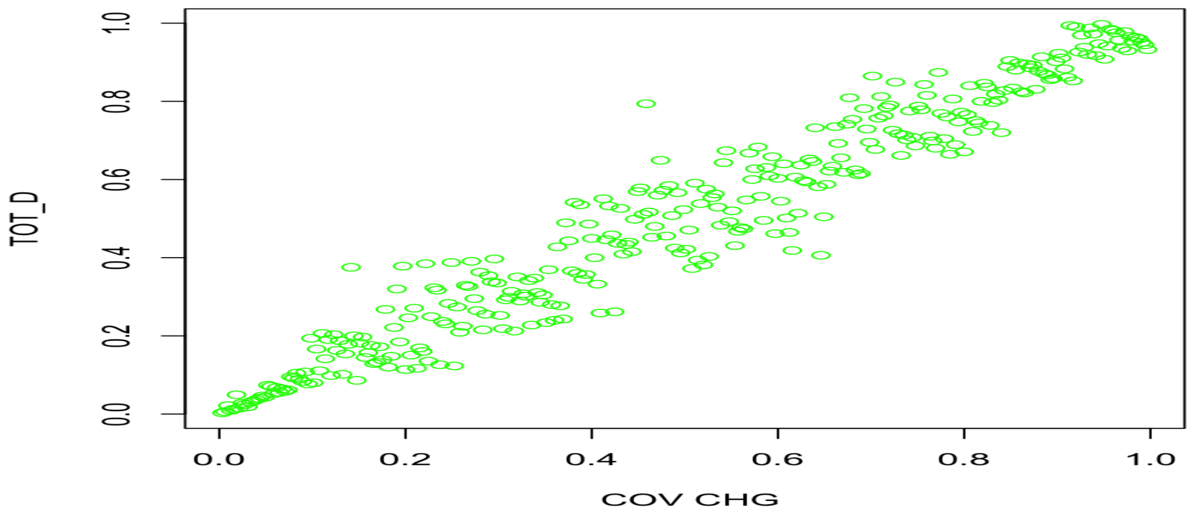

Figure 12: Survival Gumbel Copula Scatter plot for COV_CHG and TOT_D.

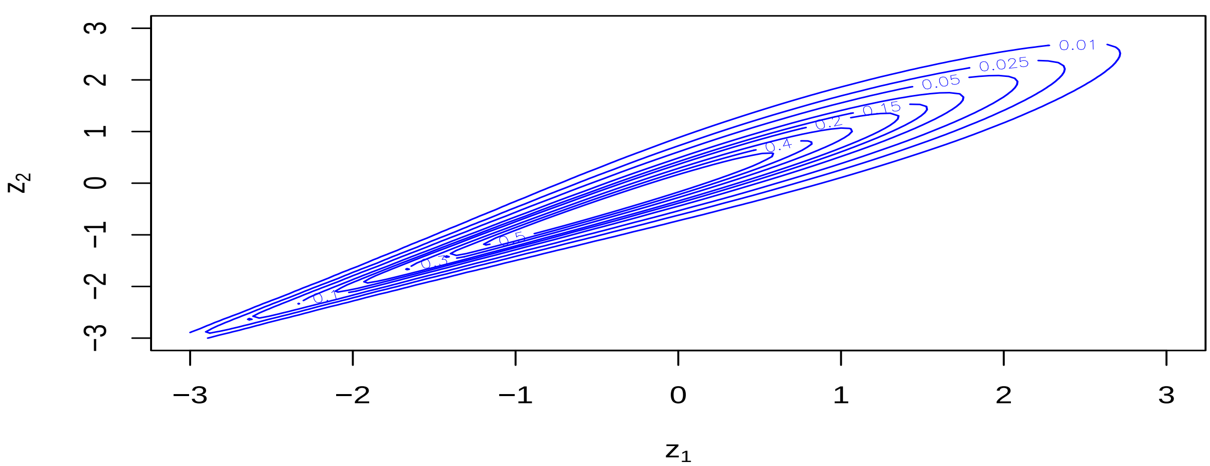

Figure 13: Survival Gumbel Copula Contour plot for COV_CHG and TOT_D.

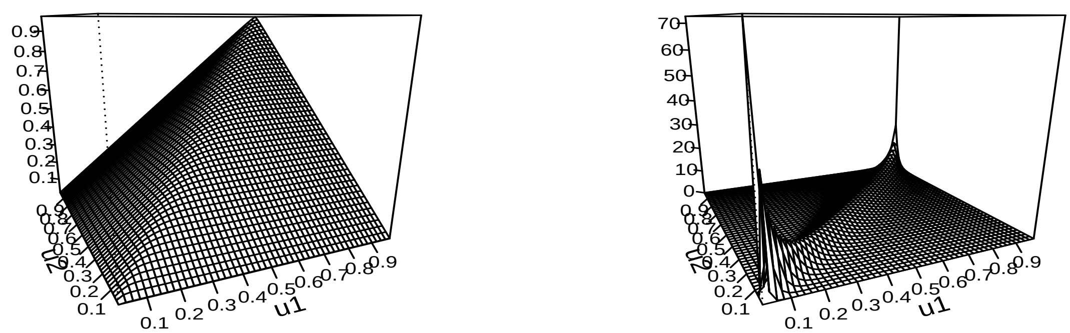

Figure 14: Survival Gumbel Copula PDF and CDF with parameter value (6.12) for COV_CHG and TOT_D.