Application of Copula Theory in Modelling Risks by Incorporating Dependence Structure

(A Case Study of the Kenyan General Insurance Business)

Joseph Kyalo Mung'atu

A thesis submitted in fulfilment for the Degree of Doctor of Philosophy in Applied Statistics in the Jomo Kenyatta University of Agriculture and Technology

2011

DECLARATION

This thesis is my original work and has not been presented for a degree in any other University.

Signature: ··························· Date: ··················

Joseph Kyalo Mung'atu

This thesis has been submitted for examination with our approval as University Supervisors.

Signature: ··························· Date: ··················

Dr. Samuel Mwalili

JKUAT, Kenya

Signature: ······························ Date: ··················

Prof. Patrick G. O. Weke

UoN, Kenya

DEDICATION

To my two daughters Immaculate and Sharon

ACKNOWLEDGEMENT

First and foremost I thank the Almighty God for giving good health and providence during my research epoch. I also would like to express my sincere appreciation to my supervisors: Dr. S.M. Mwalili (JKUAT) and Prof. P. G. O. Weke of the University of Nairobi for their guidance, advice, patience, moral and professional support during the entire learning period. Besides their tight schedules, they still had time with me over the duration. I do appreciation to Jomo Kenyatta University of Agriculture and Technology (JKUAT) for entirely funding this undertaking. This financial support came at a much deserved time. Additionally, I thank all the members of JKUAT, Nairobi Campus where I served as I undertook my research. To Prof. C. Kanali and Prof. E. Magiri, I will forever remember your encouragement and the way you nurtured me to be a better scholar.

In a special way, I'm exceedingly grateful to all the members of the Statistics and Actuarial Science (STACS) department, JKUAT, for the guidance I received at all the presentations, towards my progress, organised through the department. The support of the Chairman, Prof. P. Mwita, cannot go unnoticed as he helped me look at Copulas from a more pragmatic stand-point. I am thankful to Prof. R. Odhiambo, a senior member of the department, who yearned to see the successful completion of my work. I also felt so encouraged by H. M. Humphreys' words as he kept a keen track of my progress besides attending all the seminars relating to this work. And to my Ph.D colleagues at the department: V. Andika, J. Chelule, J. Akinyi, C.

Mugo, J. Kiingati, and J. M. Wekesa; I cherish your support recalling the way we used to discuss the developments in our respective areas. Lastly, much gratitude to Gladys, Justa and Julia who helped me forward many of my documents, through the department, including the progress reports to the Board of Post Graduate Studies (BPS). To this end it is GOD!

Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

ABBREVIATIONS

CTE	Conditional Tail Expectation
LOB	Lines of Business
VaR	Value at Risk
DomH	Domain of H
RanH	Rank of H
CDF	Cumulative density function
PDF	Probability density fuction
df	Degrees of freedom
DFA	Dynamic financial analysis
ALM	Asset-Liability Management
IBNR	Incurred but not reported
HAC	Hierarchical Archimedean Copulas
KES	Kenya Shilling
IRA	Insurance Regulatory Authority

ABSTRACT

Insurance companies maintain different lines of business as a mode of diversification which in itself aids in reducing risks of encountering ruin. The dependence structure of these lines of business cannot be ignored especially in rate making as it reduces diversification benefits. However, in practice you would find that risks are more heterogeneous than homogeneous a problem that can be solved by breaking down the risks into a number of homogeneous categories. The lines of business are considered,here, to contain sub-classes which are homogeneous. The lines will depict a hierarchical structure from the sub-classes to the main lines of business up to the portfolio level and their dependence structure is studied here by the hierarchical copulas. In risk classification, similar risks should be assigned to the same class with respect to each variable. The dependencies are examined by fitting copulas, estimating the dependence parameters and lastly using distance matrices to cluster the risks together. The distance to use in the classification is determined by the problem at hand. The empirical study derives its data from the general insurance business in Kenya where the risks are classified by the Copula based approach. This work proposed the use of the upper tail dependence, measured by the tail index, derived from the dependence parameter in determining the retention limits for a re-insurance arrangement. Though the dependence is not the only factor to consider for such re-insurance treaties the forwarding proportions should be some where proportional to ${}^{1}\!/_{(1\,-\,\text{Tail index})}$. This will ensure that the highly dependent risks in the upper tail will forward higher proportions to the re-insurer and vice versa.

Chapter 1

INTRODUCTION

1.1 Introduction to Copulas

First we look at what a copula is and why it has been gaining popularity among the risk managers, in actuarial and statistical work in the recent past. The term *copula* was first used in the work of Sklar (1959) and is derived from the Latin word *copulare*, meaning to connect or to join. The main purpose of copulas is to describe the interrelation of several random variables. Copulas, therefore, are tools for modelling dependence of several random variables.

A copula is a function that joins or couples a multivariate distribution function onto univariate marginal distribution functions and so a copula is a multivariate distribution function. Our concern in this thesis is how random variables relate to each other. Since it is possible to determine the behaviour of a univariate marginal distribution function, by use of multivariate distribution we are able to determine how the marginal distributions behave together or to fully understand their dependence in a deeper way. Copulas are becoming more popular tools in actuarial problems in measuring dependence than the commonly known linear correlation due to the limitations of the linear correlation as outlined in sub-section1.2.1.

1.2 Measures of Dependence

Measures of dependence summarize a complicated dependence structure in a single number in the bivariate case. There are three important concepts in measuring dependence. These include: the linear correlation (the classical one), rank correlation and the coefficients of tail dependence. The last two measures are general enough to give sensible measures for any dependence structure since correlation is only a suitable measure in a special class of distributions, that is, the elliptical distributions. They provide, perhaps, the best alternatives to the linear correlation coefficient as a measure of dependence for non-elliptical distributions, for which the linear correlation coefficient is inappropriate and often misleading. Copulas capture the properties of the joint distribution which are invariant, that is, they remain unchanged under strictly increasing transformations of the random variables. Linear correlation will also be considered as a measure of dependence and also state when it becomes a handy measure of dependence.

1.2.1 Linear correlation

The Pearson product-moment correlation coefficient (r) or correlation coefficient in this sequel is a measure of the degree of linear relationship between two variables, usually labeled X and Y. While in regression the emphasis is on predicting one variable from the other, in correlation the emphasis is on the degree of a linear relationship between two variables (risks). In regression the interest is directional, one variable is predicted and the other is the predictor while in correlation the interest is non-directional, the relationship is the critical aspect.

Let X and Y be some two risk variables for which we want to model their relationship or dependence, in all likelihood, we will therefore use linear correlation. However, this will fail due to the fact that correlation is not an all-purpose dependency measure in risk management. We now examine correlation and show where it is suitable and also highlight its drawbacks in comparison with copulas. The terms *dependency* and *correlation* are used interchangeably but correlation is an imperfect measure of dependency. If X and Y are the risk variables, then the linear correlation coefficient for (X, Y) is

$$\rho(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{var}(X)}\sqrt{\mathrm{var}(Y)}} \text{ and } -1 \leq \rho(X,Y) \leq 1$$

where $var(X)$ and $var(Y)$ are the variances of X and Y respectively and $cov(X,Y)$ is their covariance. Correlation is a reasonable measure of dependency when the random variables are distributed as bivariate normal and thus a correlation of +1 tells us that the two variables are positively dependent (rises or falls together), -1 tells us that they are negatively dependent (one rises and the other falls) and a correlation of zero indicates that the two variables are independent.

As a dependency measure, correlation is reasonable in elliptical distributions which include Student-t and normal distributions. The elliptical distributions are widely used because their functions can be determined from knowledge of variances and correlations alone. We should note that in Student-t distribution, uncorrelated components do not imply independence and this provides an example where zero correlation of risks does not imply independence of risks. It is only in the case of the multivariate normal where uncorrelatedness always can be interpreted as independence. Further, correlation has its limitations in a normal or elliptical world in that it cannot model asymmetries. In financial and insurance applications there is stronger dependence between big losses than between big gains.

We hereby list some of the problems of correlation as a dependency measure. First and foremost, correlation is not defined unless the variances are finite. It is, therefore, not an appropriate dependence measure for very heavy-tailed risks where variances appear infinite. Correlation is not invariant under transformations of the risks. For example, correlation between X and Y is not the same as between $log(X)$ and $log(Y)$. Hence, transformations of the data can affect the correlation estimates.

Correlation is simply an omnibus index of dependency; it cannot tell us everything we would like to know about the dependence structure of risks. Perfect positively dependent risks do not necessary have a correlation of 1 and vice versa. The values between -1 and 1 are not always attainable. All possible values of correlation depend on the marginal distribution of the risks. We will now get the alternative to correlation as a measure of dependency by use of copula. This is due to the fact that modern risk management calls for an understanding of dependence going beyond simple linear correlation.

1.2.2 Rank correlation

It is a common practice in nonparametric statistics to concentrate on the ranks of given data rather than on the data itself. *Kendall's tau* and *Spearman's rho* are key correlation coefficients derived from ranks of a data set. Considering ranks leads to scale invariant estimates, which in turn is very pleasant when working with copulas. Therefore rank correlations give a possible way of fitting copulas to data.

Definition 1.2.1 *(Concordance) Let* (x_i, y_i) *and* (x_j, y_j) *be two observations from a vector* (X, Y) *of continuous random variables. We say that they are concordant if* $x_i < x_j$ *and* $y_i < y_j$*, or* $x_i > x_j$ *and* $y_i > y_j$ *that is,* $(x_i - x_j)(y_i - y_j) > 0$*. Similarly, they are* discordant *if* $x_i < x_j$ *and* $y_i > y_j$*, or* $x_i > x_j$ *and* $y_i < y_j$ *, that is,*$(x_i - x_j)(y_i - y_j) < 0$ *. That is, if* X *and* Y *are two random variables, then they are said to be concordant if large (small) values of* X *tend to be associated with large (small) values of* Y*.*

1.2.2.1 Kendall's tau

Kendall's tau is a measure of association and is defined in terms of concordance.

Definition 1.2.2 *(Kendall's tau) Let* $\{(x_1, y_1), ..., (x_n, y_n)\}$ *denote a random sample of* n *observations from a random vector* (X, Y) *of continuous random variables. There are* $\binom{n}{2}$ *distinct pairs* (x_i, y_i) *and* (x_j, y_j) *of observations in the sample, and each pair is either concordant or discordant. Let* c *denote the number of concordant pairs and* d *the number of discordant pairs. Then, the Kendall's tau for the sample is defined as,*

$$t = \frac{c-d}{c+d} = \frac{c-d}{\binom{n}{2}}$$

The population version of Kendall's tau for a vector (X, Y) of continuous random variables with joint distribution function H is defined as follows. Let (X_i, Y_i) and (X_j, Y_j) be independent and identically distributed random vectors, each with joint distribution function H. The population version of Kendall's tau is defined as the probability of concordance minus the probability of discordance.

$$\tau = \tau_{X,Y} = P\left[(X_i - X_j)(Y_i - Y_j) > 0\right] - P\left[(X_i - X_j)(Y_i - Y_j) < 0\right]$$

1.2.2.2 Spearman's rho

As in the case for Kendall's tau, the population version of the measure of association known as Spearman's rho is also based on discordance and concordance.

Definition 1.2.3 *(Spearman's rho) Let* (X_i, Y_i) , (X_j, Y_j) and (X_k, Y_k) *be independent random vectors with a common joint distribution function* H*, whose margins are* F*,* G*. Then, the population version is defined as the difference between probabilities of concordance and discordance of the vectors* (X_i, Y_i) *and* (X_j, Y_k) *thus,*

$$\rho = \rho_{X,Y} = 3\left\{P\left[(X_i - X_j)(Y_i - Y_k) > 0\right] - P\left[(X_i - X_j)(Y_i - Y_k) < 0\right]\right\}$$

Kendall's tau and Spearman's rho are both symmetric dependence measures taking values in the interval [-1,1]. They give the value zero for independent random variables although a correlation of zero does not necessarily imply independence.

1.2.3 Tail dependence

We distinguish between *upper* and *lower* tail dependence. Upper tail dependence means intuitively, that with large values of X_1 also large values of X_2 are expected. More precisely, the probability that X_1 exceeds a given threshold q, given that X_2 has exceeded the same level q for $q \to 1$ is considered. If this probability is smaller than order q, then the random variables have no tail dependence, for example in the independent case. Otherwise they have tail dependence.

Definition 1.2.4 *(Tail dependence) For random variables* X_1 *and* X_2 *with cumulative distribution functions* $F_i, i = 1, 2$ *we define the coefficient of upper tail dependence by:*

$$\lambda_u \doteq \lim_{q \nearrow 1} P\left(X_2 > F_2^{-1}(q) \,\middle|\, X_1 > F_1^{-1}(q)\right),$$

provided that the limit exists and $\lambda_u \in [0, 1]$ *. The coefficient of lower tail dependence is defined analogously by*

$$\lambda_l \doteq \lim_{q \searrow 0} P\left(X_2 \leq F_2^{-1}(q) \,\middle|\, X_1 \leq F_1^{-1}(q)\right).$$

If $\lambda_u > 0$, we say that X_1 and X_2 have upper tail dependence, while for $\lambda_u = 0$ we say that they are *asymptotically independent* in the upper tail and analogously for λ_l .

For continuous cumulative distribution functions, quite simple expressions are obtained for the coefficients using Bayes' rule, namely

$$\begin{aligned} \lambda_l &= \lim_{q \searrow 0} \frac{P\left(X_2 \leq F_2^{-1}(q), X_1 \leq F_1^{-1}(q)\right)}{P\left(X_1 \leq F_1^{-1}(q)\right)} \\ &= \lim_{q \searrow 0} \frac{C(q, q)}{q} \end{aligned}$$

and similarly,

$$\lambda_u = 2 + \lim_{q \searrow 0} \frac{C(1-q, 1-q) - 1}{q}$$

1.3 The Copula Function

The operational definition of a copula is a multivariate distribution function defined on the unit cube $[0,1]^n$, with uniformly distributed marginals.

Definition 1.3.1 *(Uniform Distribution) A random variable X has a uniform distribution and it is referred to as a continuous uniform random variable if and only if its probability density is given*

$$U(x:a,b) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

Unlike correlation, copulas have a nice property of being invariant under strictly increasing transformations of random variables. Moreover, instead of summarizing dependence structure with a single number like correlation, we can use a model for the dependence structure that reflects more detailed knowledge of the risk management problem we are handling. Copula as a multivariate distribution function helps us to fit multivariate risk factor data, and then find the marginal models for individual risk factors and copula models for their dependence structure. We have a wide range of copula families from which to select a suitable model, see section 2.3. This enables us to choose a particular copula family depending on the random variables of the multivariate data we are trying to model. If the marginal distributions are known, a copula can be used to suggest a suitable form for the joint distribution. This means we can create multivariate distribution functions by joining arbitrary marginal distributions together and we can extract copulas from well-known multivariate distribution functions. Multivariate distribution functions have more information than the individual marginal distributions and this generally helps us to avoid the drawbacks of correlation as a measure of dependency. Copula helps to model asymmetries. In financial applications there is a stronger dependence between big losses than between big gains. The choice of appropriate model is paramount. Copula represents a way of trying to extract the dependence structure from the joint distribution function and to separate dependence and marginal behaviour.

We first concentrate on general multivariate distributions and then study the special properties of the copula subset. For a function H, we denote by DomH and RanH the domain and range of H respectively. Additionally, a function f will be called increasing whenever $x \leq y$ implies that $f(x) \leq f(y)$. We may also refer to this as f is nondecreasing. A statement about points of a set $S \subset \Re^n$, where S is typically the real line or the unit cube $[0,1]^n$, is said to hold almost everywhere if the set of points of S, where the statement fails to hold, has Lebesgue measure equal zero.

1.3.1 Mathematical introduction to copulas

Definition 1.3.2 *Let $S_1, S_2, ..., S_n$ be nonempty subsets of $\overline{\Re}$, where $\overline{\Re}$ denotes the extended real line $[-\infty, \infty]$. Let H be a real function of n variables such that $DomH = S_1 \times S_2 \times ... \times S_n$ and for $\boldsymbol{a} \leq \boldsymbol{b}$ ($a_k \leq b_k$ for all k) let $B = [\boldsymbol{a}, \boldsymbol{b}]$ $(= [a_1, b_1] \times ... \times [a_n, b_n])$ be an n-box whose vertices are in $DomH$. Then the H-volume of B is given by*

$$V_H(B) = \sum \mathrm{sgn}(\boldsymbol{c}) H(\boldsymbol{c}),$$

where the sum is taken over all vertices $\boldsymbol{c}$ of B, and sgn($\boldsymbol{c}$) is given by

$$\mathrm{sgn}(\boldsymbol{c}) = \begin{cases} 1, & \text{if } c_k = a_k \text{ for an even number of } k'\text{s}, \\ -1, & \text{if } c_k = a_k \text{ for an odd number of } k'\text{s}. \end{cases}$$

Equivalently, the H-volume of an n-box $B = [\boldsymbol{a}, \boldsymbol{b}]$ is the nth order difference of H on B

$$V_H(B) = \Delta_a^b H(\boldsymbol{t}) = \Delta_{a_n}^{b_n} ... \Delta_{a_1}^{b_1} H(\boldsymbol{t}),$$

where the n first order differences are defined as

$$\Delta_{a_k}^{b_k} H(\boldsymbol{t}) = H(t_1, ..., t_{k-1}, b_k, t_{k+1}, ..., t_n) - H(t_1, ..., t_{k-1}, a_k, t_{k+1}, ..., t_n).$$

Definition 1.3.3 *A real function H of n variables is n-increasing if $V_H(B) \geq 0$ for all n-boxes B whose vertices lie in $DomH$.*

Suppose that the domain of a real function H of n variables is given by $DomH = S_1 \times \ldots \times S_n$ where each S_k has the smallest element a_k . We say that H is grounded if $H(\mathbf{t}) = 0$ for all t in $DomH$ such that $t_k = a_k$ for at least one k . If each S_k is nonempty and has a greatest element b_k, then H has margins, and the one-dimensional margins of H are the functions H_k with $DomH_k = S_k$ and with $H_k(x) = H\left(b_1, ..., b_{k-1}, x, b_{k+1}, ..., b_n\right)$ for all x in S_k.

Lemma 1.3.4 *Let S_1 and S_2 be nonempty subsets of $\bar{\Re}$, and let H be a grounded 2-increasing function with domain $S_1 \times S_2$. Then H is nondecreasing in each argument.*

Proof: Let a_1, a_2 denote the least elements in S_1 and S_2 respectively and b_1, b_2 are their respective greatest elements. We say that a function H from $S_1 \times S_2$ into $\Re$ is grounded if $H(x, a_2) = 0 = H(a_1, y)$ for all (x, y) in $S_1 \times S_2$. Then the function $t \mapsto H\left(t, b_2\right) - H\left(t, a_2\right)$ is nondecreasing on S_1 and the function $t \mapsto H\left(b_1, t\right) - H\left(a_1, t\right)$ is nondecreasing on S_2. We then say that the function H from $S_1 \times S_2$ into $\Re$ has margins, and the margins of H are the functions F and G given by: $DomF = S_1$, and $F(x) = H(x, b_2)$ for all x in S_1; $DomG = S_2$, and $G(y) = H(b_1, y)$ for all y in S_2.

Lemma 1.3.5 *Let S_1 and S_2 be nonempty subsets of $\bar{\Re}$, and let H be a grounded*

2-increasing function with domain $S_1 \times S_2$. Let (x_1, y_1) and (x_2, y_2) be any points in $S_1 \times S_2$, then

$$|H(x_2, y_2) - H(x_1, y_1)| \leq |F(x_2) - F(x_1)| + |G(y_2) - G(y_1)|.$$

Proof: From the triangle inequality, we have

$$|H(x_2, y_2) - H(x_1, y_1)| \leq |H(x_2, y_2) - H(x_1, y_2)| + |H(x_1, y_2) - H(x_1, y_1)|.$$

We assume that $x_1 \leq x_2$. Because H is grounded, 2-increasing, and has margins, Lemma 1.3.4 yields $0 \leq H(x_2, y_2) - H(x_1, y_2) \leq F(x_2) - F(x_1)$. An analogous inequality holds when $x_2 \leq x_1$, hence it follows that for any x_1, x_2 in S_1, $|H(x_2, y_2) - H(x_1, y_2)| + |F(x_2) - F(x_1)|$. Similarly for any y_1, y_2, in S_2, $|H(x_1, y_2) - H(x_1, y_1)| + |G(y_2) - G(y_1)|$, hence the proof is complete.

Definition 1.3.6 *An n-dimensional distribution function is a function H with domain $\overline{\Re}^n$ such that H is grounded, n-increasing and $H(\infty, \infty, ..., \infty) = 1$.*

It follows from Lemma 1.3.4 that the margins of an 2-dimensional distribution function are distribution functions, which we denote by F_1 and F_2.

Definition 1.3.7 *(Copula function) A two-dimensional copula function (or a 2-copula) is defined as a binary function* $C : [0,1]^2 \to [0,1]$ *, which satisfies the following three properties:*

1. $C(u, 0) = C(0, u) = 0$ *for any* $u \in [0, 1]$.
2. $C(u, 1) = C(1, u) = u$ *for any* $u \in [0, 1]$.
3. *For all* $0 \leq u_1 \leq u_2 \leq 1$ *and* $0 \leq v_1 \leq v_2 \leq 1$

   $C([u_1, v_1] \times [u_2, v_2]) = C(u_2, v_2) - C(u_1, v_2) - C(u_2, v_1) + C(u_1, v_1) \geq 0.$

Due to the properties 1-3, when the arguments u and v are univariate distribution functions F_1 and F_2, the copula function $C(F_1; F_2)$ is a legitimate bivariate distribution function with marginals F_1 and F_2. Conversely, any bivariate distribution function $H(x; y)$ with continuous marginals F_1 and F_2 admits a unique representation as a copula function:

$$C(u, v) = H\left(F_1^{-1}(u), F_2^{-1}(v)\right)$$

In general, an n-dimensional Copula is any function $C : [0,1]^n \to [0,1]$ such that:

1. C is grounded and n-increasing
2. C has margins $C_k, k = 1, 2, ..., n$, which satisfy $C_k(u) = u$ for all u in $[0, 1]$.

It is also important to note that for any n-copula, $n \geq 3$, each k-dimensional margin of C is a k-copula. Equivalently, an n-copula C, is any function $C : [0,1]^n \to [0,1]$ with the following properties:

1. For every $\mathbf{u}$ in $[0,1]^n$, $C(\mathbf{u}) = 0$ if at least one coordinate of $\mathbf{u}$ is 0, and $C(\mathbf{u}) = u_k$ if all coordinates of $\mathbf{u}$ equal to 1 except u_k.

2. For every $\mathbf{a}$ and $\mathbf{b}$ in $[0,1]^n$ such that $a_i \leq b_i$ for all i, $V_C([a,b]) \geq 0$.

3. Since copulas are joint distribution functions on $[0,1]^n$, a copula C induces a probability measure on $[0,1]^n$ through

$$V_C([0,u_1] \times \ldots \times [0,u_n]) = C(u_1, \ldots, u_n)$$

   and a standard extension to arbitrary (not necessarily n-boxes) Borel subsets of $[0,1]^n$. Therefore, there is a unique probability measure on the Borel subset of $[0,1]^n$ which coincides with V_C on the set of n-boxes of $[0,1]^n$.

From Lemma 1.3.5 we have the following theorem.

Theorem 1.3.8 *Let C be an n-copula. Then for every $\boldsymbol{u}$ and $\boldsymbol{v}$ in $[0,1]^n$,*

$$|C(\boldsymbol{v}) - C(\boldsymbol{u})| \leq \sum_{k=1}^{n} |v_k - u_k|$$

Hence C is uniformly continuous on $[0,1]^n$. , see Nelsen (2006)

1.4 Motivation for Copula Models

Copulas are useful in the fields of finance and insurance, which are inseparable as discussed in Hipp (2007) . We now take a brief look at why we move into the copula models:

1. To start with, normal dependence is not appropriate: normally distributed random variables have a dependence completely specified by the correlation coefficient. This does not capture tail dependence, and it changes under nonlinear transforms. Copulas, therefore, allow for the inclusion of features such as fat tails and skewness for nonelliptically distributed risks.

2. Dependence matters: For asset liability management (life insurance) and dynamic financial analysis (non-life insurance) all occurring risks are considered and modelled. For the overall risk of a company also the dependence of these risks matters.

3. Tails matter: If two risks are X and Y, then these risks contribute much to the overall risk of the company if they are large at the same time, for instance, in the following sense:

$$P\{X > u \text{ and } Y > u\} > 0.5P\{X > u\}$$

4. The need for simple concepts: we need models of low complexity which can easily be calibrated to the data. Better a model that captures dependence, sufficiently well, than one which underestimates the dependence.

5. Lack of knowledge on dependence: Dependence between risks is caused by factors with simultaneous influence on these risks. However, the factors and the way they influence the risks are unknown in most situations. So a simple and more abstract modelling of dependence is needed.

6. Separation of size from dependence: risks X and Y can be large because they are heavy tailed; so they are both large with high probability, even though they are independent. To identify the influence of dependence on the overall risk one needs a dependence concept which is independent of the marginal distribution of the risks.

7. Dynamic Financial Analysis (DFA) and Asset-liability Management (ALM) approaches need concepts for simulation: an appropriate model for dependence will be used to simulate the cash flow of an insurer and to see the consequences of investment, reinsurance and new business. So the dependence model should be of a form which enables easy simulation.

1.5 Statement of the problem

Risk modelling is a vital undertaking for any company in the insurance business. Failure to sufficiently capture the underlying patterns of the probable risks (key among them emanating from the reported claims, and especially the very large claims) can have very serious implications, the worst being insolvency. When claims are reported from a single line of business to an insurance company, this form a simple univariate data set whose analysis is straight forward. However, the situation becomes more complicated when at one point in time; the company receives more than one claim and from different lines which sees a build up of a multivariate data. Additionally, when the number of assets in a portfolio is huge, it is almost unlikely to accurately estimate the joint probability distribution. The common approach to this challenge is to assume that the return on asset observes the multinormal distribution and carry on the simulation based on such assumption.

1.6 Objectives of the Study

1.6.1 General objective

Make use of copula models and statistical distances and to provide alternative procedures to risk modelling.

1.6.2 Specific objectives

1. Model the dependence between risks by the use of Copulas under complex structures.

2. Analyse the underlying claims distributions in a portfolio.

3. Model the claims in the upper tails of the distribution.

4. Since statistical distances have been used for testing similarities/dissimilarities and hence in the classification of points (or data vectors) into different sets, we will apply them to create some classification criterion for the claim severities.

5. After describing the underlying claims distribution of the company in question, the study utilize dependence measures to come up with a principle of determining an appropriate retention level in re-insurance arrangements.

1.7 Significance of the study

This study simplifies the complexities of modelling multivariate claims severities by making use of copulas and statistical distances. The work is also aimed at providing the insurance industry with yet another method at their disposal for their risk modelling and appraising. Potential investors will be able to understand the dependence between risks and hence arrive at a more diversified portfolio before their initial outlay. In addition to the further knowledge that will be created, the research will be a stepping ground for future research.

1.8 Scope and Limitations

The study aimed at formulating methods of modelling of claims severities that have a multivariate structure; which will encompass describing their distributions, estimating their dependencies, and Clustering the risks by the use of distances. In the case of risk transfer, via re-insurance arrangements, the existing criterion with be improved by the use of copulas. The limitation was that of the insurance companies were not willing to share out their data with other parties. This was tackled by using the Insurance Regulatory Authority database. Reading materials were also a major limitation.

1.9 Thesis Overview

This thesis is outlined as follows: In this closing chapter, we introduce the work by giving a background information in measures of dependence. We also state the problem, objectives, significance of the study, as well as the scope and limitation. In Chapter 2 we review the literature on the application of copulas in both finance and insurance. The Sklar's theorem, which is at the core of copula theory, is presented here together with families of copulas. Modelling the dependence between M lines of business with sub-classes is considered in Chapter 3 where the Hierarchical

Archimedean Copulas (HAC) are utilized as the main tool. We simulate data in Chapter 4 that follows four loss distributions and has an hierarchical structure. This data are applied on the theory developed in Chapter 3 as well as on the chapter that follow. Chapter 5 now deals in classification of risks. A clustering criterion is developed that is based on distances. Empirical data on insurance claims amounts experience in Kenya is the subject of Chapter 6. The methods developed based on artificial data in Chapter 5 are now applied on empirical data. Finally, we have the conclusions and recommendation in Chapter 7 that ends with an highlight of areas of further research.

Chapter 2

LITERATURE REVIEW

Copulas have found a widespread application in finance and actuarial science, particularly in the insurance industry. In this section we examine some of the previous studies that found the copula theory handy in the two areas: finance and insurance; as the insurance companies will use the copula theory to determine the intrinsic interdependencies between risks and also use the same theory to manage their finances given the risks' dependence structure.

2.1 Applications to Finance

In finance, Liu (2006) presents an application of copula methodology in modelling joint distributions with fat tails. Four widely established copulas were estimated and compared with the multivariate normal model. They are the Gaussian copula, the t-copula, the Gumbel copula and the mixture of Gumbel copula. The effect of the copula on the Value-at-Risk (VaR) is also studies. In addition, this study examines models with different portfolio allocations. Liu (2006) found that Copula is particularly useful in approximating the tails of portfolio returns. Next, when we move toward the centre of the joint distribution, this advantage is reduced. In addition, it also emerged that no uniformly best model is available when we consider portfolio allocations. This provided complementary results for other joint distribution modelling comparisons conducted in recent works and called for more specific comparison criterion.

The study observed that copula provides an alternative to the multivariate normal specification of the dependence between variables, and has gained increasing attention in asset pricing, portfolio allocation and risk management.

In risk management, more specifically option pricing, Goorbergh, Genest, and Werker (2005) examined the behaviour of bivariate option prices in the presence of association between the underlying assets. Parametric families of copulas offering various alternatives to the Gaussian dependence structure were used to model this association, which is explicitly assumed to vary over time as a function of the volatilities of the assets. These dynamic copula models were applied to better-of-two-markets and worse-of-two-markets options on the Standard and Poor's 500 and Nasdaq indexes.

Their results demonstrated that option prices implied by dynamic copula models can differ substantially from prices implied by models that fix the dependence between the underlyings, particularly in times of high volatilities. In the study, the Gaussian copula also produced option prices that differed significantly from those induced by non-Gaussian copulas, irrespective of initial volatility levels. Within the class of alternatives considered, option prices were robust with respect to the choice of copula.

Romano (2002) described some possible uses of copula functions in risk management applications and proved how some kind of copula functions are easy to implement in Monte Carlo simulation models to estimate risk measures. A practical application with a portfolio of ten Italian equities was performed and proved that the common hypothesis of multinormal distribution for asset returns (or risk factor returns) underestimates the VaR and the Expected Shortfall of a market portfolio. A Monte Carlo simulation, modelling asset returns using fat tail marginal distributions and a copula function with tail dependence was also performed and obtained a more accurate estimate for the two risk measures. The study also proved that a methodology using a multivariate Gaussian distribution of the latent variables does not capture the risk of many joint counterparty defaults. On the contrary, events of this kind are effectively captured using the Student-t copula to describe the dependence structure of the latent variables. Therefore, the Student-t copula can be very useful to model the extreme risk that worries risk managers and supervisors.

2.2 Applications of Copulas to Insurance

Besides copulas application in finance as a risk management tool, the following are just but a few applications in insurance, which is the focus of this study. Frees and Valdez (1997) introduced actuaries to the concept of copulas, as a tool for understanding relationships among multivariate outcomes. The work explored some of the practical applications, including estimation of joint life mortality and multiple decrement models. Statistical inference procedures are illustrated for those who wish to use copulas for statistical inference, by using insurance company data on losses and expenses. Using these data, Frees and Valdez (1997) showed how to fit copulas and described their usefulness by pricing a reinsurance contract and estimating expenses for pre-specified losses.

Copulas have also found their use in weather-based insurance. Filler, Odening, Okhrin, and Xu (2009) observed that systemic weather risk was a major obstacle for the formation of private (non-subsidized) crop insurance. Consequently, this study explored the possibility of spatial diversification of insurance by estimating the joint occurrence of unfavourable weather conditions in different locations. For this purpose, copula methods were employed that allow an adequate description of stochastic dependencies between multivariate random variables. The estimation procedure was applied to weather data in Germany. The results indicated that indemnity payments based on temperature as well as on cumulative rainfall exhibited strong stochastic dependence even at a national scale. Thus the possibility to reduce risk exposure by increasing the trading area of the insurance is limited. Irrespective of their economic implications the results pinpointed the necessity of a proper statistical modelling of the dependence structure of multivariate random variables. The usual approach of measuring stochastic dependence with linear correlation coefficients turned out to be questionable in the context of weather insurance as it may overestimate diversification effects considerably.

Gatzert, Schmeiser, and Schuckmann (2007) were motivated by the fact that in financial firms and insurance groups, enterprise risk management (ERM) is becoming increasingly important in controlling and managing the different independent legal entities in the group. Gatzert et al. (2007) assessed and related risk concentrations and joint default probabilities of legal entities in a corporation composed of three entities, a bank, a life insurance company, and a non-life insurance company. The procedure provided valuable insight regarding the group's risk situation, which is highly relevant for enterprise risk management purposes.

An insurance group (a conglomerate) typically consists of several legally independent entities, each with limited liability. However, diversification concepts assume that these entities are fully liable and all together meet all outstanding liabilities of each. Even if diversification is of no importance from a policyholder perspective, it is useful in determining risk concentration in an insurance group because greater diversification generally implies less risk. To determine default probabilities, Gatzert et al. (2007) focused on the case of limited liability without transfer of losses between the different legal entities within the group. Joint default probabilities only depend on individual default probabilities and the coupling dependence structure. Hence, the study dwelt on the effect of different dependence structures using the concept of copulas.

For the numerical analysis, Gatzert et al. (2007) considered an insurance group comprised of three legal entities and compared results from the Gauss, Gumbel, and Clayton copulas for normal and non-normal marginal distributions. Economic capital was adjusted for each situation to satisfy a fixed individual default probability. In contrast to the risk concentration factor, joint default probabilities only depend on individual default probabilities and on the dependence structure, but not on distributional assumptions. They further found that the risk concentration factor and the joint default probability of all three entities increase with increasing dependence between the entities, while the probability of a single default decreases. Overall, the sum of default probabilities of one, two, or three entities decreases with increasing dependence. Furthermore, one entity's large risk contribution, in terms of volatility, led to a much higher risk concentration factor for the group as a whole. The findings further demonstrated that even if different dependence structures imply the same risk concentration factor for the group, joint default probabilities for different sets of subsidiaries can vary tremendously. In particular, the lower tail dependent Clayton copula led to the lowest probability of default for all three entities, while the upper tail dependent Gumbel copula exhibited the highest default probability. The analysis showed that a simultaneous consideration of risk concentration factor and default probabilities can be of substantial value, especially for the management of the corporate group with respect to enterprise risk management.

Copulas have also been applied in credibility theory which is a form of insurance pricing. The theory of credibility has been called a "basis" of the field of actuarial science. Frees and Wang (2004) develops a direct link between credibility and loss distributions through the notion of a copula. The work develops credibility using a longitudinal data framework. In a longitudinal data framework, one might encounter data from a cross-section of risk classes (towns) with a history of insurance claims available for each risk class. For the marginal claims distributions, Frees and Wang (2004) use generalized linear models, an extension of linear regression that also encompasses Weibull and Gamma regressions. Copulas are used to model the dependencies over time; and this study was the first to propose the use of a t-copula in the context of generalized linear models. The t-copula is the copula associated with the multivariate t-distribution; like univariate t-distributions, it seems especially suitable for empirical work. Moreover, they show that the t-copula gives rise to easily computable predictive distributions that we use to generate credibility predictors. Like Bayesian methods, the copula credibility prediction methods allow us to provide an entire distribution of predicted claims, not just a point prediction. Frees and Wang (2004) present illustrative example of Massachusetts automobile claims, and compare their new credibility estimates with those currently existing in the literature.

In this study, claims from a Gamma family were used and provided the necessary theoretical underpinnings for the exponential family of distributions that also includes the normal and Weibull distributions. Although any parametric family of copulas fits within the framework described in Frees and Wang (2004), the work explores the advantages of the t-copula. They find that this is a desirable dependence structure, at least for the bodily injury liability automobile claims data investigated. Using their data, they compared the copula-based credibility predictors and found that they performed well compared to traditional credibility estimators. They even demonstrated the well-known *shrinkage* characteristic that actuaries find appealing for traditional estimators. This may not be a general characteristic of copula-based credibility predictors.

The contribution of copulas in the incurred but not reported (IBNR) claims due to an insurance company was examined by Pettere and Kollo (2006) who studied the claims of a Latvian Insurance company. The two variables investigated in this work were claim size and time from the moment when claim occurred to the moment when the payment has been reported. They found approximations to the distributions of both variables as well as the bivariate distribution modelled using the Archimedean copulas. The IBNR claim reserves were calculated using the bivariate Clayton copula model. Pettere and Kollo (2006) found that for the claim size, lognormal distribution was the best model but Wald distribution can also be used. Moreover, the development factor can be described by the lognormal distribution too. It also emerged that when approximating the bivariate distribution of the claim size and the development factor with Archimedean copulas, the Clayton copula gave the best model and so, the bivariate Clayton copula were used in simulations estimating the IBNR reserves. Applicability of the method of estimation IBNR reserves was checked on data from past of the company. The real empirical values for necessary reserves were in good accordance with the predicted estimate using Clayton copula. Owing to the increase in popularity of copulas to measure dependent risks, generating multivariate copulas has become a very crucial exercise. Wu, Valdez, and Sherris (2006) noted that multivariate exchangeable Archimedean copulas are one of the most popular classes of copulas that are used in actuarial science and finance for modelling risk dependencies and for using them to quantify the magnitude of tail dependence. Current methods for generating multivariate Archimedean copulas can again become a very difficult task as the number of dimension increases. Wu et al. (2006) presented an algorithm for generating multivariate exchangeable Archimedean copulas based on a multivariate extension of a bivariate result after observing that the resulting analytical procedures suggested in the existing literature did not offer much guidance for practical implementation. Genest and Rivest (1993) proposed a procedure for generating bivariate Archimedean copulas and again later described in Nelsen (2006) and Embrechts, McNeil, and Straumann (2002). Wu et al. (2006) were able to extend the bivariate result into the multivariate case and hence developing an interesting algorithm to generate exchangeable Archimedean copulas using a proof that is simply based on fundamental Jacobian techniques for deriving distributions of transformed random variables. They were, again, able to derive the distribution function of an n-dimensional Archimedean copula, a result already known in Genest and Rivest (2001) but their approach of proving this result was based on a different perspective. In order to demonstrate the usefulness and the reasonableness of the results, they considered illustrative examples of generating from the Gumbel-Hougaard and the Frank family of Archimedean copulas. It was found that the simulation results appeared to be reasonable as expected. In terms of practicality, they also provided illustration of evaluating the extra capital required for the addition of a new line of business. Their illustrations show that both the dependency risk and the choice of the correlation coefficient will have significant impact on the amount of capital required for an insurance company running multiple lines of business.

Frees and Valdez (2008) demonstrated actuarial applications that can be performed when modern statistical methods are applied to detailed, micro-level automobile insurance records. They tested their model using 1993-2001 data consisting of policy, claims and payment files from a major Singapore insurance company. The model allowed them to study the accident frequency, loss type and severity jointly and to incorporate individual characteristics such as age, gender and driving history that explain heterogeneity among policyholders. Based on this hierarchical model, one can analyze the risk profile of either a single policy (micro-level) or a portfolio of business (macro-level). In their work Frees and Valdez (2008) investigate three types of actuarial applications. First, the calculation of the predictive mean of losses for individual risk rating as this allows the actuary to differentiate prices based on policyholder characteristics. The second application was that of the predictive distribution of a portfolio of business. They demonstrated the calculation of various risk measures, including value at risk and conditional tail expectation that are useful in determining economic capital for insurance companies.

For a sensitivity analysis, they incorporated the copulas in two ways: first, assumed that the specified copula was consistently used for the estimation and prediction portions and second they assumed that the (correct but more complex) t-copula was used for estimation with the specified copula used for prediction. The idea behind the second assumption was that a statistical analysis unit of a company may perform a more rigorous analysis using a t-copula and another unit within a company may wish to use this output for quicker calculations about their financial impact. The study by Frees and Valdez (2008) found that the copula effect was large and increased with the percentile. The upper percentiles are the most important to the actuary for many financial implications.

A large difference existed between assuming independence among coverages and using a t-copula to quantify the dependence. They also found that when re-estimating the full model under alternative copulas, the marginal parameters changed to produce significant differences in the risk measures. The results for the independence copula were somewhat counterintuitive. For most portfolios, with positive correlations among claims, one typically needs to go out further in the tail to achieve a desired percentile, suggesting that the VaR should be larger for the t-copula than the independence copula.It is also important to note that Frees and Valdez (2008) found that the VaR is not affected by the choice of copula. In contrast, for the CTEs, the normal and t-copula give higher values than the independence copula. This result was due to the higher losses in the tail under the normal and t-copula models. The third application was the examination of the effects of several reinsurance treaties, that is, to show the predictive loss distributions for both the insurer and reinsurer under quota share and excess-of-loss reinsurance agreements.

Motivated by the fact that most of the Economic Capital assessment models encounter difficulties when trying to incorporate the dependence of claim costs between different Lines of Business (LOBs), Faivre (2003) suggested the use of copula theory as a solution to this problem. When a copula is applied to marginal distributions that are subject to some minor technical requirements, but do not necessarily belong to the same distribution family it results in a proper multivariate distribution. The study used this property to model the overall distribution of claim costs of a four-LOB company. By using different copulas Faivre (2003) was able to show that the dependence structure has a substantial impact on the Economic Capital of that firm and in particular the research paid attention to situations in which attritional losses from different LOBs compensate for each other to produce stable total results, while at the same time extreme losses tends to occur simultaneously across different LOBs. Faivre (2003) observed that this type of dependence could not be modelled by the multivariate normal distribution, which is at the root of many current Economic Capital assessment models. With the help of the copula model Faivre (2003) showed that the assumption that claim costs are independent from one LOB to another, when they are not, could result in an underestimation of the required Economic Capital for high percentiles. In particular, among all the scenarios that could lead to a capital underestimation the study pointed out that by using a Student-t copula, we can model a situation where attritional losses produce technical results that tend to compensate each other, while extreme losses tend to occur during the same accident year across different LOBs. It was also observed that even if the Gumbel copula may not reflect a realistic situation, it can produce stress scenarios useful to test the solvency of a company.

2.2.1 Sklar's Theorem

The following theorem is known as Sklar's Theorem (Sklar, 1958) which is perhaps the most important result regarding copulas, and is used in essentially all applications of copulas.

Theorem 2.2.1 *(Sklar's Theorem) Let H be an n-dimensional distribution function with margins $F_1, \ldots, F_n$. Then there exists an n-copula C such that for all $\boldsymbol{x}$ in $\overline{\Re}^n$,*

$$H(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n)).$$

If $F_1, \ldots, F_n$ are all continuous, then C is unique; otherwise C is uniquely determined on $RanF_1 \times \ldots \times RanF_n$. Conversely, if C is an n-copula and $F_1, \ldots, F_n$ are distribution functions, then the function H defined above is an n-dimensional distribution function with margins $F_1, \ldots, F_n$.

From the Sklar's Theorem we see that for continuous multivariate distribution functions, the univariate margins and the multivariate dependence structure can be separated, and the dependence structure can be represented by a copula.

Corollary 2.2.2 *Let H be an n-dimensional distribution function with continuous margins $F_1, \ldots, F_n$ and copula C. Then for any $\boldsymbol{u}$ in $[0, 1]^n$,*

$$C(u_1, ..., u_n) = H(F_1^{-1}(u_1), ..., F_n^{-1}(u_n)). \tag{2.2.1}$$

, see Nelsen (2006)

2.2.2 Copula densities

As per the definition, a copula is a cumulative distribution function. It is quite typical for these monotonically increasing functions, that even though being theoretically very powerful their graphs are hard to interpret. Owing to this, typically plots of densities are used to illustrate distributions, rather than plots of the cumulative distribution function. Certainly, not in all cases copulas do have densities as will be pointed out later. However, if the copula is sufficiently differentiable the copula density can be computed:

$$c(\mathbf{u}\ ) = \frac{\partial^n C(u_1, ..., u_n)}{\partial u_1 ... \partial u_n}$$

Contrary to this, it may be the case that besides an absolutely continuous component (represented by the density) the copula also has a singular component. If the copula is given in Corollary 2.2.2 we obtain the copula density in terms of the joint density together with marginal cumulative distribution functions and marginal densities. Note that it is necessary that the cumulative distribution function is differentiable. Denoting the joint density by f and the marginal densities by $f_i, i = 1, 2, ..., n$, it follows, by the chain rule, that

$$c(\mathbf{u}) = \frac{f\left(F_1^{-1}(u_1), ..., F_n^{-1}(u_n)\right)}{f_1\left(F_1^{-1}(u_1)\right) ... f_n\left(F_n^{-1}(u_n)\right)}$$

2.2.3 Conditional distributions

Dependency is an important concept inferring outcomes from a random variable based on the knowledge of a related factor. Let us consider two uniform random variables U_1 and U_2 with known copula C and U_1 is observed. The goal is to deduce the conditional distribution which can then be used for predicting or estimating U_2. Assuming sufficient regularity, we obtain for the conditional cumulative distribution function

$$\begin{aligned} P\left(U_2 \leq u_2 \left| U_1 = u_1\right.\right) &= \lim_{\delta \to 0} \frac{P\left(U_2 \leq u_2, U_1 \in \left(u_1 - \delta, u_1 + \delta\right]\right)}{P\left(U_1 \in \left(u_1 - \delta, u_1 + \delta\right]\right)} \\ &= \lim_{\delta \to 0} \frac{C\left(u_1 + \delta, u_2\right) - C\left(u_1 - \delta, u_2\right)}{2\delta} \\ &= \frac{\partial}{\partial u_1} C\left(u_1, u_2\right) \end{aligned}$$

Hence, the conditional cumulative distribution function may be derived directly from the copula itself. The conditional density function is obtained by deriving once more with respect to u_2 . In most cases, the best estimator of U_2 will be the conditional expectation which of course is directly obtained from the conditional density function.

2.2.4 Fréchet-Hoeffding copula boundaries

Hoeffding and Fréchet independently showed that a copula always lies in between certain bounds. The reason for this is the existence of some extreme cases of dependency, Nelsen (2006). *Minimum copula:* This is the lower bound for all copulas. In the bivariate case only, it represents perfect negative dependence between variates:

$$W\left(u,v\right)=\max\left(0,u+v-1\right)$$

For n-variate copulas, the lower bound is given by

$$W\left(u_1,u_2,...,u_n\right)=\max\left\{1-n+\sum_{i=1}^{n}u_i,0\right\}.$$

Maximum copula: This is the upper bound for all copulas. It represents perfect positive dependence between variates:

$$M\left(u,v\right)=\min\left(u,v\right)$$

For n-variate copulas, the upper bound is given by

$$C\left(u_1,u_2,...,u_n\right)=\min_{j\in\{1,...,n\}}u_j=M\left(u_1,u_2,...,u_n\right)$$

For all 2-copulas $C(u,v)$,

$$W(u,v) \leq C(u,v) \leq M(u,v)$$

while in the multivariate case, the corresponding inequality is

$$W(u_1,u_2,...,u_n) \leq C(u_1,...,u_n) \leq M(u_1,...,u_n).$$

The random variables X and Y are *comonotonic* if they have the Fréchet-Heoffding upper copula, $M(x,y) = Min(x,y)$. Comonotonic is an example of concordance where X and Y are perfectly positive dependent. Further, they are *counter-monotonic* if they have the Fréchet-Heoffding lower copula, $W(x,y) = max(x+y-1,0)$. Counter-monotonic is an example of discordance where X and Y are perfectly negative dependent.

From Figure 2.1 the surface given by the bottom and back side of the pyramid (the lower bound) is the counter-monotonic copula $C(u,v) = \max\{u+v-1,0\}$, while the front side is the upper bound, $C(u,v) = \min(u,v)$.

2.2.5 Important copulas

From the entire set of copulas, three are of special nature: $C^{-}, C^{\perp}$ and C^{+} , which are respectively the copulas of *Countermonotony*, *Independence*, and *Comonotony*. As noted earlier, countermonotony is the extreme negative dependence, independence

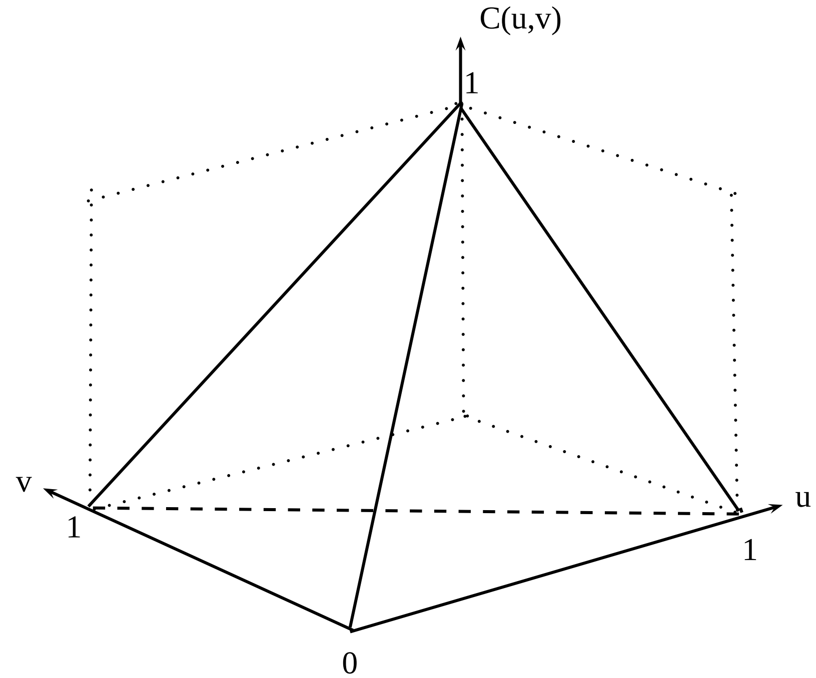


Figure 2.1: *The Fréchet-Hoeffding bounds showing that every copula has to lie inside of the shown pyramid*

is the absence of dependence, and comonotony is the extreme positive dependence structure. The independence copula is shown in Figure 2.2. Mathematically:

$$C^{-}(u_1, u_2, ..., u_n) = \max\left\{1 - n + \sum_{i=1}^{n} u_i, 0\right\}$$

$$C^{\perp}(u_1, u_2, ..., u_n) = \prod_{i=1}^{n} u_i$$

$$C^{+}(u_1, u_2, ..., u_n) = \min_{j \in t\{1,...,n\}} u_j$$

where $u_1, u_2, ..., u_n \in [0; 1]$.

$C^{\perp}$ and C^{+} are copulas whatever the dimension n, but C^{-} is only a copula in the

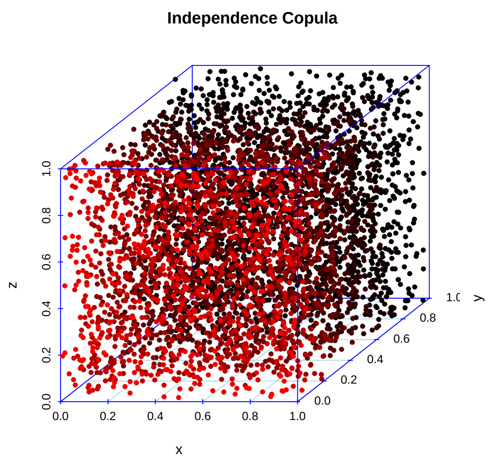


Figure 2.2: *Scatter plot for the independence copula*

bivariate case. However C^- is for any dimension n (n greater than two) the point-wise best possible bound, that is for any u in $DomC^-$ there exists a copula C such that $C(u) = C^-(u)$.

Definition 2.2.3 *A copula C_1 is said to be smaller than a copula C_2 if*

$$\forall \left(u_1, u_2, ..., u_n\right) \in [0,1]^n\,,\;\; C_1\left(u_1, u_2, ..., u_n\right) \leq C_2\left(u_1, u_2, ..., u_n\right)$$

written $C_1 \prec C_2$ where $\prec$ is called the concordance ordering, and is related to the first order stochastic dominance over cumulative distribution functions.

It is important to note that the ordering $C^- \prec C \prec C^+$ holds for any copula C and from Definition 2.2.3, a copula is said to be a positive dependence structure if $C^\perp \prec C \prec C^+$ and a copula is a negative dependence structure if $C^- \prec C \prec C^\perp$. Since $\prec$

is a partial ordering and that there exist copulas that are neither positive nor negative dependence structures, Definition 2.2.3 provides us with a first classification among copulas, though it remains theoretical. That is the reason we introduce measures of the dependence induced by the copula.

2.3 Families of Copula

Copulas have been distinguished in the *Elliptical* and *Archimedean* families until Alfonsi and Brigo (2005) who introduced new families of copulas based on *periodic* functions.

2.3.1 Elliptical copulas

Elliptical copulas are the copulas with elliptical distributions. They have an elliptical form and therefore symmetry in the tails. Important copulas in this family are the Gaussian and the Student's t-copula. One way of obtaining families of copula is by inversion from known bivariate distribution families. For an elliptical copula, the distribution is equation 2.2.1. Evaluation of equation 2.2.1 needs implementation of the joint CDF of the elliptical distribution and univariate quantile functions for each margin. Differentiating equation 2.2.1 gives the density of an elliptical copula

$$c\left(u_1,...,u_n\right)=\frac{f\left[F_1^{-1}\left(u_1\right),...,F_n^{-1}\left(u_n\right)\right]}{\prod\limits_{i=1}^{n} f_i\left[F_i^{-1}\left(u_i\right)\right]}$$

where f is the joint PDF of the elliptical distribution, and $f_1, \ldots, f_n$ are marginal density functions.

2.3.1.1 The Gaussian copula

For 2-dimensional case, the Gaussian copula may be represented as:

$$\begin{aligned} C^{Ga}(u, v; \rho) &= \Phi_\Sigma^2 \left(\Phi^{-1}(u), \Phi^{-1}(v)\right) \\ &= \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right\} dxdy \end{aligned}$$

where $\Phi(x)$ is the standard normal distribution function and $\Phi_\Sigma^2(x, y)$ is the bivariate normal distribution with zero mean and correlation ρ between the marginals, that is, Σ is a 2 x 2 matrix with 1 on the diagonal and ρ otherwise.

For normal distribution independence is equivalent to zero correlation. Hence for $\rho = 0$, the Gaussian copula equals the independence copula. On the other side, if $\rho = 1$ we obtain the comonotonicity copula, while for $\rho = -1$ the counter-monotonicity copula is obtained. The intuition of positive or negative dependence recurs in the form of positive/negative linear dependence. Thus the Gaussian copula interpolates between these three fundamental dependency structures via one simple parameter, the correlation ρ . A copula with this feature is said to be comprehensive.

We also note that the covariance matrix Σ, used here, is not arbitrary. It is a correlation matrix, which is obtained from an arbitrary covariance matrix by scaling each component to variance 1, which does not change the resultant copula as it is scale invariant.The copula of the n-variate normal distribution with covariance matrix Σ is

$$C_{\rho}^{Ga}(\mathbf{u}) = \Phi_{\rho}^{n}\left(\Phi^{-1}(u_1), ..., \Phi^{-1}(u_n)\right),$$

where Φ_{ρ}^{n} denotes the joint distribution function of the n-variate standard normal distribution function with covariance matrix Σ, and Φ^{-1} denotes the inverse of the distribution function of the univariate standard normal distribution. This can explicitly be written as:

$$c(u_1, ..., u_n \,|\Sigma) = |\Sigma|^{-\frac{1}{2}} \exp\left\{\frac{1}{2}\mathbf{c}^T\left(I_n - \Sigma^{-1}\right)\mathbf{c}\right\}$$

where $\mathbf{c} = (q_1, ..., q_n)^T$ with $q_i = \Phi^{-1}(u_i)$ for $i = 1, ..., n$ and Φ is the CDF of $N(0,1)$.

Gaussian copulas do not have upper tail dependence. Since elliptical distributions are radially symmetric, the coefficient of upper and lower tail dependence are equal. Hence Gaussian copulas do not have lower tail dependence either. This can be shown as follows:

Let $(X,Y)^T$ have the bivariate standard normal distribution function with linear correlation coefficient $\rho = 1$. That is $(X, Y)^T \sim C(\Phi(x), \Phi(y))$, where C is a member of the Gaussian family given above with $\rho_{12} = \rho$. Since copulas in this family are exchangeable,

$$\lambda_U = 2 \lim_{u \nearrow 1} P\{V > u \,|U = u\},$$

and because Φ is a distribution function with infinite right endpoint,

$$\lim_{u \nearrow 1} P\{V > u \,|U = u\} = \lim_{x \to \infty} P\left\{\Phi^{-1}(V) > x \,\middle|\Phi^{-1}(U) = x\right\} = \lim_{x \to \infty} P\{X > x \,|Y = x\}.$$

Using the fact that $Y \,|X = x \sim N(\rho x, 1 - \rho^2)$ we obtain

$$\lambda_U = 2 \lim_{x \to \infty} \bar{\Phi}\left((x - \rho x) \Big/ \sqrt{1 - \rho^2}\right) = 2 \lim_{x \to \infty} \bar{\Phi}\left(x\sqrt{1 - \rho} \Big/ \sqrt{1 + \rho}\right),$$

from which it follows that $\lambda_U = 0$ for $\rho_{12} < 1$. Hence the Gaussian copula C with $\rho < 1$ does not have upper tail dependence.

2.3.1.2 The t - copula

If $\mathbf{X}$ has the stochastic representation

$$\mathbf{X} = d\mu + \frac{\sqrt{\nu}}{S}\mathbf{Z}, \tag{2.3.1}$$

where $\mu \in \Re^n$, $S \sim \chi^2_\nu$ and $\mathbf{Z} \sim N_n(\mathbf{0}, \Sigma)$ are independent, then $\mathbf{X}$ has an n-variate

t_ν-distribution with mean μ (for $\nu > 1$) and covariance matrix $\frac{\nu}{\nu-2}\Sigma$ (for $\nu > 2$). If $\nu \leq 2$ then $\mathrm{Cov}(\mathbf{X})$ is not defined. In this case we just interpret Σ as being the shape parameter of the distribution of $\mathbf{X}$. The copula of $\mathbf{X}$ given by equation 2.3.1 can be written as

$$C_{\nu,\rho}^{t}\left(\mathbf{u}\right) = t_{\nu,\rho}^{n}\left(t_{\nu}^{-1}\left(u_1\right), ..., t_{\nu}^{-1}\left(u_n\right)\right),$$

where $\rho_{ij} = \Sigma_{ij} \Big/ \sqrt{\Sigma_{ii}\Sigma_{jj}}$ for $i, j \in \{1, \ldots, n\}$ and where $t_{\nu,\rho}^{n}$ denotes the distribution function of $\sqrt{\nu}\mathbf{Y} \Big/ \sqrt{S}$ where $S \sim \chi_\nu^2$ and $\mathbf{Y} \sim N_n\left(\mathbf{0}, \rho\right)$ are independent. Here t_ν denotes the (equal) margins of $t_{\nu,\rho}^{n}$, that is, the distribution function of $\sqrt{\nu}Y_1 \Big/ \sqrt{S}$. In the bivariate case the copula expression can be represented as:

$$\begin{aligned} C^t(u, v; \rho, \nu) &= t_{\rho,\nu}^{2}\left(t_\nu^{-1}(u), t_\nu^{-1}(v)\right) \\ &= \int_{-\infty}^{t_\nu^{-1}(u)} \int_{-\infty}^{t_\nu^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho_{12}^2}} \frac{1}{\left\{1 + \frac{x^2 - \rho_{12}xy + y^2}{\nu(1-\rho_{12}^2)}\right\}^{1+\nu/2}} dxdy \end{aligned}$$

where $t_\nu(x)$ is t-distribution with ν degrees of freedom and $t_{\rho,\nu}^{2}(x, y)$ is bivariate t-distribution with correlation ρ. Note that ρ_{12} is simply the usual linear correlation coefficient of the corresponding bivariate t_ν-distribution if $\nu > 2$. If $(X_1, X_2)^T$ has a standard bivariate t-distribution with ν degrees of freedom and linear correlation matrix ρ, then $X_2 \left| X_1 = x\right.$ is t-distributed with $\nu + 1$ degrees of freedom and

$$E\left(X_2 \left| X_1 = x\right.\right) = \rho_{12}x,\ Var\left(X_2 \left| X_1 = x\right.\right) = \left(\frac{\nu + x^2}{\nu + 1}\right)\left(1 - \rho_{12}^2\right).$$

This result can be used to show that the t-copula has upper tail dependence (and because of radial symmetry) equal lower tail dependence:

$$\begin{aligned}
\lambda_U &= 2 \lim_{x\to\infty} P\left(X_2 > x \,|X_1 = x\right) \\
&= 2 \lim_{x\to\infty} \bar{t}_{\nu+1}\left(\left(\frac{\nu+1}{\nu+x^2}\right)^{1/2} \frac{x-\rho_{12}x}{\sqrt{1-\rho_l^2}}\right) \\
&= 2 \lim_{x\to\infty} \bar{t}_{\nu+1}\left(\left(\frac{\nu+1}{\nu/x^2+1}\right)^{1/2} \frac{\sqrt{1-\rho_{12}}}{\sqrt{1+\rho_{12}}}\right) \\
&= 2\bar{t}_{\nu+1}\left(\sqrt{\nu+1}\sqrt{1-\rho_{12}} \Big/ \sqrt{1+\rho_{12}}\right)
\end{aligned}$$

The coefficient of upper tail dependence is increasing in ρ_{12} and decreasing in ν, as one would expect. Furthermore, the coefficient of upper (lower) tail dependence tends to zero as the number of degrees of freedom tends to infinity for $\rho_{12} > 1$.

2.3.1.3 The Cauchy copula

The Cauchy copula is actually a special case of the Student-t copula where the degrees of freedom is $\nu = 1$. The copula generated by a multivariate Cauchy distribution with linear correlation matrix Σ is given by

$$C\left(u_1, ..., u_n\right) = T_1\left(t_1^{-1}\left(u_1\right), ..., t_1^{-1}\left(u_n\right)\right)$$

where T_1 then is the joint distribution function of a standard Cauchy random vector expressed as

$$T_1(x_1, ..., x_n) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\pi^{n/2}\sqrt{|\Sigma|}} \int_{-\infty}^{x_n}\int_{-\infty}^{x_{n-1}} ... \int_{-\infty}^{x_1} \left(1 + z^T\Sigma^{-1}z\right)^{-\frac{(n+1)}{2}} dz_1...dz_n$$

and $t_1^{-1}(.)$ is the inverse of a standard Cauchy distribution with

$$t_1(z) = \int_{-\infty}^{z} \frac{1}{\pi} \cdot \left(\frac{1}{1+\omega}\right)^z d\omega.$$

Due to its relationship with the Student-t copulas, the Cauchy copulas also have non-zero tail dependence.

2.3.2 Elliptical copulas graphics

The following graphicals were plotted from simulated data in order to aid in the visualisation of the elliptical copulas.

Notice the dependence at the tails of each distribution. The t-distribution has a more flexible structure since increasing the degrees of freedom, the scatter plot will approach the normal copula as displayed Figure 2.4.

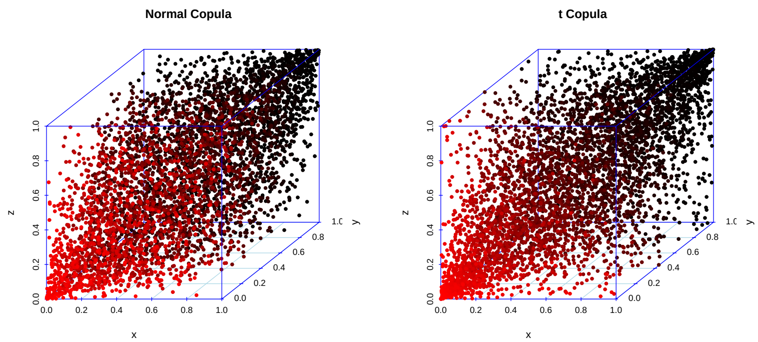


Figure 2.3: *Scatter plots of 5000 random numbers from a normal copula and a t-copula. Both have a dependence parameter of 0.5 with the t-distribution having 3 df.*

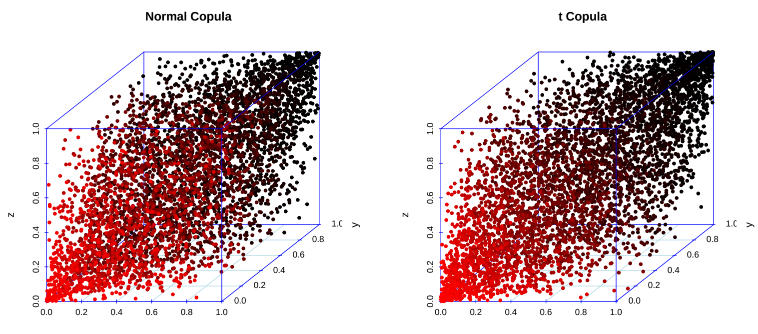


Figure 2.4: *Scatter plots of 5000 random numbers from a normal copula and a t-copula. Both have a dependence parameter of 0.5 with the t-distribution having 15 df.*

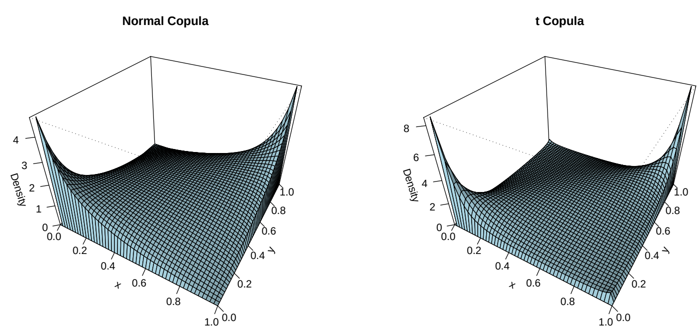


Figure 2.5: *Perspective plots for the Normal and t-copulas for the densities of random variables both of which have a dependence parameter of 0.5 and the t-distribution has 3df.*

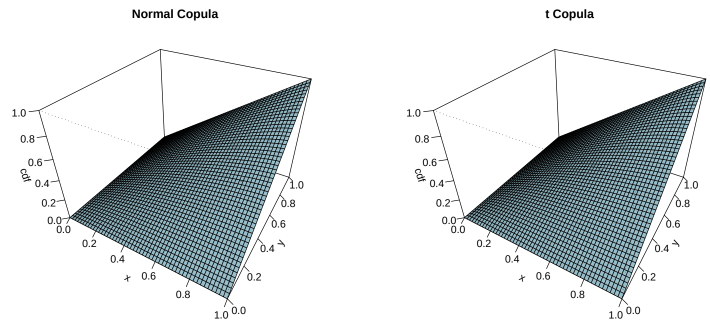


Figure 2.6: *CDFs for the Normal and t-copulas in figure 2.5 above.*

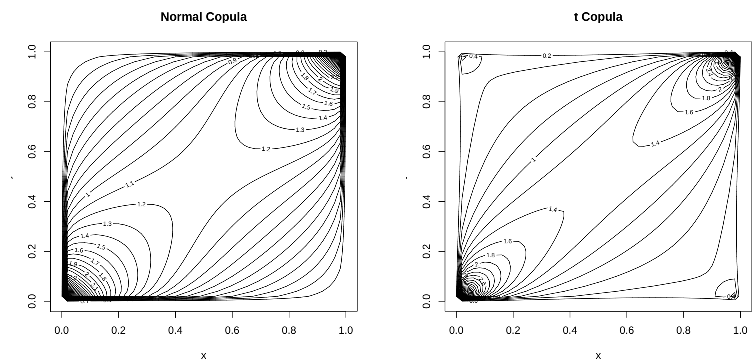


Figure 2.7: *Contour plots for the data in figure 2.5 above.*

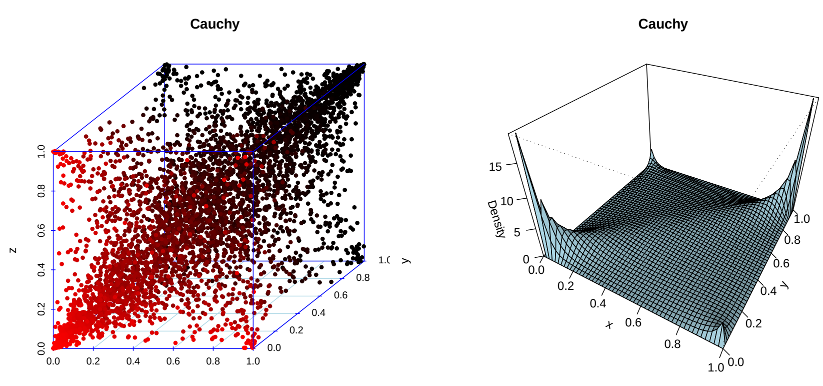


Figure 2.8: *Scatter and perspective plots for Cauchy distributed random variables with a dependence of 0.5.*

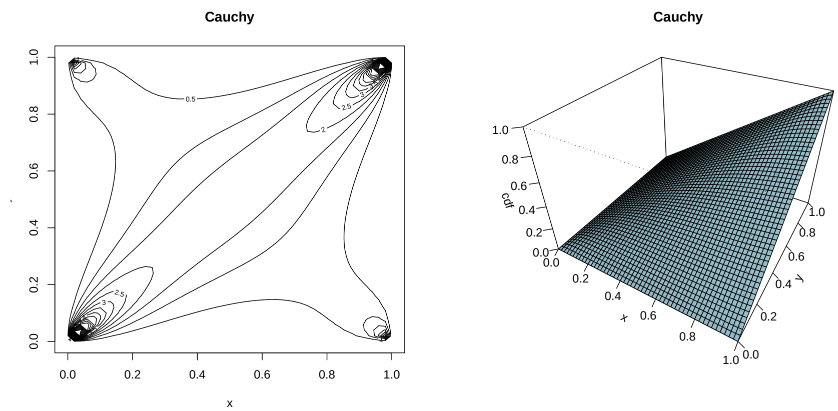


Figure 2.9: *Contour plot and the CDF for the data in figure 2.8 above.*

2.3.3 Archimedean copulas

Archimedean copulas are widely applied, because they are not difficult to construct. In comparison to Elliptical copulas, Archimedean copulas have only one dependency parameter (instead of a dependency matrix) and have many different forms. Unlike elliptical copulas (for instance, Gaussian), most of the Archimedean copulas have closed-form solutions and are not derived from the multivariate distribution functions using Sklar's Theorem. In this sub-section we bring together important definitions and properties of Archimedean copulas, that turn out to be necessary for our approach.

Definition 2.3.1 *Let $\varphi : [0,1] \to [0,\infty]$ be a continuous, strictly decreasing and convex function such that $\varphi(1) = 0$ and $\varphi(0) = \infty$. The function φ has an inverse $\varphi^{-1} : [0,\infty] \to [0,1]$ with the same properties like φ, except that $\varphi^{-1}(0) = 1$ and $\varphi^{-1}(\infty) = 0$.*

Definition 2.3.2 *The function $C : [0,1]^n \to [0,1]$ defined by*

$$C(u_1,\ldots,u_n) = \varphi^{-1}(\varphi(u_1) + \ldots + \varphi(u_n)) \tag{2.3.2}$$

is called n-dimensional Archimedean copula if and only if φ^{-1} is completely monotonic on $[0,\infty)$, that is

$$(-1)^k \frac{\partial^k}{\partial u^k}\varphi^{-1}(u) \geq 0 \text{ for } k \in \aleph$$

The function φ is called the generator of the copula. We assume that the generator φ has only one parameter, denoted as θ.

If $\varphi(0) = \infty$ we say that φ is a strict generator, and the copula is said to be a strict Archimedean copula. Since the most useful copulas are strict we will assume that this condition is satisfied in what follows. Archimedean copulas arise naturally in the context of Laplace transforms.

Let $\phi = \varphi^{-1}$. Thus, $\phi(s)$ is the Laplace transform

$$\phi(s) = \int_0^\infty e^{-sw} dM(w)$$

of some univariate cumulative distribution function M of a positive random variable (that is, $M(0) = 0$), then the function 2.3.2 is ensured to be a proper distribution function. In other words, the necessary condition for function 2.3.2 to be a cumulative distribution function is the complete monotonicity of the inverse generator φ^{-1}. Let us introduce the class of functions

$$L_n = \left\{ \phi : [0, \infty) \to [0,1] \middle| \phi\left(0\right) = 1, \phi\left(\infty\right) = 0, \left(-1\right)^j \phi^{(j)} \geq 0, j = 1, ..., n \right\}$$

$n \in \aleph$, L_∞ being the class of Laplace transforms of strictly positive random variables. With this notation, the necessary and sufficient conditions for function 2.3.2 to be a copula is $\varphi^{-1} \in L_\infty$ and that, if function 2.3.2 is a copula for all $n \in \aleph$, then φ^{-1} must be completely monotone and hence a Laplace transform of a strictly positive random variable.

Archimedean copulas given by function 2.3.2 have some algebraic properties which are very useful for their generalisation and are found in Theorem 2.3.3.

Theorem 2.3.3 *Let C be an Archimedean copula with generator φ. Then*

1. *C is symmetric, that is, $C\left(u_1, u_2\right) = C\left(u_2, u_1\right) \forall u_1, u_2 \in [0,1]$;*
2. *C is associative, that is, $C\left(C\left(u_1, u_2\right), u_3\right) = C\left(u_1, C\left(u_2, u_3\right)\right) \forall u_1, u_2, u_3 \in [0,1]$;*
3. *If $c > 0$ is any constant then $c\varphi$ is also a generator of C.*

It is, however, important to note that a Copula C is exchangeable if it is associative. Archimedean copulas are permutation-symmetric in the n arguments, thus they are distribution functions of n exchangeable uniform random variates. For this reason, these copulas suffer from a very limited dependence structure since all k-dimensional marginal distributions are identical ($k < n$). One particularly simple form of a n-dimensional copula is

$$H(x_1, x_2, ..., x_n) = \varphi^{-1}\left(\sum_{i=1}^{n} \varphi(F_i(x_i))\right)$$

where φ is known as a *generator function.* Such copulas are known as *Archimedean.* Any generator function is a valid copula provided that it satisfies the properties below:

$$\varphi(1) = 0;\ \lim_{x\to 0} \varphi(x) = \infty;\ \varphi^{'}(x) < 0;\ \varphi^{''}(x) > 0.$$

Product copula: Also called the independent copula, is copula with no dependence between variates and defined:

$$\varphi(x) = -\ln(x)$$

Where the generator function is indexed by a parameter, a whole family of copulas may be Archimedean. Its density function is unity everywhere.

If X and Y are independent random variables, then the product of their individual margins F and G equal to their joint distribution function H. That is,

$$H(x, y) = F(x)\, G(y) \text{ for all } x, y \in \Re$$

.

Then, the independence copula of the independent random variables is denoted as, $C = \prod$ and we can write $\prod(u,v) = uv$. For n independent random variables, $C(u_1, u_2, ..., u_n) = \prod(u_1, u_2, ..., u_n) = \prod_{i=1}^{n} u_i$.

The general relationship between Kendall's tau, τ, and the generator of an Archimedean copula $\varphi_\theta(x)$ for a bivariate data set can be written as:

$$\tau = 1 + 4\int_0^1 \frac{\varphi_\theta(t)}{\varphi_\theta(t)} dt$$

For instance, the relationship between Kendall's tau, τ, and the Clayton copula parameter θ for a bivariate data set is given by:

$$\hat{\theta} = \frac{2\tau}{1-\tau}$$

The definition does not extend to a multivariate data set of n variables because there will be multiple values of *tau*, one for each pairing. However, one can calculate *tau* for each pair and use the average, that is:

$$\hat{\theta} = \frac{2\bar{\tau}}{1-\bar{\tau}}, \quad \bar{\tau} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n, i\neq j} \tau_{ij}}{n(n-1)}$$

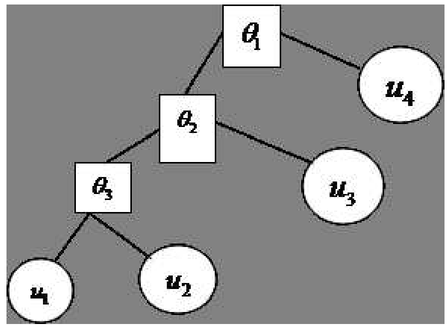


Figure 2.10: *Fully nested copula with a dimension of $n = 4$*

The three often used Archimedean copulas include: Gumbel, Clayton and Frank. Multivariate copulas, on the other hand, allow for non-exchangeability. Since any multivariate generalisation should contain function 2.3.2 as a special case, clearly the necessary conditions for function 2.3.2 have to be satisfied. A simple generalisation of the multivariate Archimedean copula can be found in Embrechts, Lindskog, and McNeil (2003) and Whelan (2004). The copula for the n-dimensional case requires $n-1$ generators $\varphi_1, ..., \varphi_{n-1}$,

$$\begin{aligned} C\left(u_1, ..., u_n\right) &= \varphi_{n-1}^{-1}(\varphi_{n-1} \circ \varphi_{n-2}^{-1}[...(\varphi_2 \circ \varphi_1^{-1}[\varphi_1\left(u_1\right) + \varphi_1\left(u_2\right) + \varphi_2\left(u_3\right)) \\ &+ ...+\varphi_{n-2}\left(u_{n-2}\right)] + \varphi_{n-1}\left(u_n\right)) \qquad (2.3.3) \end{aligned}$$

This is referred to as fully nested, since a higher dimensional copula is obtained by adding one dimension step by step. The resulting dependence structure is more general than in function 2.3.2, it is one of partial exchangeability.

There are $n(n-1)/2$ distinct bivariate margins and (only) $n-1$ several copulas with corresponding parameters. The expression given by expression 2.3.3 will be a proper n-copula if, in addition to the property of complete monotonicity for the inverse generators, other conditions concerning the composite functions $\omega = \varphi_{i+1} \circ \varphi_i^{-1}$ are satisfied. Define

$$L_n^* = \left\{ \omega : [0, \infty) \to [0, \infty) \middle| \omega(0) = 0, \omega(\infty) = \infty, (-1)^{j-1} \omega^{(j)} \geq 0, j = 1, ..., n \right\}$$

with $n \in \aleph$. The functions in L_∞^* are compositions of the form $\varphi_{i+1} \circ \varphi_i^{-1}$ with $\varphi_{i+1}^{-1}, \varphi_i^{-1} \in L_1$.

The second generalisation method of multivariate Archimedean copulas is more flexible than the fully nested one. It is a mixture of exchangeable and fully nested copulas and is referred to as partially nested. The lowest dimension in which a copula from this class exists is $n = 4$, and the copula is

$$C(u_1, ..., u_4) = \varphi^{-1} \left(\varphi \circ \varphi_{12}^{-1} [\varphi_{12}(u_1) + \varphi_{12}(u_2)] + \varphi \circ \varphi_{34}^{-1} [\varphi_{34}(u_3) + \varphi_{34}(u_4)] \right) \tag{2.3.4}$$

with three generators φ, φ_{12} and φ_{34} . This copula is thus generated by three different generating functions. The random vector $(U_1, ..., U_4)$ with distribution function 2.3.4 is only partially exchangeable. The random variates U_1 and U_2 are exchangeable, and so are U_3 and U_4, but the remaining pairs are not. However, the joint distribution of (U_1, U_3) is equal to the joint distribution of (U_2, U_3), (U_1, U_4) and (U_2, U_4). Higher dimension nesting can be found in Whelan (2004).

2.3.3.1 The Clayton copula

The Clayton copula is an asymmetric Archimedean copula, Nelsen (2006), exhibiting greater dependence in the negative tail than in the positive. This copula is given by:

$$C_\theta(u, v) = \max\left(\left[u^{-\theta} + v^{-\theta} - 1\right]^{-1/\theta}, 0\right)$$

and its generator is:

$$\varphi_\theta(x) = \frac{1}{\theta}\left(x^{-\theta} - 1\right)$$

where, $\theta \in [-1, \infty) \setminus \{0\}$. The relationship between Kendall's tau, τ, and the Clayton copula parameter θ is given by:

$$\hat{\theta} = \frac{2\tau}{1 - \tau}$$

For $\theta = 0$ in the Clayton copula, the random variables are statistically independent. The generator function approach can be extended to create multivariate copulas, by simply including more additive terms.

2.3.3.2 The Gumbel copula

The Gumbel copula (also referred to as the Gumbel-Hougard copula) is an asymmetric Archimedean copula, Nelsen (2006), exhibiting greater dependence in the positive tail than in the negative. This copula is given by:

$$C_\theta(u,v) = \exp\left\{-\left[(-\ln u)^\theta + (-\ln v)^\theta\right]^{1/\theta}\right\}$$

and its generator is:

$$\varphi_\theta(x) = (-\ln x)^\theta$$

where, $\theta \in [1, \infty)$. The relationship between Kendall's tau, τ, and the Gumbel copula parameter θ is given by:

$$\hat{\theta} = \frac{1}{1-\tau}$$

2.3.3.3 The Frank copula

The Frank copula is a symmetric Archimedean copula, Nelsen (2006), given by:

$$C_\theta\,(u,v) = -\frac{1}{\theta}\ln\left\{1 + \frac{\left(e^{-\theta u}-1\right)\left(e^{-\theta v}-1\right)}{e^{-\theta}-1}\right\}$$

and its generator is:

$$\varphi_\theta(x) = -\ln\left\{\frac{\exp(-\theta x)-1}{\exp(-\theta)-1}\right\}$$

where, $\theta \in (-\infty, \infty) \setminus \{0\}$. The relationship between Kendall's tau, τ, and the Frank copula parameter θ is given by:

$$\frac{[D_1\,(\theta)-1]}{\theta} = \frac{1-\tau}{4}$$

where, $D_1(\theta) = \frac{1}{\theta}\int\limits_0^\theta \frac{t}{e^t-1}dt$, is a Debye function of the first kind. A Debye function is defined as:

$$D_n(x) = \frac{n}{x^n}\int\limits_0^x \frac{t^n}{e^t-1}dt$$

where n , a non-negative integer, is the order of the Debye function.

2.3.4 Graphics for the Archimedean copulas

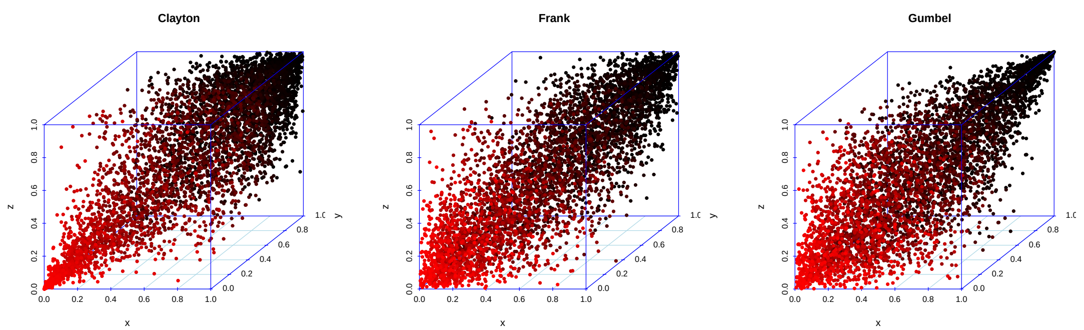


Figure 2.11: *Scatter plots of 5000 random variables for the Clayton, Frank and the Gumbel copulas.*

Scenarios were simulated with standard normal marginals and a Kendall's tau of 0.5. Notice the differences in the dependence structure in the Figure 2.11. The Clayton has lower tail dependence, Frank that resembles a 'sausage' has no tail dependence while the Gumbel captures the upper tail dependence. Similar observation, but from different perspectives, can be seen in Figure 2.12 and Figure 2.13.

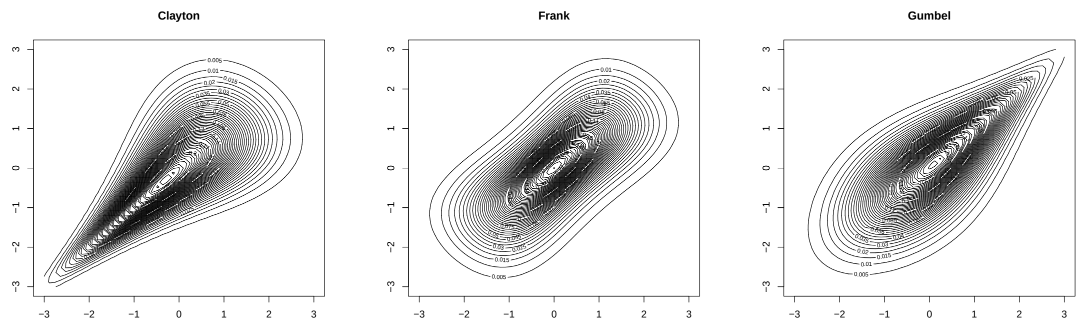


Figure 2.12: *Contour plots for the data found in figure 2.11 to enables us visualise the dependence structure on two-dimension.*

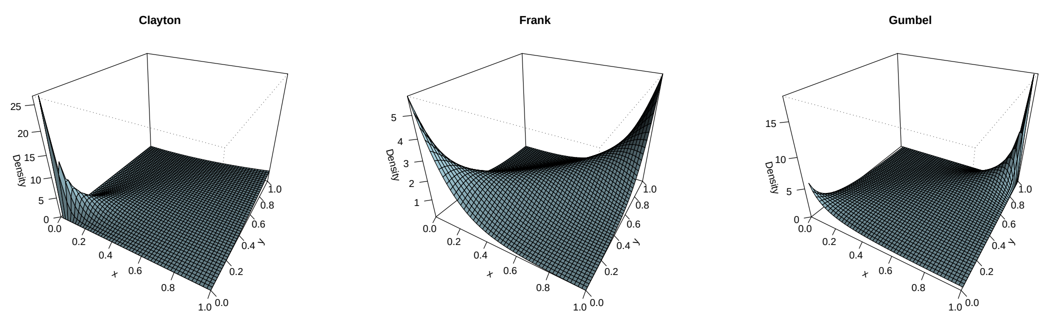


Figure 2.13: *Perspective plots of our data in figure 2.11. Notice the strengths of each of the three Archimedean copulas.*

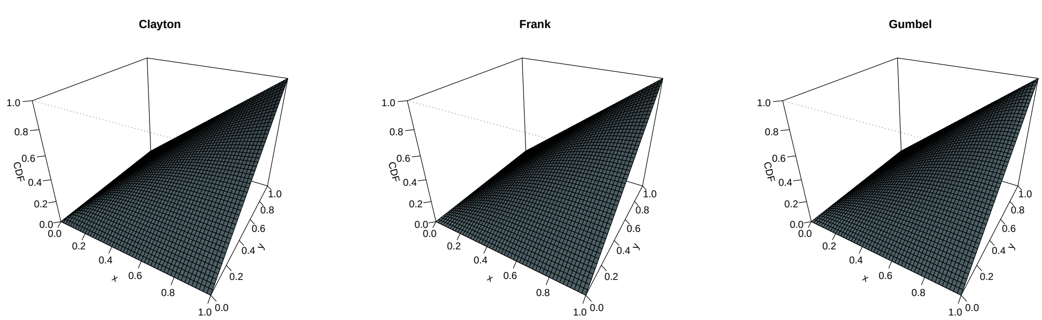


Figure 2.14: *The CDFs of our data in figure 2.11 above.*

2.3.5 Periodic copula

Alfonsi and Brigo (2005) introduced new families of copulas based on periodic functions. They noticed that if $\tilde{c}$ is a 1-periodic non-negative function that integrates to 1 over $[0, 1]$ that is,

$$\int_0^1 \tilde{c}(u)du = 1$$

and ϕ is a double primitive of $\tilde{c}$, then both

$$\tilde{C}^-(u_1, u_2) = \int_0^{u_1}\int_0^{u_2} \tilde{c}(x_1 + x_2)\, dx_1 dx_2 = \phi(u_1 + u_2) - \phi(u_1) - \phi(u_2)$$

$$\tilde{C}^+(u_1, u_2) = \int_0^{u_1}\int_0^{u_2} \tilde{c}(x_1 - x_2)\, dx_1 dx_2 = \phi(u_1) + \phi(-u_2) - \phi(u_1 - u_2)$$

are copula functions, with the second one not necessarily exchangeable. This may be a tool to introduce asymmetric dependence, which is absent in most known copula functions.

2.3.6 Empirical copulas

When analysing data with an unknown underlying distribution, one can transform the empirical data distribution into an empirical copula, Nelsen (2006), by warping such that the marginal distributions become uniform. Mathematically the empirical copula frequency function is calculated by

$$C_n\left(\frac{i}{n},\frac{j}{n}\right)=\frac{\text{Number of pairs } (x,y) \text{ such that } x\leq x_{(i)} \text{ and } y\leq y_{(j)}}{n}$$

$$\text{where } 1\leq i\leq n,\ 1\leq j\leq n$$

where $x_(i)$ represents the ith order statistic of x. Less formally, simply replace the data along each dimension with the data ranks divided by n.

Classification of risks generally forms the basis of rate making in practically all branches of insurance. Once a risk is insured, it is reasonable that the standards for classifying that risk can and should be different from those of marketing or underwriting. Furthermore, once the classifications are established, there are also guidelines to follow in establishing the prices, or classification differentials, for the system. We focus on the appropriate rules regarding selection of classification variables and the definition of classes at the very start of the classification rating process. Given the preceding, the variables comprising a classification system should be chosen so that the following guidelines or conditions in addition, of course, to any legal requirements regarding fair discrimination, are generally adhered to according to Walters (1981) who writes on risk classification standards, viz:

1. Similar risks should be assigned to the same class with respect to each variable. Conversely, dissimilar risks should be assigned to different classes, so that there

are no clearly identifiable subsets with a significantly different loss potential or expected loss in the same class.

2. The common characteristics used to identify insureds as similar should reasonably relate to the potential for, or hazard of, loss.

3. The classes should be exhaustive and mutually exclusive; that is, each insured should belong to at least one, but only one class with respect to each rating variable.

4. There should be clear and objective phraseology in the definition of classes, with no ambiguity as to what class an individual insured belongs.

5. An insured should not be easily able to misrepresent or manipulate his classification.

6. The cost of administering a rating variable should be reasonable in relation to the benefits received.

7. The class rating factors should be susceptible to measurement by actual experience data.

The first guideline is what is meant by homogeneous classes. Classes that are homogeneous will take fewer risks to obtain reasonable estimates of expected costs, and will minimize the ability of the competition to skim off better than average risks, thus changing the ultimate costs. Secondly or "the reasonable relationship" guideline serves to maintain homogeneous classes by avoiding spurious measures which likely have potentially identifiable subsets. Of course, if a strong statistical correlation persists over time, with no emergence of practical subdivisions then the degree of perceived reasonableness may be enhanced over time as well.

The third, fourth, and fifth guidelines deal with classes being well-defined, and help to ensure that each risk is actually placed in the right classification and to avoid unequal application of the classification system. The "exhaustive" quality allows more risks to be accepted and, once accepted, gives a complete method of rating them. "Exclusivity" precludes two different rates for the exact same risk. "No ambiguity" also prevents unequal treatment of the same risk, while protection from misrepresentation by insureds will keep the statistical data consistent as well as enhancing the equal treatment of insureds. The last two guidelines touch on efficiency and effectiveness of the classification process.

However, there are some examples of classifications which do not meet the guidelines above and they may include the following

1. The use of occupation as a rating variable for auto liability insurance may result in a problem with regard to meeting the ambiguity guideline, both in splitting the population into exhaustive categories, as well as not having all cells likely

being reasonably related to the hazard of loss.

2. Similarly, national origin, if not already proscribed by law, would have problems meeting the mutually exclusive and exhaustive guidelines.

3. Using unverifiable criteria or too subjective wording, such as with psychological profiles, would also present major problems. The use of characteristics which are easily circumvented by some insureds and not others can favour the pricing of some to the detriment of others.

In addition, a class plan would not be homogeneous if it failed to reflect premium differences for identifiable and rateably different subsets within broader classifications. The degree of failure would depend upon the cost of determining the necessary information. From the insured's standpoint, the pricing impact of not subdividing depends upon the size of the subsets and the resulting differences in price for each of the subclasses. It may be that only a small amount of premium can be saved by refinement, if one of the subclasses is very large and also the lowest priced, such as rating by past accident record in auto insurance where accident-free or claim-free drivers usually save at most five percent over the cost of not having such a program. After attaining the classes that an industry intends to hold, a portfolio of diverse risks is achieved. Diversification of risks is a risk management technique that mixes a wide variety of risks within a portfolio. The rationale behind this technique contends that a portfolio of different kinds of risks will, on average, yield higher returns and pose a lower risk than any individual investment found within the portfolio. Diversification strives to smooth out unsystematic risk events in a portfolio so that the positive performance of some investments will neutralize the negative performance of others. Therefore, the benefits of diversification will hold only if the risks in the portfolio are not perfectly correlated.

This technique endeavours to maximize return by venturing in different areas that would each react differently to the same event. Most risk managers agree that, although it does not guarantee against loss, diversification is the most important component of meeting high financial goals while minimizing risk. We now propose a criterion to restructure a company's classes of risks using their dependence structure. Burgi, Dacorogna, and Iles (2008) noted that modern portfolio theory is based on correlation as a measure of dependence but the criterion below is based on the copula theory which is superior to the correlation as a measure of the intrinsic relatedness of different risks. Dependence between risks reduces the benefits of diversification. Often dependence increases when diversification is most needed like in case of stress and it is thus non-linear.

Similarity and dissimilarity: Distances are used to measure similarity and dissimilarity. Similarity is a quantity that reflects the strength of relationship between two objects. This quantity is usually having range of either -1 to +1 or normalized into 0 to 1. If the similarity between feature i and feature j is denoted by S_{ij} , we can measure this quantity in several ways depending on the scale of measurement, or data type, that we have.

Dissimilarity is a measure of disorder between two objects. It may also be viewed as a measure of the discrepancy between two objects based on several features. Let the normalized dissimilarity between object i and object j be denoted by δ_{ij}. The relationship between dissimilarity and similarity is given by

$$S_{ij} = 1 - \delta_{ij}$$

for similarity bounded by 0 and 1. When similarity is one, that is, exactly similar, the dissimilarity is zero and when the similarity is zero, that is, very different, the dissimilarity is one. If the value of similarity has range of -1 to +1, and the dissimilarity is measured with range of 0 and 1, then

$$S_{ij} = 1 - 2\delta_{ij}$$

When dissimilarity is one, that is, very different, the similarity is minus one and when the dissimilarity is zero, that is, very similar, the similarity is one. In many cases, measuring dissimilarity, that is, distance, is easier than measuring similarity. Once we can measure the dissimilarity, we may easily normalize it and convert it to similarity measure.

There are many types of distance and similarity. Each similarity or dissimilarity has its own characteristics. For instance, we have distances for Binary variables, Nominal/Categorical Variables, Ordinal Variables, and Quantitative Variables. The most commonly used of binary dissimilarity (distance) are Simple Matching distance, Jaccard's distance and the Hamming distance. For the categorical variables, if the number of category is only two, then we can use distance for binary variables such as simple matching, Jaccard's or Hamming distance. If the number of category is more than two, we need to transform these categories into a set of dummy variables that has binary value. To compute dissimilarity or distance between two Ordinal Variables, that is, ranked or two orded or two rating vectors, the most common methods are: Normalized Rank Transformation, Spearman Distance, Footrule Distance, Kendall Distance, Cayley Distance, Hamming Distance, Ulam Distance, Chebyshev or Maximum Distance and the Minkowski Distance. Distance for ordinal variables is a measure of spatial disorder between two rank or ordering vectors. Since our dependence parameters are quantitative variables, we shall have a close look at some of the common used distances for the quantitative variables below.Let us have a brief look at some of the distances that are mainly used.

2.4 Distances

Distance is a numerical description of how far apart objects are. In everyday discussion, distance may refer to a physical length, a period of time, or estimation based on other criteria, for instance, two counties over. In mathematics and statistics, distance must meet more rigorous criteria. In most cases there is symmetry and "distance from X to Y" is interchangeable with "distance between Y and X". A distance or metric function is a function which defines a distance between elements of a set. A set with a metric is called a metric space.

Definition 2.4.1 *(Distance function) A metric on a set X is a function, called the distance function or simply distance, $d : X \times X \to \Re$ (where $\Re$ is the set of real numbers). For all $x, y, z \in X$, this function is required to satisfy the following conditions:*

1. $d(x, y) \geq 0$ *(non-negativity)*

2. $d(x, y) = 0$ *if and only if* $x = y$ *(identity of indiscernibles. Note that condition 1 and 2 together produce* positive definiteness*)*

3. $d(x, y) = d(y, x)$ *(symmetry)*

4. $d(x, z) \leq d(x, y) + d(y, z)$ *(subadditivity / triangle inequality).*

These conditions express intuitive notions about the concept of distance. For example, from 1 the distance between distinct points is always positive or zero. Property 2 implies that distance is zero if and only if it is measured to itself and the distance from x to y is the same as the distance from y to x, that is, distance is symmetric. The triangle inequality means that the distance traversed directly between x and z, is not larger than the distance to traverse in going first from x to y, and then from y to z. Property 1 ($d(x, y) \geq 0$) follows from properties 2 and 4 and does not have to be required separately.

2.4.1 The Mahalanobis distance

This distance measure was introduced by P. C. Mahalanobis in 1936. It is based on correlations between variables by which different patterns can be identified and analysed. It is a useful way of determining similarity of an unknown sample set to a known one. It differs from Euclidean distance in that it takes into account the correlations of the data set and is scale-invariant, i.e. not dependent on the scale of measurements.

Formally, the Mahalanobis distance from a group of values with mean $\mu = (\mu_1, \mu_2, \mu_3 ..., \mu_p)^T$ and covariance matrix Σ for a multivariate vector $\mathbf{x} = (x_1, x_2, x_3, ..., x_p)^T$ is defined as

$$D_M(x) = \sqrt{(x-\mu)^T \Sigma^{-1} (x-\mu)}$$

Mahalanobis distance can also be defined as dissimilarity measure between two random vectors $\vec{x}$ and $\vec{y}$ of the same distribution with the covariance matrix P:

$$d(\vec{x}, \vec{y}) = \sqrt{(\vec{x}-\vec{y})^T P^{-1} (\vec{x}-\vec{y})}$$

If the covariance matrix is the identity matrix, then the Mahalanobis distance reduces to the Euclidean distance. If the covariance matrix is diagonal, then the resulting distance measure is called the *normalized Euclidean distance*:

$$d(\vec{x}, \vec{y}) = \sqrt{\sum_{i=1}^{p} \frac{(x_i - y_i)^2}{\sigma_i^2}}$$

where σ_i is the standard deviation of the x_i over the sample set.

2.4.2 Euclidean distance

The Euclidean distance or Euclidean metric is the "ordinary" distance between two points that one would measure with a ruler, which can be proven by repeated application of the Pythagorean Theorem. In most cases when people talk about distance, they will refer to Euclidean distance. The most well-known distance is the Euclidean distance which is defined as:

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x}, \mathbf{y}\| = \sqrt{(x-y)^T(x-y)} = \sqrt{\sum_i (x_i - y_i)^2}$$

(with $\|\mathbf{x}\|$ being the norm of $\mathbf{x}$, and x_i and y_i being the i-th element of $\mathbf{x}$ and $\mathbf{y}$). Expressed as a squared distance (in an Euclidean world, it is always more practical to work with squared quantities because of the Pythagorean theorem), it is computed as:

$$d^2(\mathbf{x}, \mathbf{y}) = (x-y)^T(x-y)$$

2.4.3 Manhattan distance

The Manhattan distance function computes the distance that would be travelled to get from one data point to the other if a grid-like path is followed, that is, the distance between two points measured along axes at right angles. The Manhattan distance between two items is the sum of the differences of their corresponding components. It is also known by various names as City Block Distance, boxcar distance, absolute value distance, rectilinear distance, Minkowski's L_1 distance, or taxi cab metric and it is given by:

$$d_i = \sum_i^n |x_i - y_i|$$

where n is the number of variables, and x_i and y_i are the values of the i-th variable, at points X and Y respectively. Other distances include: the Canberra Distance, Bray

Curtis Distance and Angular Separation which represents cosine angle between two vectors. We also have the Correlation coefficient, for the quantitative variables, which is standardized angular separation by centering the coordinates to its mean value. Correlation coefficient measures similarity rather than distance or dissimilarity.

2.5 Clustering

Clustering is a technique to group objects based on distance or similarity. It is therefore the assignment, grouping or segmenting of a set of observations, individuals, cases, or data rows into subsets, called clusters, so that observations in the same cluster are similar in some sense. The cardinal objective of clustering is the notion of degree of similarity (or dissimilarity) between the individual objects being clustered.

The types of clustering algorithm include, first the Hierarchical algorithms that find successive clusters using previously established clusters. They are usually either agglomerative, the "bottom-up", or divisive the "top-down". Agglomerative algorithms begin with each element as a separate cluster and merge them into successively larger clusters. Divisive algorithms begin with the whole set and proceed to divide it into successively smaller clusters. The second lines of algorithms are the Partitional algorithms which typically determine all clusters at once, but can also be used as divisive algorithms in the hierarchical clustering. The Density-based clustering algorithms are devised to discover arbitrary-shaped clusters. In this approach, a cluster is regarded as a region in which the density of data objects exceeds a threshold. Finally, the Subspace clustering methods look for clusters that can only be seen in a particular projection, subspace or manifold, of the data.

In this work we utilize the agglomerative approach under the Hierarchical clustering. The algorithm of agglomerative approach to compute hierarchical clustering is as follows:

1. Convert object features to distance matrix, in our case we have the matrix of the rank correlation coefficients and the tail dependence.

2. Set each object as a cluster, thus for the sixteen objects, we will have sixteen clusters in the beginning.

3. Iterate until number of cluster is one, that is, by merging the two closest clusters and continuously updating the distance matrix.

2.5.1 Linkages between objects

Given a distance matrix, linkages between objects can be computed through a criterion to compute distance between groups. The most common and basic criteria are: Single Linkage (minimum distance criterion), Complete Linkage (maximum distance criterion), Average Group (average distance criterion), Centroid distance criterion and the Ward (which is to minimize variance of the merge cluster).

2.5.2 Cophenetic correlation coefficient

After the formation of the clusters, the question now is how good is the clustering? There is an index called Cross Correlation Coefficient or Cophenetic Correlation Coefficient that shows the goodness-of-fit of our clustering similar to the Correlation Coefficient of regression. To compute the Cophenetic Correlation Coefficient of hierarchical clustering, we need a distance matrix and a Cophenetic matrix. To obtain Cophenetic matrix, we need to fill the distance matrix with the minimum merging distance that we obtain in the previous cluster objects. Cophenetic Correlation Coefficient is simply correlation coefficient between distance matrix and Cophenetic matrix.

The dependence structure of different lines of business cannot be ignored especially in Economic Capital assessment and overly too in Enterprise Risk Management (ERM). It is in this light that Faivre (2003) used Copulas to model the overall distribution of claim costs of a four-Lines of Business company. The work utilized different copulas to show that the dependence structure has a substantial impact on the Economic Capital of that firm. This study extends the work by Faivre (2003) and makes further realistic assumptions. The study by Wu, Valdez, and Sherris (2007) that used Value-at-Risk and the Conditional Tail Expectation (CTE) is extended here to take the proposed risk framework.

Suppose that there exists an ideal situation where all risks are the same. This would imply that the portfolio is completely homogeneous. However, in real world you would find that risks are more heterogeneous than homogeneous a problem that can be solved by breaking down the risks into a number of homogeneous categories, Straub (1997). We can now think of the lines of business to contain sub-classes which are homogeneous. The lines will depict a hierarchical structure from the sub-classes to the main lines of business and their dependence structure will be studied by the hierarchical copulas. The following chapter dedicates its scope to measuring the dependence in data that has an hierarchical structure.

Chapter 3

MODELLING THE DEPENDENCE BETWEEN M-LINES OF BUSINESS WITH SUBCLASSES

3.1 Introduction

Traditionally, insurance companies maintain different lines of business, in their operations, as a mode of diversification which in itself aids in reducing risks of encountering ruin. These could be the portfolio profiles for the companies. For instance, there are companies operating various lines in the general insurance, others in life insurance while some have composite arrangements. In this chapter, we consider a case of a company with M-lines of business and each of this line contains some sub-classes. Therefore,the portfolio will have an hierarchical structure, that is, from the sub-classes to the main lines of business and finally to the portfolio level.

3.2 Modelling the aggregate loss

For the portfolio of business classes, let N_i denote the number of claims during a fixed time period for policies in class i and $X_{i1}, X_{i2}, ...$ denote the amounts of successive claims in that business class. The aggregate loss of each class in the portfolio during the period can be expressed as

$$S_i = \sum_{j=1}^{N_i} X_{ij} \tag{3.2.1}$$

where N_i is a random variable. This is a compound risk model with the distribution function of S_i in equation 3.2.1 denoted by F_{S_i}. In the compound risk model, we assume that $X_i, X_{i1}, X_{i2}, ...$ are independent identically distributed with common severity distribution function (sdf), denoted as F_{X_i} The sequence $X_i, X_{i1}, X_{i2}, ...$ are also independent of the number of claims N_i.

The expected value of S_i can be obtained by using the conditional distribution of S_i, given N_i. First, we use the condition $N_i = n_i$ to substitute outcome n_i for the random variable N_i. We then use the independence of X_{ij} and N_i to get rid of the condition $N_i = n_i$. This gives the following computation: Let $\mu_{ik} = E\left[X_i^k\right]$

Now,

$$\begin{aligned} E[S_i] &= E[E[S_i \,|N_i]] \\ &= \sum_{n_i=0}^{\infty} E[X_{i1} + ... + X_{iN_i} \,|N_i = n_i] \Pr[N_i = n_i] \\ &= \sum_{n_i=0}^{\infty} E[X_{i1} + ... + X_{in_i}] \Pr[N_i = n_i] \\ &= \sum_{n_i=0}^{\infty} n_i \mu_{i1} \Pr[N_i = n_i] = \mu_{i1} E[N_i] \end{aligned} \tag{3.2.2}$$

The variance can be determined with the formula of the conditional variance, see Bowers, Gerber, Jones, and Nesbitt (1997),

$$
\begin{aligned}
Var\left[S_i\right] &= E\left[Var\left[S_i\left|N_i\right.\right]\right] + Var\left[E\left[S_i\left|N_i\right.\right]\right] \\
&= E\left[N_i Var\left[X_i\right]\right] + Var\left[N_i \mu_{i1}\right] \\
&= E\left[N_i\right] Var\left[X_i\right] + \mu_{i1}^2 Var\left[N_i\right]
\end{aligned}
\tag{3.2.3}
$$

The moment generating function for the compound distribution can be obtained as follows:

$$
\begin{aligned}
M_{S_i}\left(t\right) &= E\left[E\left[e^{tS_i}\left|N_i\right.\right]\right] \\
&= \sum_{n_i=0}^{\infty} E\left[e^{t\left(X_{i1}+\ldots+X_{iN_i}\right)}\left|N_i = n_i\right.\right] \Pr\left[N_i = n_i\right] \\
&= \sum_{n_i=0}^{\infty} E\left[e^{t\left(X_{i1}+\ldots+X_{in_i}\right)}\right] \Pr\left[N_i = n_i\right] \\
&= \sum_{n_i=0}^{\infty} \left\{M_{X_i}\left(t\right)\right\}^{n_i} \Pr\left[N_i = n_i\right] \\
&= E\left[\left(e^{\log M_{X_i}(t)}\right)^{N_i}\right] = M_{N_i}\left(\log M_{X_i}\left(t\right)\right)
\end{aligned}
\tag{3.2.4}
$$

The aggregate loss for the portfolio can now be obtained from:

$$\begin{aligned} S &= S_1 + S_2 + ... + S_N \\ &= \sum_{i=1}^{N} S_i \end{aligned}$$

where N denotes the total number of claims and S_i is the claims amount emanating from the i-th business class, and S is taken to be zero if $N = 0$.

3.3 Hierarchical Copula

The Archimedean copulas are a good choice for modelling bivariate distributions. However, for high dimensional case, they tend to be too restrictive as they imply exactly the same dependence structure between all pairs of variables. To make Archimedean copulas less restrictive in describing dependence structures of n-dimensional distributions, we can resort to a hierarchical structure, Kang (2007).

Consider, for a simple case, a company with only two LOB with each having two homogenous sub-classes. This gives rise to four sub-classes in the company. We can divide the four sub-classes into two pairs: the first pair consisting of one LOB and then the second includes the other LOB. Ideally, we can find a proper bivariate copula to model each pair of LOB respectively and then nest these two copulas into another bivariate copula to form a joint distribution of the four sub-classes, see Figure 3.1 for illustration.

Let $C_1\left(u_1, u_2\right)$ and $C_2\left(u_3, u_4\right)$ be the two copulas governing the two pairs of LOB and then the final joint distribution can be given as

$$F\left(x_1, x_2, x_3, x_4\right) = C_3\left(C_1\left(u_1, u_2\right), C_2\left(u_3, u_4\right)\right) \tag{3.3.1}$$

where $u_i = F_i(x_i)$ and its density form is

$$f(x_1, x_2, x_3, x_4) = c_3(C_1(u_1, u_2), C_2(u_3, u_4))\, c_1(u_1, u_2)\, c_2(u_3, u_4) \prod_{i=1}^{4} f_i(x_i) \quad (3.3.2)$$

However, equation 3.3.1 does not always hold for any copulas and for certain choices of C_1 and C_2, C_3 will not satisfy the definition of copulas, Nelsen (2006). Nevertheless, Archimedean copulas have certain properties that facilitate constructing hierarchical copulas as shown in equation 3.3.1.

3.4 Hierarchical Archimedean Copulas (HAC)

The idea of hierarchical Archimedean Copulas has been mentioned in the literature some of which include Embrechts et al. (2003), Whelan (2004) and later in Savu and Trede (2006) which is to build a hierarchy of Archimedean copulas. The model relies on a notational framework based on nested multivariate and generalised Archimedean copulas. At each level we aggregate Archimedean copulas from the previous level, finally ending at the top level with a hierarchical Archimedean copula, being the joint distribution function of n standard uniformly distributed random variables $U_1, U_2, ..., U_n$. The joint distribution function is evaluated at $\mathbf{u} = (u_1, u_2, ..., u_n) \in [0, 1]^n$. Let there be L hierarchy levels indexed by l. At each level $l = 1, ..., L$ we have m_l distinct objects with index $j = 1, ..., m_l$. The $u_1, u_2, ..., u_n$ are located at the lowest level, $l = 0$. At level $l = 1$ the $u_1, u_2, ..., u_n$ are grouped into

m_1 ordinary multivariate Archimedean copulas $C_{1,j}, j = 1, ..., m_1$, of the form

$$C_{1,j}\left(\mathbf{u}_{1,j}\right) = \varphi_{1,j}^{-1}\left(\sum_{u_{1,j}} \varphi_{1,j}\left(\mathbf{u}_{1,j}\right)\right)$$

$\varphi_{2,j}$ denotes the generator of copula $C_{2,j}$ and $\mathbf{C}_{2,j}$ represents the set of all copulas from level $l = 1$ entering copula $C_{2,j}$ for $j = 1, ..., m_2$. We can proceed in this manner until attaining level L with the hierarchical Archimedean copula $C_{L,1}$ as single object.

The following notation will be useful in representing the hierarchical Archimedean copula $C_{l,j}$ at level $l = 1, ..., L$. We let $C_{l,j}$ have either the argument $\mathbf{u}_{l,j}$ denoting the set of all $u_1, u_2, ..., u_n$ entering (directly or indirectly) the copula $C_{l,j}$, or the argument $\mathbf{C}_{l,j}$ denoting the set of all copulas from level $l-1$ entering $C_{l,j}$ at level $l = 2, ..., L$. Therefore, $C_{2,j}\left(\mathbf{C}_{2,j}\right)$ and $C_{2,j}\left(\mathbf{u}_{2,j}\right)$ are just two ways of writing the same thing.

The conditions to ensure that the resulting structure is in fact a hierarchy include: The number of copulas must decrease at each level, that is $m_l < m_{l-1}\ \forall\ l = 2, ..., L$. The top level contains a single object $(C_{L,1})$, hence $m_L = 1$. Let $n_{l,j}$ denote the dimension of copula $C_{l,j}$, which we define as the cardinality of $\mathbf{u}_{l,j}$ (rather than the cardinality of $\mathbf{C}_{l,j}$). It must hold $\sum_{j=1}^{m_l} n_{l,j} = n\ \forall\ l = 1, ..., L$, that is, at each level the dimensions of the copulas need to add up to the dimension n of the hierarchical copula. At the top level L we must arrive at dimension $n_{L,1} = n$.

For the hierarchical Archimedean copula $C_{L,1}$ to be a proper cumulative distribution function the following conditions have to be satisfied. To start with, all inverse generator functions $\varphi_{l,j}^{-1}$ should be completely monotone. And secondly, the composite functions $\varphi_{l+1,i} \circ \varphi_{l,j}^{-1} \in L_\infty^* \; \forall \; l = 1, ..., L$ and $j = 1, ..., m_l, \; i = 1, ..., m_{l+1}$ such that $C_{l,j} \in \mathbf{C}_{l+1,i}$.

In the case of fully nested copulas, the degree of dependence, as expressed by the copula parameter θ, has to be greatest for the most deeply nested copulas in order to satisfy the conditions for a proper n-dimensional distribution. This condition has been shown for the Gumbel and the Cook-Johnson copula families in Embrechts et al. (2003). Hence, for the fully nested n-copula in equation 2.3.3 the condition $\theta_1 \geq \theta_2 \geq \ldots \geq \theta_{n-1}$ has to be satisfied. Transferring this result to our setting, the parameters have to satisfy the following condition to ensure that the resulting hierarchical copula is a proper n-dimensional copula: $\theta_{l+1,i} < \theta_{l,j} \forall l = 1, ..., L$ and $j = 1, ..., m_l$, $i = 1, ..., m_{l+1}$ such that $C_{l,j} \in C_{l+1,i}$ where $\theta_{l,j}$ is the parameter belonging to the generator $\varphi_{l,j}$. This condition means that there is a higher degree of dependence for variates linked at a lower level than between those linked only at a higher level; the dependence diminishes with increasing level.

To illustrate let us consider our example of two LOB each having two sub-classes

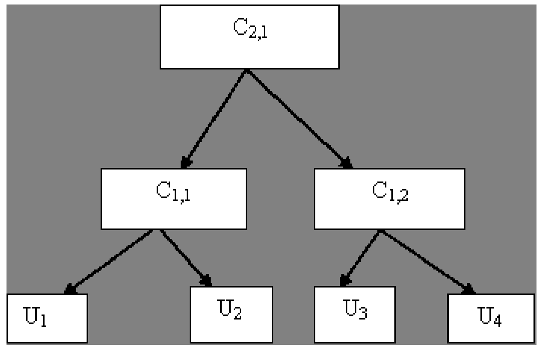


Figure 3.1: *Partially nested copula of dimension $n = 4$*

giving rise to four random variates. For simplicity, the four random variates are uniformly distributed, though this may not hold for more complex cases. Our hierarchical copula is the joint distribution function of four standard uniform random variates $U_1, ..., U_4$ spanning two levels. At the lower level we couple two pairs of random variables U_1, U_2 and U_3, U_4 with distinct copulas $C_{1,1}$ and $C_{1,2}$ generated by $\varphi_{1,1}$ and $\varphi_{1,2}$ respectively. These two copula functions will then be coupled at the upper level using a third generator $\varphi_{2,1}$, as shown in Figure 3.1.

The resulting hierarchical Archimedean copula has the following analytical form

$$\begin{aligned} C_{2,1}(u) &= C_{2,1}(u_1, u_2, u_3, u_4) \\ &= C_{2,1}(C_{1,1}(u_1, u_2), C_{1,2}(u_3, u_4)) \\ &= \varphi_{2,1}^{-1}\left(\varphi_{2,1} \circ \varphi_{1,1}^{-1}[\varphi_{1,1}(u_1) + \varphi_{1,1}(u_2)] + \varphi_{2,1} \circ \varphi_{1,2}^{-1}[\varphi_{1,2}(u_3) + \varphi_{1,2}(u_4)]\right) \end{aligned} \tag{3.4.1}$$

We find here that there are two composite functions $\varphi_{2,1} \circ \varphi_{1,1}^{-1}$ and $\varphi_{2,1} \circ \varphi_{1,2}^{-1}$ that enter the model. If $\varphi_{1,1}, \varphi_{1,2}$ and $\varphi_{2,1}$ are completely monotone and both composite functions $\varphi_{2,1} \circ \varphi_{1,1}^{-1}$ and $\varphi_{2,1} \circ \varphi_{1,2}^{-1}$ are elements of L_{∞}^{*}, then $C_{2,1}(\mathbf{u})$ is a hierarchical Archimedean copula. These conditions are fulfilled if $\theta_{2,1} < \theta_{1,1}$ and $\theta_{2,1} < \theta_{1,2}$.

3.4.1 Deriving the density

For multivariate Archimedean copulas of the form in equation 2.3.2 the copula density can be expressed in terms of the generator functions, as

$$c\left(u_1, ..., u_n\right) = \varphi^{-1(n)}\left(\varphi\left(u_1\right) + ... + \varphi\left(u_n\right)\right) \prod_{i=1}^{n} \varphi'\left(u_i\right)$$

where $\varphi^{-1(n)}$ denotes the n-th derivative of the inverse generator function. Unfortunately, for hierarchical copulas such a simple expression for the density is not available any more. Due to the more complex structure of hierarchical copulas, we pursue a recursive approach. We differentiate the n-dimensional top level copula $C_{L,1}$ with respect to its arguments $\mathbf{u}_{L,1}$, using the chain rule. It is at this point that the two different notations of the copula arguments prove useful. Note that at the top level

$$\begin{aligned}
C_{L,1}\left(\mathbf{u}_{L,1}\right) &= C_{L,1}\left(\mathbf{C}_{L,1}\right) \\
&= C_{L,1}\left(C_{L-1,1}, ..., C_{L-1,m_{L-1}}\right) \\
&= C_{L,1}\left(C_{L-1,1}\left(\mathbf{C}_{L-1,1}\right), ..., C_{L-1,m_{L-1}}\left(\mathbf{C}_{L-1,m_{L-1}}\right)\right) \\
&= C_{L,1}\left(C_{L-1,1}\left(\mathbf{u}_{L-1,1}\right), ..., C_{L-1,m_{L-1}}\left(\mathbf{u}_{L-1,m_{L-1}}\right)\right) \qquad (3.4.2)
\end{aligned}$$

In order to derive the density

$$c_{L,1}\left(\mathbf{u}\right) = \frac{\partial^n C_{L,1}}{\partial u_1 ... \partial u_n} \qquad (3.4.3)$$

of the hierarchical copula we apply the chain rule,

$$\begin{aligned}
\frac{\partial^n C_{L,1}}{\partial u_1 \ldots \partial u_n} &= \sum \frac{\partial^{n-i} C_{L,1}}{\partial C_{L-1,1}^{k_1} \ldots \partial C_{L-1,m_{L-1}}^{k_{m_{L-1}}}} \\
&\quad \times \prod_{r=1}^{m_{L-1}} \sum_{\mu=\{\upsilon_1,...,\upsilon_r\}} \frac{\partial^{|\upsilon_1|} C_{L-1,r}}{\partial \upsilon_1} \cdots \frac{\partial^{|\upsilon_r|} C_{L-1,r}}{\partial \upsilon_r} \qquad (3.4.4)
\end{aligned}$$

where the outer sum extends over all sets of integers $k_1, ..., k_{m_{L-1}} \in \aleph \cup \{0\}$ such that $\max_j k_j \leq n_{L-1,j}$ and $\sum_{j=1}^{m_{L-1}} k_j = n - i,\ \forall i = 0, 1, ..., , n - m_{L-1}$. These terms are the outer derivatives of the copula with respect to the elements of $\mathbf{C}_{L,1}$, that is, the m_{L-1} copulas from level $L-1$. The second part of the formula are the inner derivatives, corresponding to the derivatives of the copulas at level $L-1$ with respect to their arguments $\mathbf{u}_{L-1,j}$. The summation in the inner derivative is over all $k_r \in \{0, 1, ..., n_{L-1,r}\}$ distinct subsets $\{v_1, ..., v_r\}$ from $\mathbf{u}_{L-1,r}$. To obtain the inner derivative we have to aggregate all m_{L-1} copulas at level $L-1$ over all possible combinations of extracting k_r elements from an $n_{L-1,r}$-dimensional object (the arguments $\mathbf{u}_{L-1,r}$ of copula $C_{L-1,r}$). The number of possible combinations for extracting k_r elements from an $n_{L-1,r}$-dimensional object is

$$\prod_{s=0}^{n_{L-1,r-1}} \frac{\begin{pmatrix} n_{L-1,r} - \sum_{s=0}^{n_{L-1,r-1}} q_s \\ q_{s+1} \end{pmatrix}}{(\#q_s)!}$$

where q_s represents derivatives of different orders such that the total number of q's equals k_r , that is, $(\#q_s) = k_r$ for all s, with $q = 0$ and $\sum_{s=0}^{n_{L-1,r}} q_s = n_{L-1,r}$ for $q_s \in \{0, 1, ..., n_{L-1,r}\}$. We compute the inner derivative by multiplying the sums over all M_{L-1} copulas from level $L-1$. The algorithm is recursive: the inner derivatives of $C_{L,1}$ involve the partial derivatives of $C_{L-1,1}, ..., C_{L-1,m_{L-1}}$, which in turn can be computed using equation 3.4.4 for the next lower level. The recursion ends at the lowest level $C_{L,r}$ when only partial derivatives (of different orders) of ordinary Archimedean copulas are required.

3.4.2 Simulation

Random numbers generation from Archimedean copulas is trivial even in higher dimensions. However, the method of simulation does not carry over to hierarchical Archimedean copulas. We now have to go back to first principles and generate random numbers from hierarchical Archimedean copulas by the conditional inversion method.

Consider a (not necessarily Archimedean) copula $C = C(u_1, ..., u_n)$. The task is to generate n-tuples $u_1, ..., u_n$ of observations of $U(0,1)$ distributed random variables $U_1, ..., U_n$ whose joint distribution function is the copula C, that is, $U_1, ..., U_n \sim C$. To this end one has to invert the conditional distribution. Let $C_k(u_1, u_2, ..., u_k) = C(u_1, u_2, ..., u_k, 1, ..., 1), k = 1, .., n$ denote the k-dimensional margin of C, with $C_1(u_1) = u_1$ and $C_n(u_1, u_2, ..., u_n) = C(u_1, ..., u_n)$. The conditional distribution of U_k given the values of $U_1, ..., U_{k-1}$ is given by

$$\begin{aligned} C_k(u_k | u_1, u_2, ..., u_{k-1}) &= P(U_k \leq u_k | U_1 = u_1, ..., U_{k-1} = u_{k-1}) \\ &= \frac{\partial^{k-1} C_k(u_1, u_2, ..., u_k)}{\partial u_1 ... \partial u_{k-1}} \Bigg/ \frac{\partial^{k-1} C_k(u_1, u_2, ..., u_{k-1})}{\partial u_1 ... \partial u_{k-1}} \end{aligned}$$

for $k = 2, ..., n$.

The simulation algorithm may be given as:

1. Generate n independent uniform random variates $\upsilon_1, ..., \upsilon_n$.

2. Set $u_1 = \upsilon_1$

3. For $k = 2, ..., n$ evaluate the inverse of the conditional distribution function, that is, the conditional quantile function, at υ_k to generate $u_k = C_k^{-1}(\upsilon_k | u_1, u_2, ..., u_{k-1})$. The inverse can be derived either analytically or numerically.

The result of this simulation algorithm is a vector of random numbers $(u_1, u_2, ..., u_n)$ with joint distribution function C. A simple Enterprise Risk Management model entails setting up various risk sub-models and creating a dependency relationship between these risks. After this is accomplished all one needs to do is to simulate for a given number of trials (say 5,000) and aggregate the dependent risk values. From these aggregated results one can then either determine VaR or conditional tail expectation (CTE) at a specific percentile.

3.4.3 Estimating hierarchical copulas

Hierarchical Archimedean copulas are estimated by maximum likelihood if the margins are known (up to a vector of parameters), and canonical maximum likelihood method if the margins are estimated nonparametrically by their empirical distribution function, Savu and Trede (2006). For the canonical maximum likelihood method, the density $c_{L,1}(\mathbf{u})$ of the hierarchical copula depends on the unknown parameter vector θ with elements $\theta_{l,j}$ for $l = 1, ..., L$ and $j = 1, ..., m_l$. The number of elements of θ

corresponds to the number of generators used to construct the hierarchical copula. Let $(U_{11},...,U_{n1}),....,(U_{1T},...,U_{nT})$ be a random sample of size T from the n-dimensional random vector $(U_1,...,U_n)$.The (canonical) loglikelihood function is

$$\ln L\left(\theta; u_{it}, i=1,...,n, t=1,...,T\right) = \sum\nolimits_{t=1}^{T} \ln c_{L,1}\left(u_{1t},...,u_{nt};\theta\right)$$

and the canonical maximum likelihood estimator of θ is

$$\hat{\theta} = \arg\max \ln L\left(\theta\right)$$

Since the density of hierarchical Archimedean copulas is quite complex, a closed form expression for the maximum likelihood estimators cannot be given. It is, therefore, necessary to apply numerical optimisation algorithms.

3.4.4 Mixed copulas

We build up a mixture of copulas where each copula bears a certain weight and features the dependence structure between one particular pair of variables. In particular, with 4 random variables, we need six copulas to characterize all the dependence relations. In each copula, we assume that only one pair of variables has a dependence structure and the other variables are all independent with each other and with the dependent pair. For instance, we first model the dependence between x_1 and x_2 by a copula $C_1(u_1;u_2)$ and then construct a copula with four variables by multiplying $C_1(u_1; u_2)$ with u_3 and u_4: $C_1(u_1; u_2)u_3u_4$ is a copula by Theorem 3.5.3 in Nelsen (2006) and in this copula x_3 and x_4 are independent with each other and independent with x_1 and x_2: Similarly, we can construct the other five copulas as $C_2(u_3; u_4)u_1u_2$; $C_3(u_1; u_4)u_2u_3$; $C_4(u_1; u_3)u_2u_4$; $C_5(u_2; u_4)u_1u_3$ and $C_6(u_2; u_3)u_1u_4$. Consequently, the mixture of copulas C_M can be given as

$$\begin{aligned} C_M\left(u_1, u_2, u_4; \pi, \rho\right) &= \pi_1 C_1\left(u_1, u_2; \rho_1\right) u_3 u_4 + \pi_2 C_2\left(u_3, u_4; \rho_2\right) u_1 u_2 + \pi_3 C_3\left(u_1, u_4; \rho_3\right) \\ &\quad + \pi_4 C_4\left(u_1, u_3; \rho_4\right) u_2 u_4 + \pi_5 C_5\left(u_2, u_4; \rho_5\right) u_1 u_3 \\ &\quad + \left(1 - \pi_1 - \pi_2 - \pi_3 - \pi_4 - \pi_5\right) C_6\left(u_2, u_3; \rho_6\right) u_1 u_4 \end{aligned}$$

where $\pi = [\pi_1, \pi_2, \pi_3, \pi_4, \pi_5, \pi_6]'$ with $\sum_{i=1}^{6} \pi_i = 1$ accounts for the weights for each copula and $\rho = [\rho_1, \rho_2, \rho_3, \rho_4, \rho_5, \rho_6]'$ is the vector of parameters in each copula. Accordingly, the density form C_M is given by

$$\begin{aligned} c_M\left(u_1, u_2, u_4; \pi, \rho\right) &= \pi_1 c_1\left(u_1, u_2; \rho_1\right) + \pi_2 c_2\left(u_3, u_4; \rho_2\right) + \pi_3 c_3\left(u_1, u_4; \rho_3\right) \\ &\quad + \pi_4 c_4\left(u_1, u_3; \rho_4\right) + \pi_5 c_5\left(u_2, u_4; \rho_5\right) \\ &\quad + \left(1 - \pi_1 - \pi_2 - \pi_3 - \pi_4 - \pi_5\right) c_6\left(u_2, u_3; \rho_6\right) \end{aligned}$$

where c_i is the density form of C_i for $i = 1, ..., 6$. The approach above was first discussed in Tasfack (2006) where as one component of his model, he uses a mixture of copulas to model the joint distribution of 4 international assets. In his model, each copula component of the mixture copula simultaneously characterizes the dependence structure of two pairs of variables.

Chapter 4

SIMULATION STUDY

4.1 Introduction

We considered application in General Insurance and in particular the short - term policies. Short term policies are those that run for a short and fixed duration of time, with one year being a typical duration. General Insurance is chosen due to complexities of other forms of insurance. Data were simulated following loss distributions that are popular in credibility rate making in the insurance sector. The loss distributions are peculiar due to their unique characteristics that include:

1. Their range is all the non-negative real numbers due to the fact that the possibility of a negative claim has no practical meaning.

2. They are positively skewed since the very large claims are atypical and they have a lesser chance than the ordinary one.

3. The very long upper tail is another general characteristic in order to accommodate the chance of the extremely large claims.

4.2 Simulation of the Loss Distributions

Four lines of business were simulated in R Development Core Team (2009) with each line having four sub-classes which gives a profile of sixteen risks. The first line was from the log-normal distribution. A log-normal distribution is a probability distribution of a random variable whose logarithm is normally distributed. If X is a random variable with a normal distribution, then $Y = exp(X)$ has a log-normal distribution; likewise, if Y is log-normally distributed, then $X = log(Y)$ is normally distributed. The log-normal distribution has density:

$$f_X(x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}, x > 0$$

where μ and σ are the mean and standard deviation of the logarithm. It consists of x_1, x_2, x_3 and x_4 risks. The parameters were: $meanlog = 1.9$ and $sdlog = exp(-1)$, $meanlog = 2.4$ and $sdlog = exp(-0.6)$, $meanlog = 2.4$ and $sdlog = exp(-0.8)$, and $meanlog = 1.9$ and $sdlog = exp(-0.6)$ respectively the data being expressed in KES 100,000.

The second line had x_5, x_6, x_7 and x_8 risks which were from the Burr distribution. The Burr distribution with parameters shape1= α, shape2 = γ and scale = θ has the density:

$$f(x) = \frac{\alpha\gamma (x/\theta)^\gamma}{x \left[1 + (x/\theta)^\gamma\right]^{\alpha+1}}$$

for $x > 0, \alpha > 0, \gamma > 0$ and $\theta > 0$. The two shape parameters, rate and scale parameters were: 3, 1.5, rate = 5 , scale = 2.5, for x_5 , 1.5, 3, rate = 5 , scale = 2.5 being for x_6 while 1.5, 2, rate = 5 , scale = 2.5, and 2, 1.5, rate = 10 , scale = 5 were for x_7 and x_8 respectively. The data were again in KES 100,000 as in the case of the first line of business.

Next, we had the third line whose data where from the single parameter Pareto distribution otherwise known as the European Pareto, Faivre (2003), consisting of x_9, x_{10}, x_{11} and x_{12} risks. The Single-parameter Pareto distribution with parameter shape = α has density:

$$f(x) = \frac{\alpha\theta^{\alpha}}{x^{\alpha+1}}$$

for $x > \theta, \alpha > 0$ and $\theta > 0$. Although there appears to be two parameters, only shape is a true parameter. The following were the parameters 3.2, 2.6, 2.1 and 1.8 all starting from KES 100,000. Lastly, the fourth line followed the inverse Weibull distributions with the risks being x_{13}, x_{14}, x_{15} and x_{16}. The inverse Weibull distribution with parameters shape = τ and scale = θ has density:

$$f(x) = \frac{\tau(\theta/x)^{\tau} e^{-(\theta/x)^{\tau}}}{x}$$

for $x > 0, \tau > 0$ and $\theta > 0$. The parameters were: 1.8, rate = 5, scale = 2.5; 1.5, rate = 5, scale = 2.5; 2.1, rate = 10, scale = 5; and 1.2, rate = 10, scale = 5 respectively. Once more the data were in KES 100,000 as in the case of the previous lines of business.

4.3 Data Exploration

The simulated data were explored in order to ascertain that they were fit for the problem at hand. This was done by plotting the densities of each line of business and the plots are as shown in Figures 4.1 and 4.2.

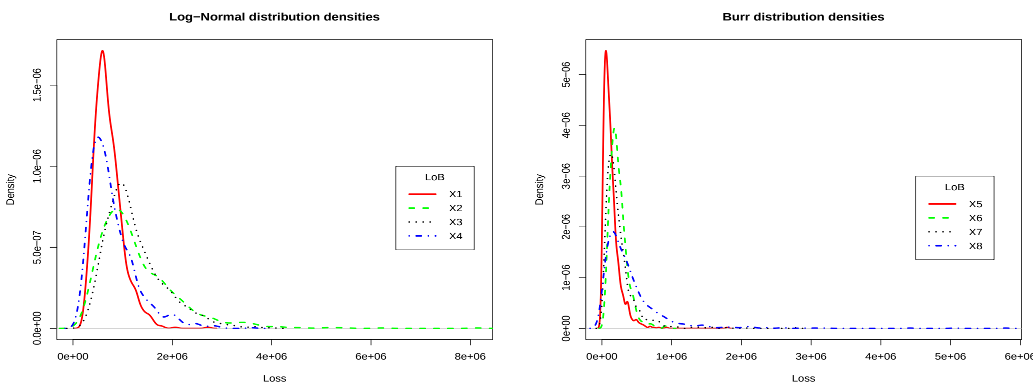


Figure 4.1: *Densities for the first and second LOBs*

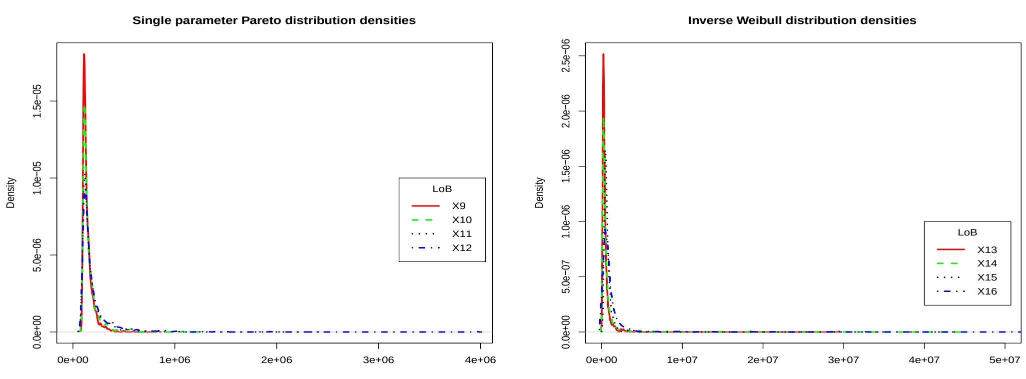


Figure 4.2: *Densities for the third and fourth LOBs*

From the densities in Figures 4.1 and 4.2, we can see that they abide by the unique characteristics of loss distributions outlined above.

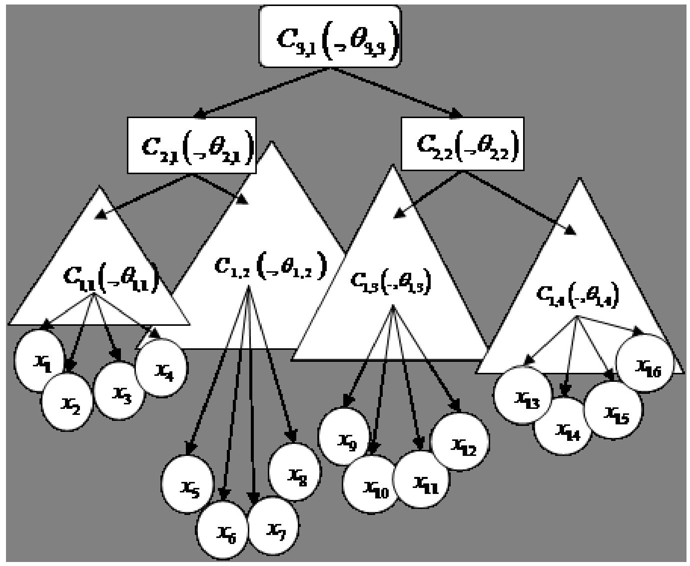


Figure 4.3: *The hierarchical structure of the simulated scenario*

4.4 Results and Discussion

The Gumbel Copulas were fitted to explore the dependence between the various LOB at the lower level ($l = 1$) in Figure 4.3 with the generators $\varphi_{l,j}(x) = (-\ln x)^{\theta_{l,j}}$, $j = 1, 2, 3, 4$. It (Gumbel) was chosen owing to its strength in capturing the upper tail dependence (see section 2.3.3) and our loss distributions so obtained are *active* in the upper tails. The other Archimedean copulas (that is, Frank and Clayton) would under estimate the dependence in the upper tail since they have no upper tail dependence. The fitted dependence parameters were $\theta_{1,1} = 2.076$, $\theta_{1,2} = 8.829$, $\theta_{1,3} = 16.512$, and $\theta_{1,4} = 9.778$.

The copulas modelled the dependence structure as well as isolating the marginals as shown in Figures 4.4, 4.5, 4.6 and 4.7.

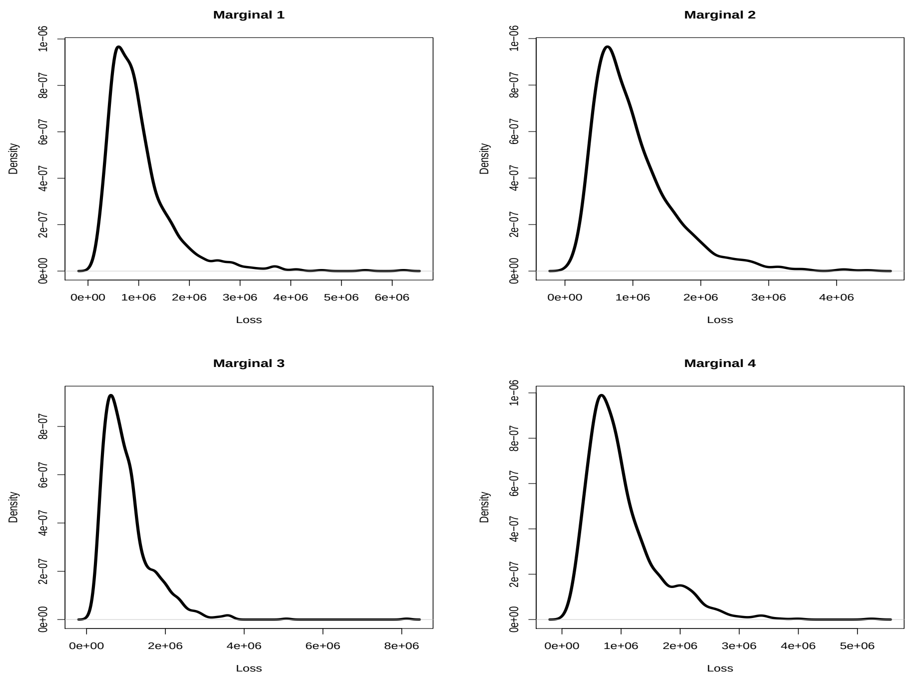


Figure 4.4: *The marginal distributions for the classes in the first LOB*

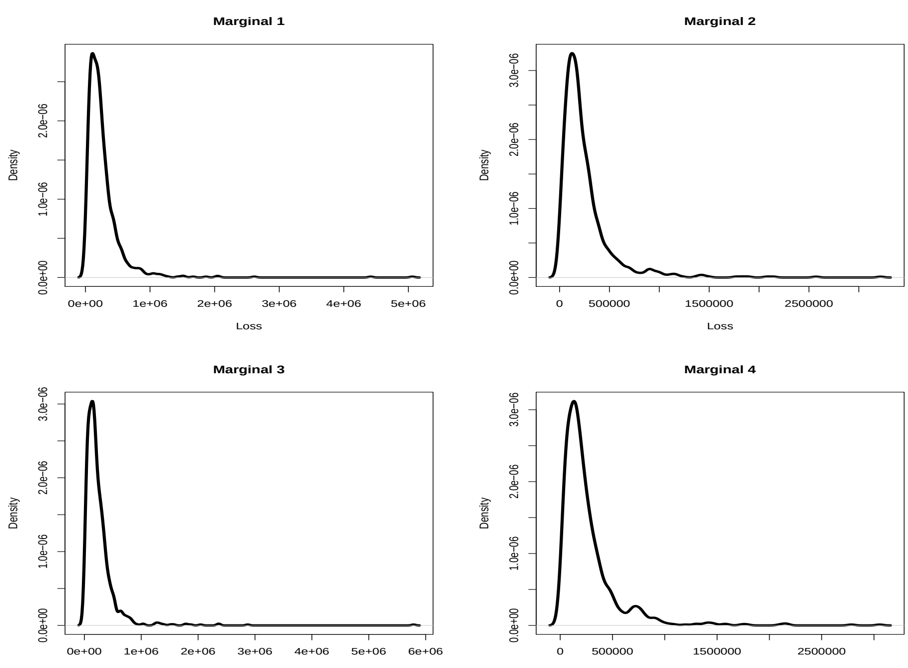


Figure 4.5: *The marginal distributions for the classes in the second LOB*

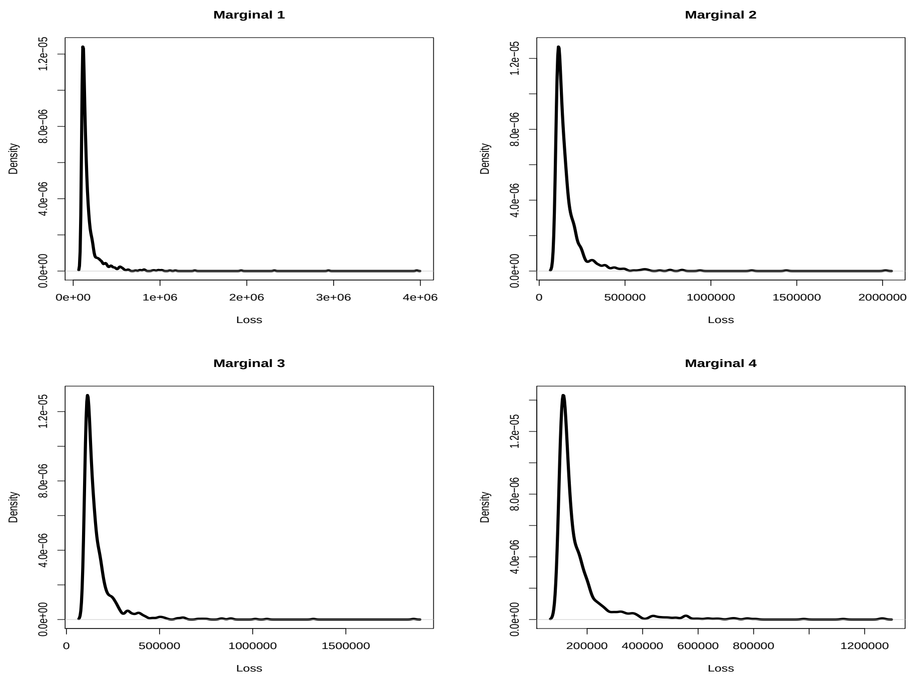


Figure 4.6: *The marginal distributions for the classes in the third LOB*

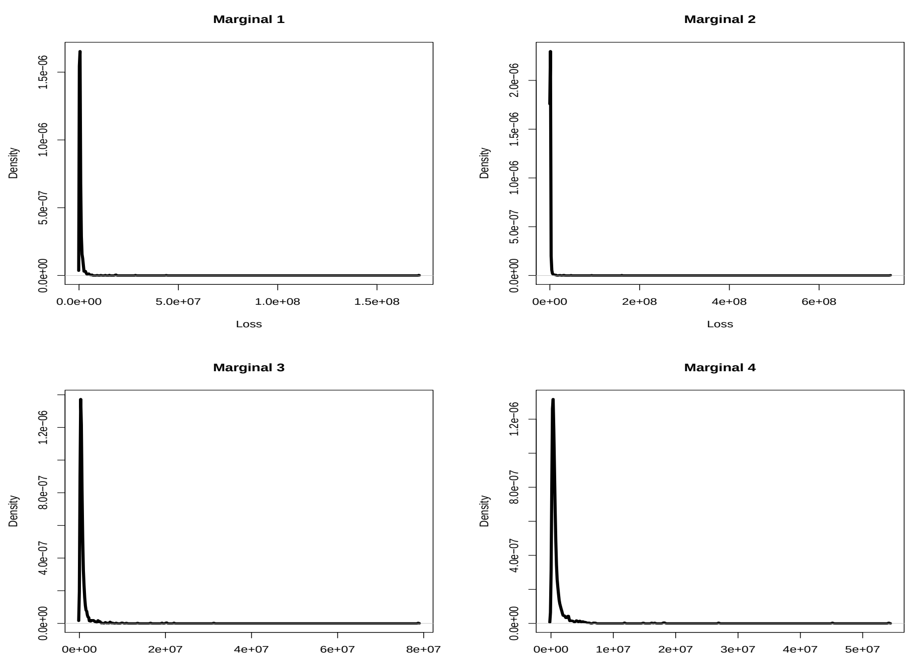


Figure 4.7: *The marginal distributions for the classes in the fourth LOB*

Note, from the figures, that the PDFs are not symmetric, which is a characteristic of the Gumbel Copula since it is only defined in the right tail. This makes the Gumbel Copula the most appropriate copula to define the dependencies at this level. The properties of the data emanating from level one, in Figure 4.3, were examined and the Frank copula was found to be the more preferred to model the dependence structure. These data now enter into the model in level two ($l = 2$). The generators were

$$\varphi_{l,j}(x) = -\ln\left\{\frac{\exp(-\theta_{l,j}x) - 1}{\exp(-\theta_{l,j}) - 1}\right\}$$

and the fitted dependence parameters were $\theta_{2,1} = 1.500$ and $\theta_{2,2} = 2.037$. The marginals also support the use of the Frank copula which has no tail dependence as shown in Figure 4.8. In the third level, which was our upper level, the Frank copula was again used and the parameter being $\theta_{3,1} = 0.328$.

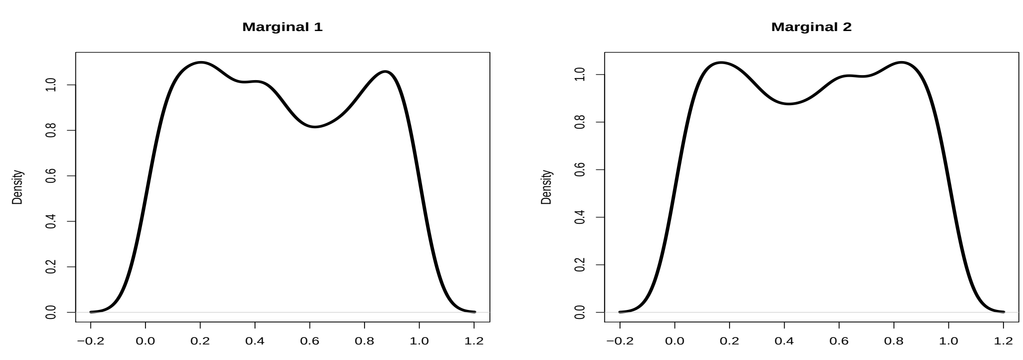


Figure 4.8: *The marginal distributions at level $l = 2$*

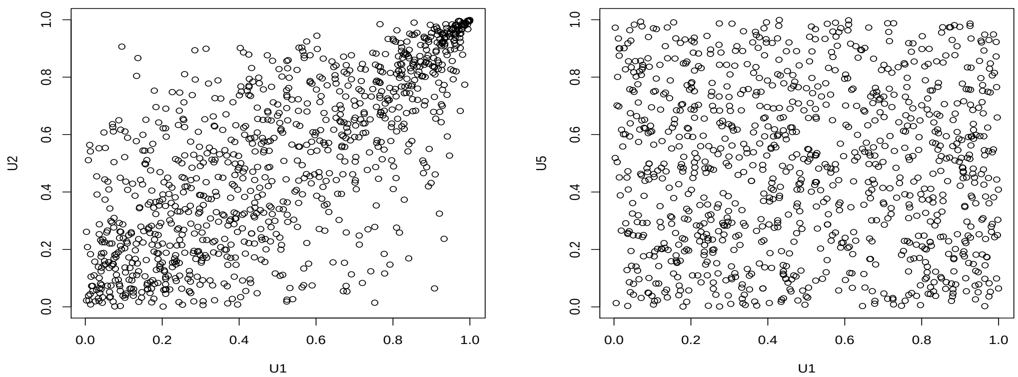


Figure 4.9: *Scatterplots of two bivariate marginal distribution*

The results from this chapter are summarised in Table 4.1.

Table 4.1: *The fitted dependence parameter θ for the different levels in the hierarchy*

	$j=1$	$j=2$	$j=3$	$j=4$
$l=1$	2.076	8.829	16.512	9.778
$l=2$	1.500	2.037		
$l=3$	0.328			

Since $\theta_{2,1}, \theta_{2,2} < \theta_{1,1}, \theta_{1,2}, \theta_{1,3}, \theta_{1,4}$, and $\theta_{3,1} < \theta_{2,1}, \theta_{2,2}$ the conditions, for a proper hierarchical copula, $\varphi_{l+1,i} \circ \varphi_{l,j}^{-1} \in L_\infty^* \forall\ l = 1, ..., L$ and $j = 1, ..., m_l,\ i = 1, ..., m_{l+1}$ such that $C_{l,j} \in C_{l+1,i}$ are satisfied. Let $U_i, i = 1, ..., 16$ be the marginals at the lower level, then the dependence between U_1 and U_2 is stronger than that between U_1 and U_5 as shown by Figure 4.9. The pairs in the same line have dependence similar to U_1 and U_2 while those in different lines have dependence similar to U_1 and U_5 .

Chapter 5

COPULA BASED CLASSIFICATION OF RISKS

5.1 The proposed criterion for grouping business lines using the dependence structures

This section presents the proposed algorithm for grouping business classes into various lines. For the different business classes follow the algorithm below to cluster them into their respective lines (or departments):

1. Fit the Copula function (section 1.3) for each pair of business classes

2. Estimate the dependence parameter (as found in sub-section 3.4.3)

3. Calculate the measures of dependence, the rank correlation (Spearman's rho or the Kendall's tau) and the tail dependence, using the relationships in section 1.4.

4. Compare closeness of these measures to each other by calculating appropriate distances culminating in a distance matrix (see section 2.4).

5. Cluster the business classes into the various homogeneous lines or departments using the minimum distance approach (see section 2.5).

This will result in the highly related classes being in one line while the less dependent classes will be in different lines. The lines of business will form a portfolio and hence increase the diversification benefits.

5.2 Application of the clustering algorithm on simulated data

The data simulated in Chapter 4 are used here to cluster the various business classes into LOB as they would 'naturally' cluster together. We compare the Euclidiean distance to the Manhattan distance on their performance towards clustering of the lines of business since our data are quantitative. The Mantel statistic is used here to measure the strength of the relationship between the cophenetic matrix and the distance matrix.

Euclidean distances were used as the criterion to cluster the business classes with respect to the spearman's rho, Kendall's tau and the Tail index so as to compare their performance against the Manhattan distances. Comparing the clustering based on the Euclidean distances and the Manhattan distances, the Cophenetic correlation coefficient (Mantel statistic) comes in handy in choosing between the best distance to use. The Manhattan distance performed better than the Euclidean distances (see Table 5.1) and so our clustering was based on the Manhattan distances.

From the dependence parameter, $theta$, matrix (see Table 5.2), it may not be straight forward to examine the dependence between two variables since $theta$ can be in the range $(-\infty, \infty)$ like in the case of the Frank copula. It is therefore important to re-scale this dependence into a definite scale such as that for Kendall's tau or Spearman's rho $[-1, 1]$. This is the motivation behind Table 5.3 on the Kendall's tau calculated from the fitted dependence parameter and Table 5.4, for the Spearman's rho, from Table 5.2. For the purpose of clustering the various LOB, with respect to their dependence, one needs a criterion of telling how close every two lines are. Here we calculate the distances between each pair giving rise to the distance matrices in Table 5.6, Table 5.7 and Table 5.8.

The cluster dendrograms in Figure 5.1, Figure 5.2 and Figure 5.3 utilised the Euclidean distances.

If the objective is that of establishing four LOB, then the departments (or lines) will be as shown in Figure 5.1. It is also worth noting that any particular dependence measure will yield similar results as Figure 5.2 and Figure 5.3 depict about the Euclidean distances. Comparing clusters for the four lines, in the case of Euclidean and the Manhattan distances; the four clusters are $(X_6, X_{15}, X_2, X_{11}, X_4, X_{12}, X_{10}, X_3, X_1 \textit{ and } X_9)$, $(X_{16}, X_7, X_{13} \textit{ and } X_{14})$, (X_5) and (X_8) for the Euclidean distance while from the Manhattan distances we have $(X_{10}, X_3, X_1 \textit{ and } X_9)$, $(X_2, X_{11}, X_{15}, X_4, X_{12}, X_6 \textit{ and } X_{13})$, $(X_{16}, X_7 \textit{ and } X_{14})$ and $(X_5 \textit{ and } X_8)$. It can be seen that when the Mantel statistic is low, the clustering criterion may miss out on some dependence information.

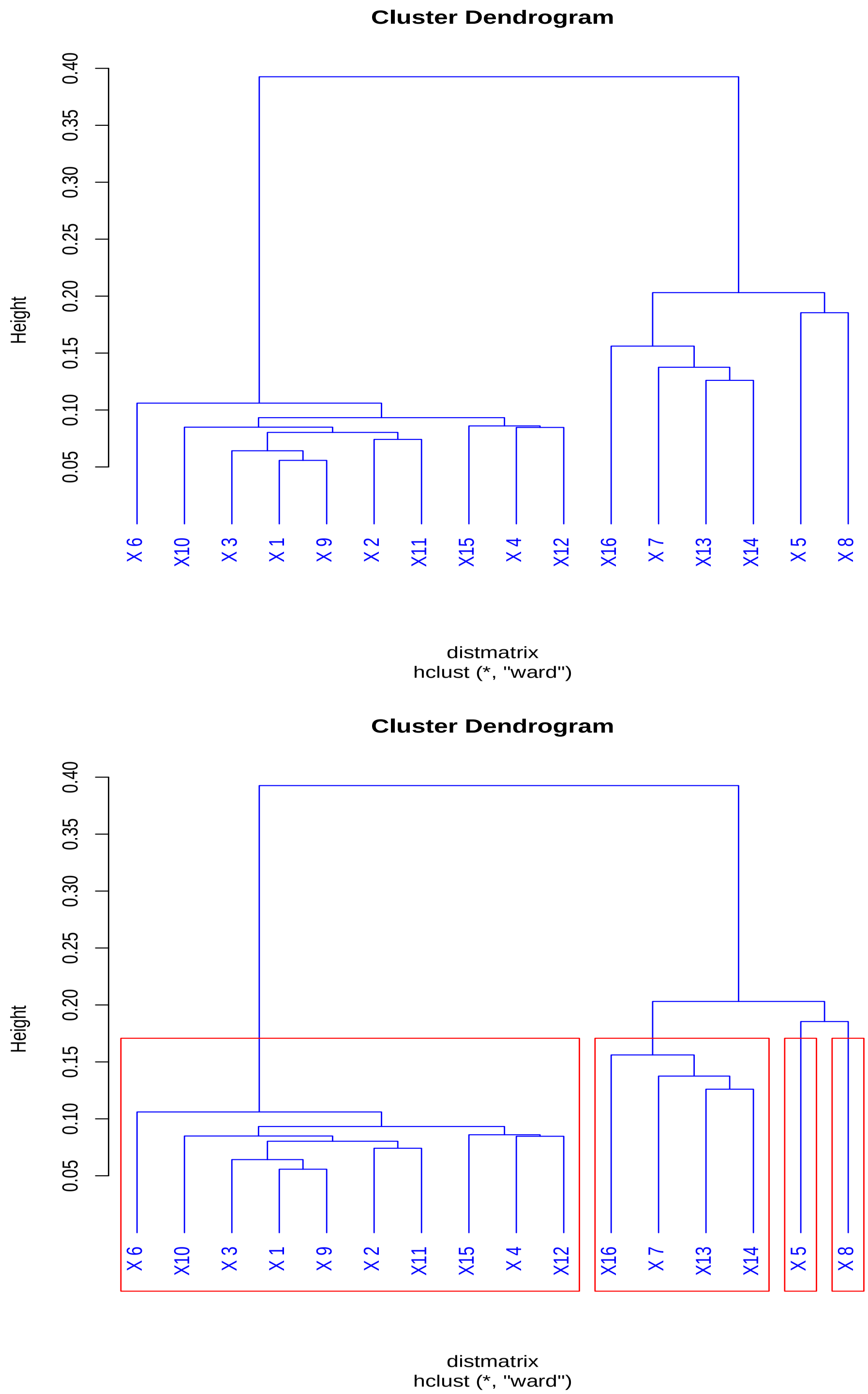


Figure 5.1: *Dendrograms clustered by the Euclidean distances for the Kendall's tau* τ

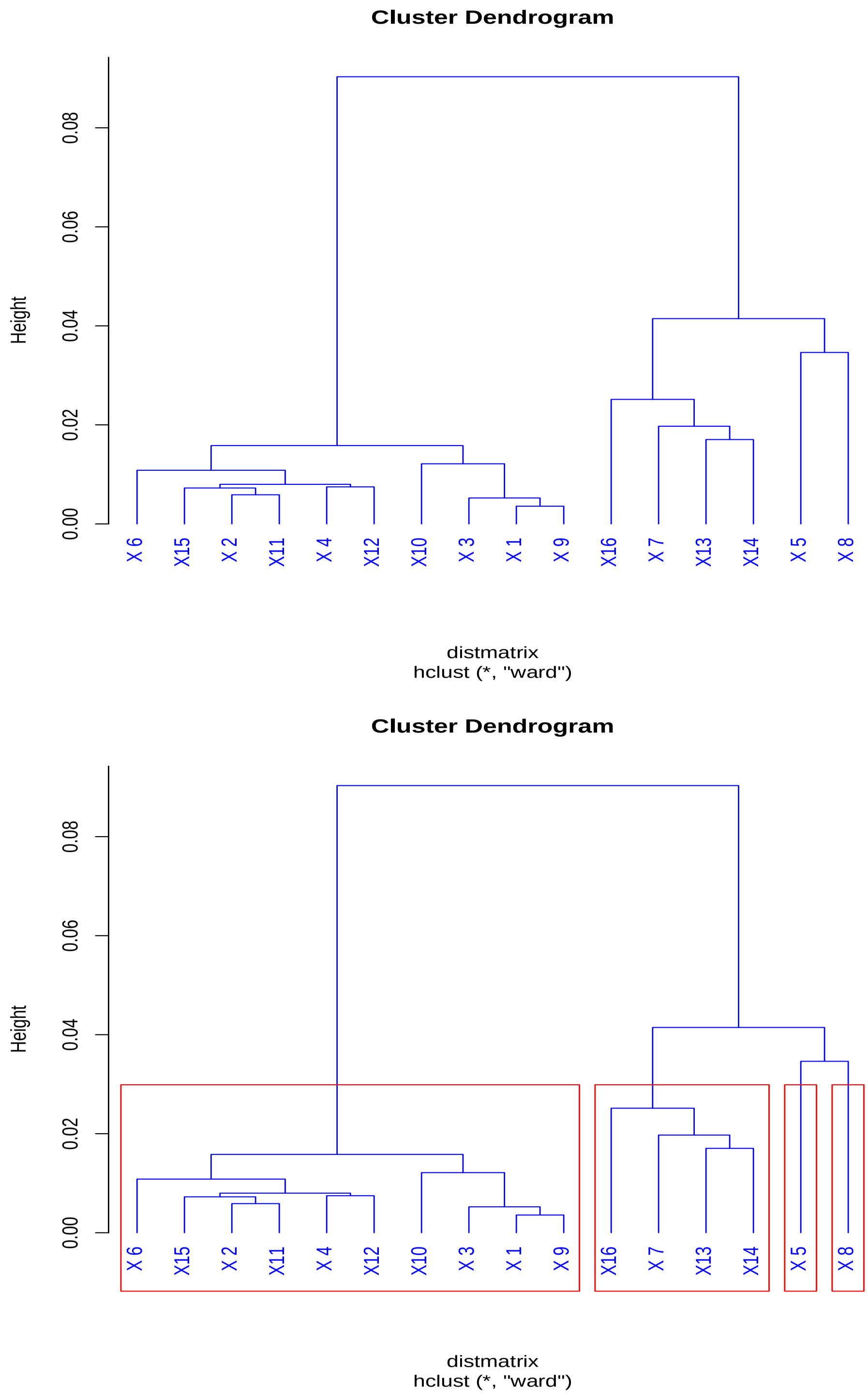
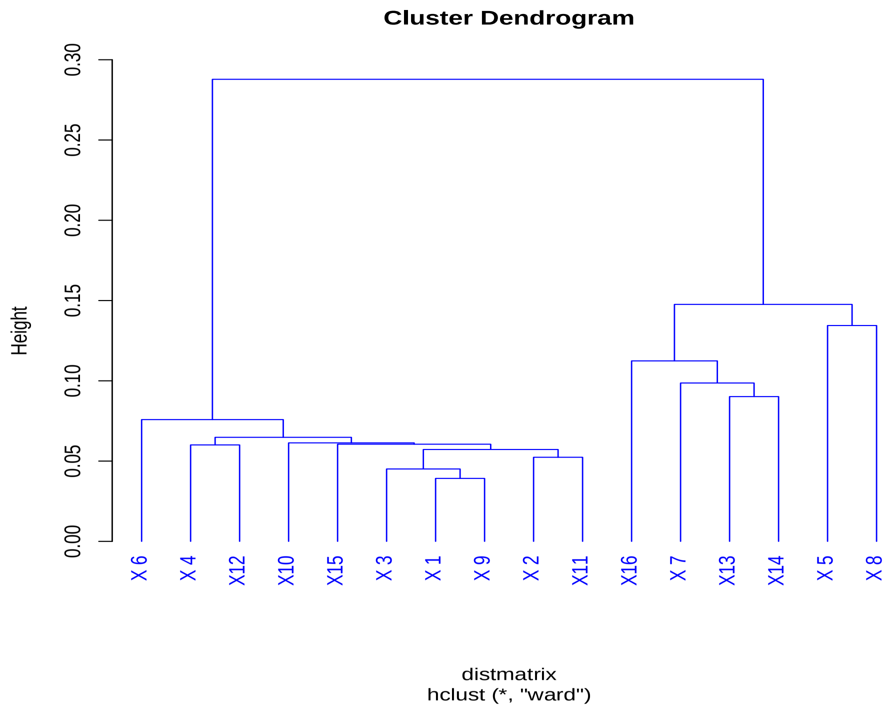


Figure 5.2: *Dendrograms clustered by the Euclidean distances for the Spearman's rho* ρ



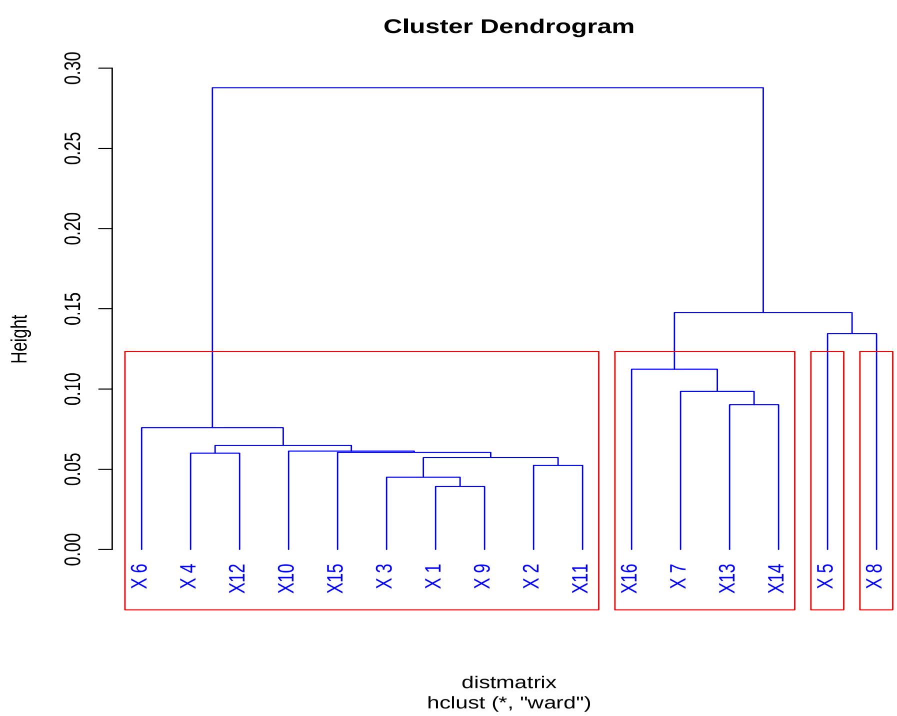


Figure 5.3: *Dendrograms clustered by the Euclidean distances for the tail index*

Table 5.1: *Comparison between the performance of the Euclidean and the Manhattan distances*

Distance				
	Manhattan		Euclidean	
Measure	Mantel statistic	Significance	Mantel statistic	Significance
Kendall's tau	0.8479	<0.01	0.7209	<0.01
Spearman's Rho	0.8479	<0.01	0.7209	<0.01
Tail index	0.8477	<0.01	0.7272	<0.01

From Table 5.1 despite both the Manhattan and Euclidean distances having significance correlation to their corresponding cophenetic matrices, the Manhattan distance has a stronger association as compared to the Euclidean distance. We therefore consider the clusters in the dendrograms emanating from the Manhattan distances in Figure 5.4, Figure 5.5 and Figure 5.6.

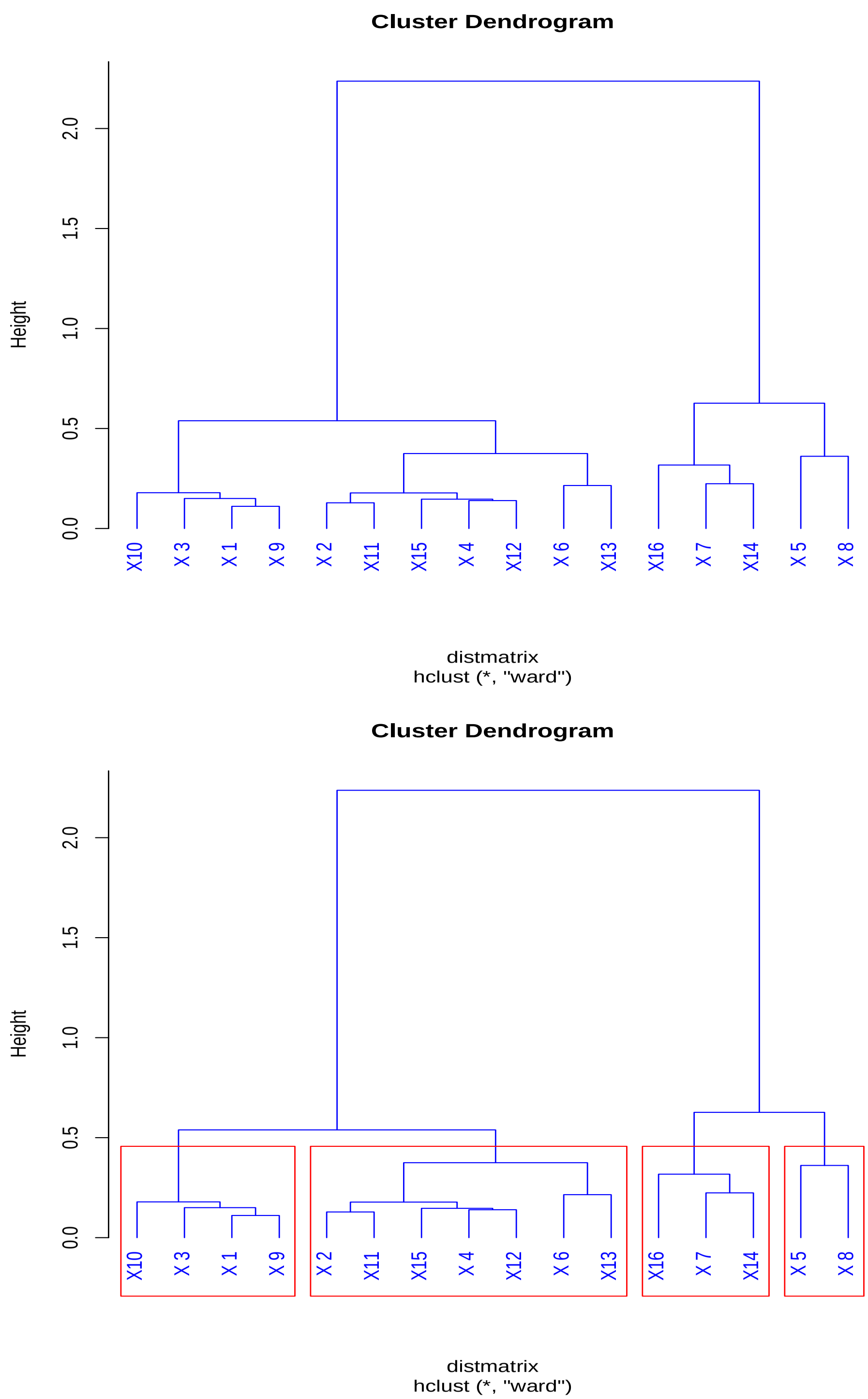
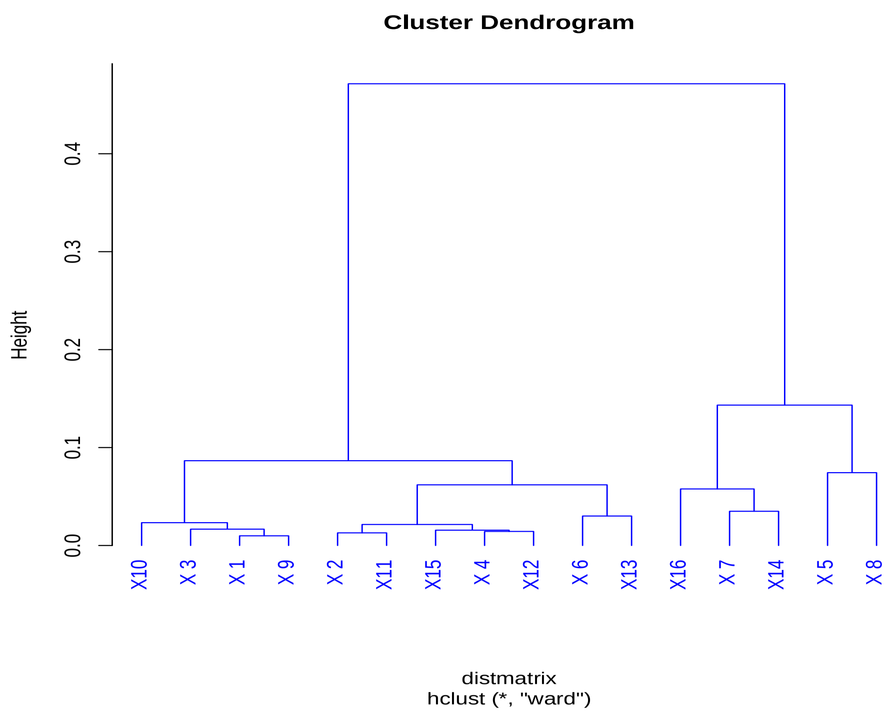


Figure 5.4: *Dendrograms clustered by the Manhattan distances for the Kendall's tau* τ



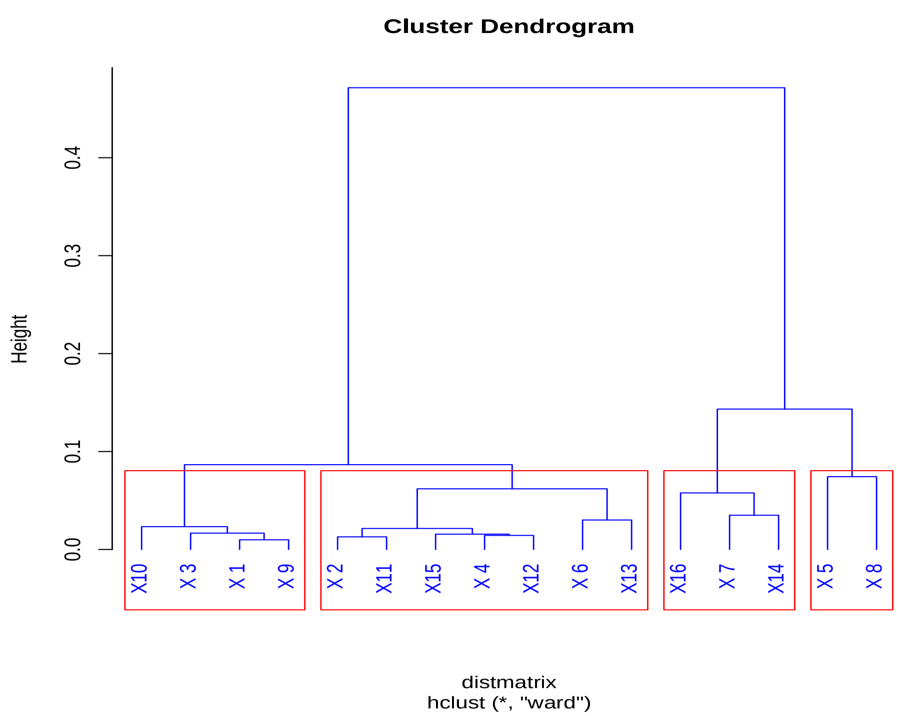
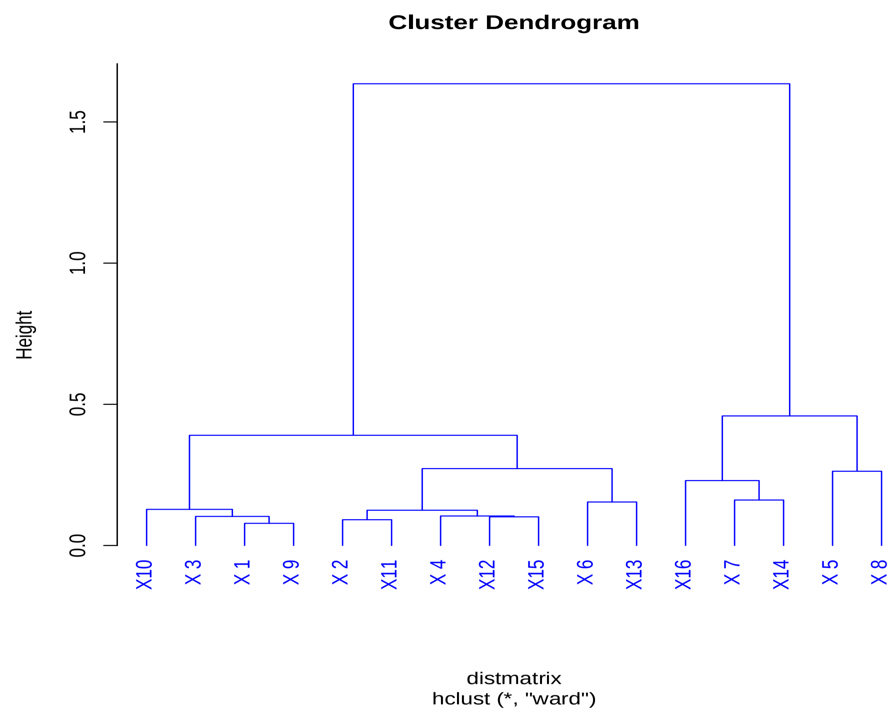


Figure 5.5: *Dendrograms clustered by the Manhattan distances for the Spearman's rho* ρ



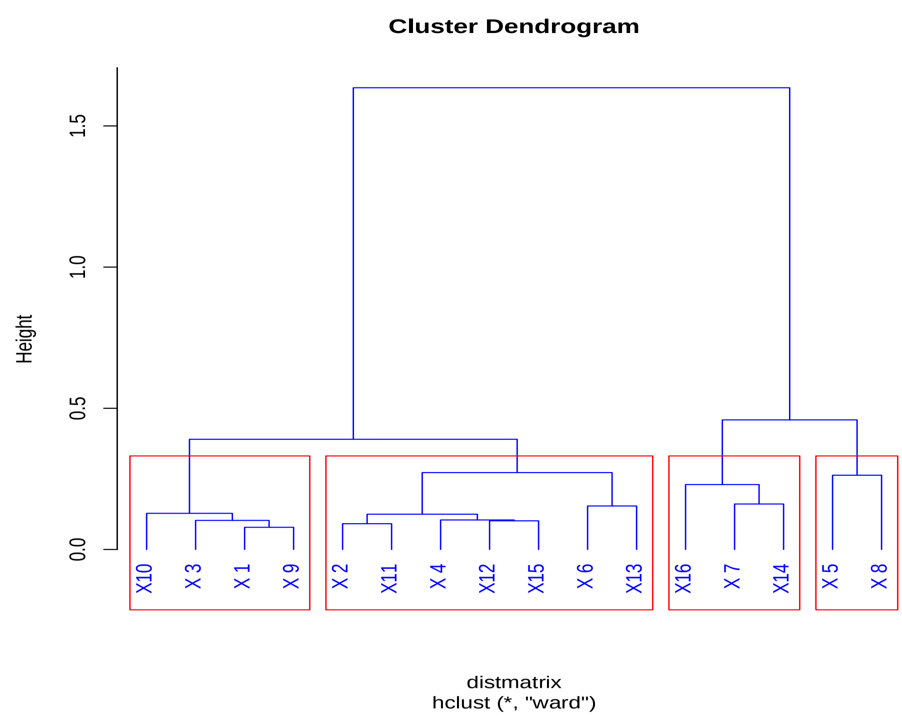


Figure 5.6: *Dendrograms clustered by the Manhattan distances for the tail index* τ

It can be concluded that the choice of distances for clustering is very crucial as they can vary depending on the problem at hand. This problem can be surmounted by using the Cophenetic correlation coefficient (the Mantel statistic). The strength of the relationship, from the Mantel statistic, dictates the strength of the clustering criterion.

Table 5.2: *The dependence parameter θ estimated for each pair of business classes*

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}
X_2	20.45														
X_3	23.44	19.83													
X_4	19.81	16.90	18.39												
X_5	11.31	10.59	10.38	10.10											
X_6	17.96	16.37	16.88	16.09	9.69										
X_7	14.35	13.11	13.77	12.43	8.55	11.74									
X_8	11.86	11.28	11.45	10.75	7.83	10.00	9.03								
X_9	25.86	21.92	23.88	20.72	10.87	19.58	14.51	11.78							
X_{10}	22.25	19.89	22.06	18.60	19.30	16.93	13.80	11.28	25.07						
X_{11}	21.47	19.18	19.69	17.83	10.44	16.56	12.86	10.73	22.16	20.20					
X_{12}	19.98	17.71	18.49	16.76	10.05	15.40	12.30	10.95	20.71	19.93	17.60				
X_{13}	17.47	15.04	16.09	14.84	9.29	13.42	11.22	9.77	16.98	15.84	14.93	14.87			
X_{14}	14.86	13.12	13.59	12.29	8.63	12.12	10.34	8.78	14.59	14.07	13.79	12.81	11.68		
X_{15}	19.41	17.28	19.41	16.45	10.41	16.07	12.57	10.95	21.83	19.46	18.15	16.69	14.37	12.88	
X_{16}	12.79	12.11	12.88	11.98	8.30	11.34	9.61	8.63	12.93	12.59	11.85	11.64	10.31	9.95	11.60

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}
X_2	0.951														
X_3	0.957	0.950													
X_4	0.950	0.941	0.946												
X_5	0.912	0.906	0.904	0.901											
X_6	0.944	0.939	0.941	0.938	0.897										
X_7	0.930	0.924	0.927	0.920	0.883	0.915									
X_8	0.916	0.911	0.913	0.907	0.872	0.900	0.889								
X_9	0.961	0.954	0.958	0.952	0.908	0.949	0.931	0.915							
X_{10}	0.955	0.950	0.955	0.946	0.948	0.941	0.928	0.911	0.960						
X_{11}	0.953	0.948	0.949	0.944	0.904	0.940	0.922	0.907	0.955	0.950					
X_{12}	0.950	0.944	0.946	0.940	0.900	0.935	0.919	0.909	0.952	0.950	0.943				
X_{13}	0.943	0.934	0.938	0.933	0.892	0.925	0.911	0.898	0.941	0.937	0.933	0.933			
X_{14}	0.933	0.924	0.926	0.919	0.884	0.917	0.903	0.886	0.931	0.929	0.927	0.922	0.914		
X_{15}	0.948	0.942	0.948	0.939	0.904	0.938	0.920	0.909	0.954	0.949	0.945	0.94	0.93	0.922	
X_{16}	0.922	0.917	0.922	0.917	0.880	0.912	0.896	0.884	0.923	0.921	0.916	0.914	0.903	0.899	0.914

Table 5.3: *Kendalls tau τ calculated from the fitted dependence parameter θ*

Table 5.4: *Spearmans rho ρ calculated from the fitted dependence parameter θ*

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}
X_2	0.996														
X_3	0.997	0.996													
X_4	0.996	0.995	0.996												
X_5	0.989	0.987	0.986	0.986											
X_6	0.995	0.995	0.995	0.994	0.985										
X_7	0.993	0.992	0.992	0.991	0.980	0.989									
X_8	0.990	0.989	0.989	0.987	0.976	0.986	0.982								
X_9	0.998	0.997	0.997	0.997	0.988	0.996	0.993	0.990							
X_{10}	0.997	0.996	0.997	0.996	0.996	0.995	0.992	0.989	0.998						
X_{11}	0.997	0.996	0.996	0.995	0.987	0.995	0.991	0.987	0.997	0.996					
X_{12}	0.996	0.995	0.996	0.995	0.986	0.994	0.990	0.988	0.997	0.996	0.995				
X_{13}	0.995	0.994	0.994	0.993	0.983	0.992	0.988	0.985	0.995	0.994	0.993	0.993			
X_{14}	0.993	0.992	0.992	0.990	0.980	0.990	0.986	0.981	0.993	0.993	0.992	0.991	0.989		
X_{15}	0.996	0.995	0.996	0.995	0.987	0.994	0.991	0.988	0.997	0.996	0.996	0.995	0.993	0.991	
X_{16}	0.991	0.990	0.991	0.990	0.979	0.989	0.984	0.981	0.991	0.991	0.990	0.989	0.986	0.985	0.989

Table 5.5: *Tail index calculated from the fitted dependence parameter θ*

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}
X_2	0.966														
X_3	0.970	0.964													
X_4	0.964	0.958	0.962												
X_5	0.937	0.932	0.931	0.929											
X_6	0.961	0.957	0.958	0.956	0.926										
X_7	0.951	0.946	0.948	0.943	0.916	0.939									
X_8	0.940	0.937	0.938	0.933	0.907	0.928	0.920								
X_9	0.973	0.968	0.971	0.966	0.934	0.964	0.951	0.939							
X_{10}	0.968	0.965	0.968	0.962	0.963	0.958	0.948	0.937	0.972						
X_{11}	0.967	0.963	0.964	0.960	0.931	0.957	0.945	0.933	0.968	0.965					
X_{12}	0.965	0.960	0.966	0.958	0.929	0.954	0.942	0.935	0.966	0.965	0.960				
X_{13}	0.960	0.953	0.956	0.952	0.923	0.947	0.936	0.927	0.958	0.955	0.952	0.952			
X_{14}	0.952	0.946	0.948	0.942	0.916	0.941	0.931	0.918	0.951	0.949	0.948	0.944	0.939		
X_{15}	0.964	0.959	0.964	0.957	0.931	0.956	0.943	0.935	0.968	0.964	0.961	0.958	0.951	0.945	
X_{16}	0.944	0.941	0.945	0.940	0.913	0.937	0.925	0.916	0.945	0.943	0.940	0.939	0.930	0.928	0.938

Table 5.6: *Manhattan distances calculated for the Kendalls tau* τ

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}
X_2	0.190														
X_3	0.132	0.151													
X_4	0.236	0.161	0.198												
X_5	0.808	0.727	0.778	0.698											
X_6	0.306	0.225	0.268	0.184	0.657										
X_7	0.547	0.468	0.507	0.433	0.499	0.383									
X_8	0.767	0.684	0.728	0.650	0.361	0.604	0.413								
X_9	0.111	0.203	0.149	0.251	0.834	0.316	0.565	0.788							
X_{10}	0.160	0.196	0.153	0.246	0.738	0.317	0.556	0.780	0.176						
X_{11}	0.182	0.128	0.148	0.161	0.734	0.227	0.474	0.697	0.198	0.191					
X_{12}	0.229	0.150	0.192	0.140	0.701	0.195	0.441	0.652	0.245	0.233	0.158				
X_{13}	0.375	0.302	0.340	0.261	0.608	0.215	0.325	0.543	0.398	0.391	0.306	0.266			
X_{14}	0.522	0.447	0.489	0.415	0.514	0.357	0.224	0.440	0.543	0.533	0.443	0.414	0.298		
X_{15}	0.225	0.146	0.180	0.147	0.701	0.197	0.445	0.659	0.233	0.228	0.147	0.143	0.278	0.420	
X_{16}	0.643	0.560	0.596	0.519	0.440	0.468	0.287	0.344	0.661	0.649	0.567	0.529	0.420	0.301	0.537

Table 5.7: *Manhattan distances calculated for the spearmans rho* ρ

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}
X_2	0.023														
X_3	0.014	0.016													
X_4	0.032	0.019	0.025												
X_5	0.169	0.156	0.165	0.151											
X_6	0.046	0.032	0.039	0.024	0.142										
X_7	0.099	0.085	0.091	0.079	0.109	0.068									
X_8	0.155	0.141	0.148	0.134	0.074	0.126	0.085								
X_9	0.010	0.025	0.016	0.034	0.173	0.047	0.101	0.157							
X_{10}	0.019	0.025	0.019	0.035	0.156	0.049	0.101	0.158	0.022						
X_{11}	0.023	0.013	0.016	0.018	0.156	0.031	0.086	0.143	0.025	0.025					
X_{12}	0.031	0.017	0.024	0.014	0.151	0.026	0.080	0.134	0.033	0.033	0.018				
X_{13}	0.061	0.048	0.054	0.039	0.133	0.030	0.057	0.113	0.063	0.064	0.048	0.040			
X_{14}	0.093	0.080	0.087	0.074	0.112	0.062	0.035	0.092	0.096	0.096	0.079	0.074	0.050		
X_{15}	0.030	0.016	0.022	0.016	0.151	0.026	0.081	0.136	0.031	0.032	0.016	0.015	0.043	0.075	
X_{16}	0.122	0.107	0.113	0.099	0.095	0.089	0.051	0.069	0.123	0.124	0.108	0.101	0.080	0.053	0.103

Table 5.8: *Manhattan distances calculated for the tail index*

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}
X_2	0.135														
X_3	0.092	0.111													
X_4	0.169	0.115	0.145												
X_5	0.589	0.531	0.571	0.509											
X_6	0.220	0.161	0.196	0.132	0.480										
X_7	0.396	0.339	0.371	0.314	0.364	0.277									
X_8	0.558	0.498	0.533	0.473	0.263	0.440	0.301								
X_9	0.078	0.145	0.102	0.180	0.608	0.227	0.408	0.573							
X_{10}	0.114	0.140	0.107	0.176	0.538	0.228	0.403	0.567	0.126						
X_{11}	0.130	0.091	0.109	0.114	0.535	0.163	0.343	0.507	0.142	0.137					
X_{12}	0.160	0.106	0.130	0.103	0.515	0.144	0.323	0.479	0.172	0.163	0.111				
X_{13}	0.270	0.217	0.248	0.187	0.444	0.154	0.235	0.395	0.287	0.282	0.220	0.195			
X_{14}	0.377	0.324	0.357	0.300	0.375	0.258	0.161	0.320	0.393	0.386	0.321	0.303	0.215		
X_{15}	0.161	0.104	0.132	0.105	0.512	0.141	0.322	0.480	0.167	0.164	0.105	0.101	0.200	0.304	
X_{16}	0.466	0.406	0.436	0.376	0.321	0.340	0.208	0.250	0.479	0.471	0.411	0.388	0.305	0.217	0.390

Chapter 6

EMPIRICAL STUDY OF CLAIMS DATA

6.1 Introduction

This chapter is dedicated to empirical investigations on the theoretical results established in Chapter 5. The data were on aggregate losses per company in a year. We will, first, explore the data that were collected from thirty-five insurance companies which are members of the Insurance Regulatory Authority (IRA) of Kenya and participating in some class of general insurance for the period 2006 to 2009. They include:

AIG (K)	First Assuarance	Madison
Amaco	Gateway	Mayair
APA	Geminia	Mercantile
Blue Shield	General Accident	Occidental
British American	Heritage AII	Pacis
Cannon	ICEA	Phoenix
CFC Life	Intra Africa	Real
Concord	Jubilee	Tausi
Cooperative	Kenindia	The Monarch
Corporate	Kenya Orient	Trident
Directline	Kenyan Alliance	UAP Provincial
Fidelity Shield	Lion of Kenya	

Other companies like Invesco and The Standard insurance companies were excluded as they were under statutory management. A general business insurer, in Kenya, can be registered to transact any or all the twelve classes of general insurance business namely:

Aviation	Liability	Personal accident
Engineering	Marine	Theft
Fire-domestic	Motor-private	Workmens compensation
Fire-industrial	Motor-commercial	Miscellaneous

6.2 Exploration of the Empirical Data

Exploring the data gave rise to the densities in Figure 6.1 which assume similar distributions' characteristic as those for the loss distributions simulated previously in chapter five. They, therefore, capture the same distribution shapes. The individual densities were isolated, for the various business classes and are found in Figure 6.2.

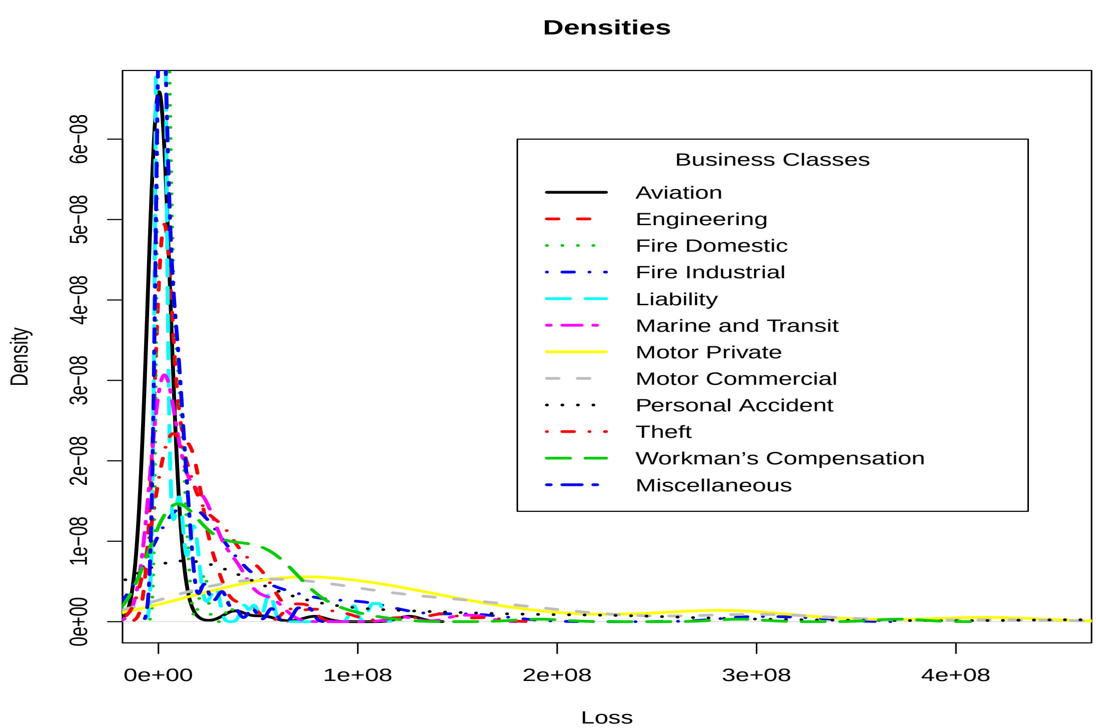


Figure 6.1: *Densities for the twelve classes of business in general insurance, Kenya*

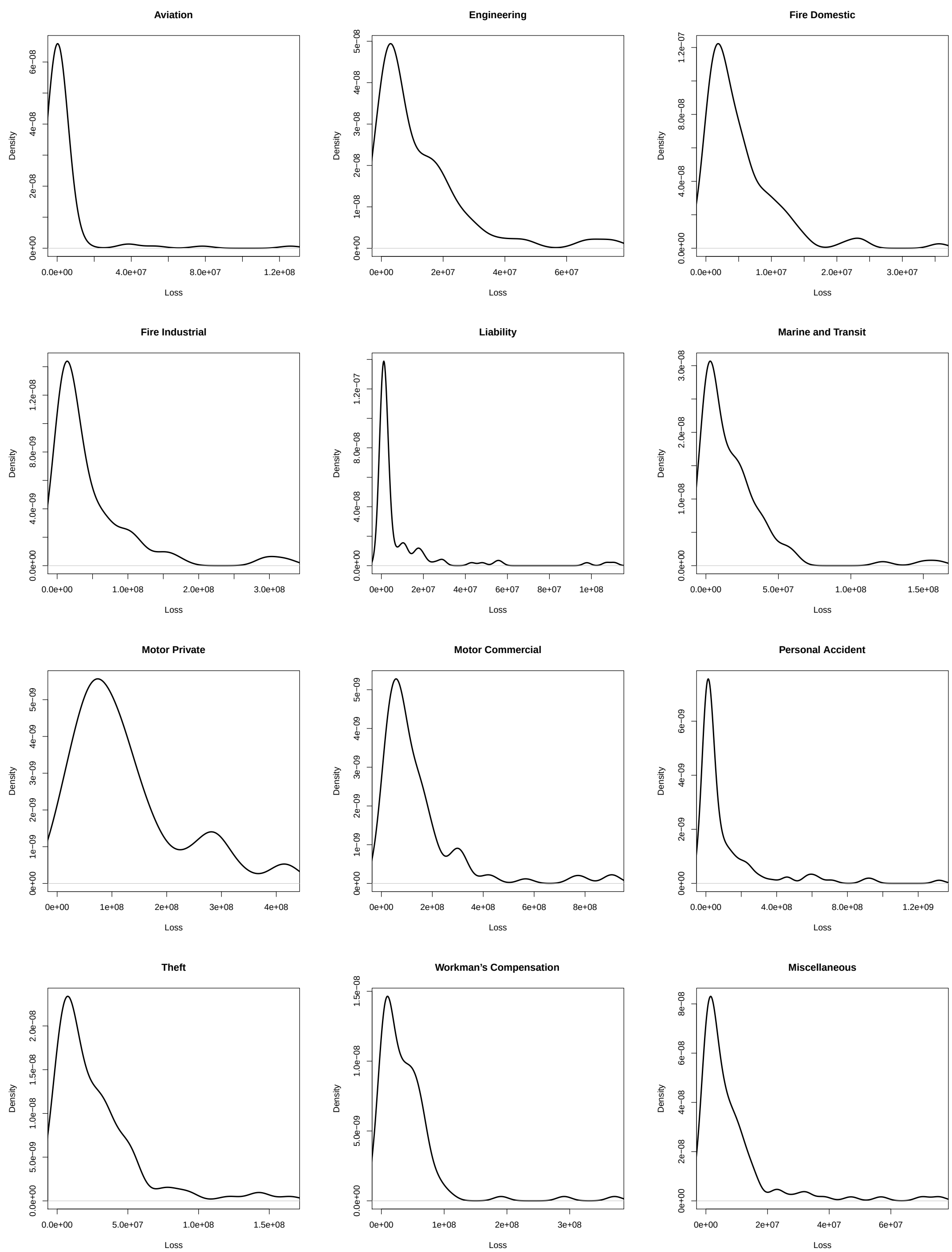


Figure 6.2: *Densities for the general insurance classes in Kenya*

6.2.1 Fitting the Copula function for each pair of business classes

As explained in Chapter five, we first fit the copula function to our data so as to estimate the dependence parameter θ. The Gumbel copulas were fitted to each pairs of business classes and θ is tabulated in Table 6.1.

The fitted dependence parameter, θ, matrix may not give all the insight to examine the dependence between two variables since $theta$ can be in the range $(-\infty, \infty)$ like in the case of the Frank copula. The re-scaling of this dependence parameter into a definite scale such as that for Kendall's tau or Spearman's rho $[-1, 1]$ is given in the next sub-section.

Table 6.1: *The dependence parameter θ estimated for each pair of general insurance classes*

	Av	Eng	F D	F I	Liab	M & T	M P	M C	P A	Theft	W C
Eng	5.814										
F D	6.025	6.696									
F I	6.173	6.906	6.859								
Liab	5.770	6.508	6.535	7.097							
M & T	5.574	6.171	6.293	6.654	6.580						
M P	6.961	7.212	8.219	8.033	7.643	7.085					
M C	8.068	8.280	8.764	8.945	8.438	8.112	10.102				
P A	6.153	6.815	7.124	7.079	6.832	6.668	8.210	8.814			
Theft	6.451	6.957	7.053	7.322	6.911	6.712	7.936	8.536	6.913		
W C	6.243	6.976	7.265	7.338	7.170	6.727	8.296	9.260	7.408	7.187	
Misc	5.496	6.398	6.566	6.410	6.664	5.917	7.474	7.864	6.669	6.568	6.681

KEY: Av - Aviation, Eng - Engineering, F D - Fire Domestic, F I - Fire Industrial, Liab - Liability, M & T - Marine and Transit, M P - Motor Private, M C - Motor Commercial, P A - Personal Accident, W C - Workman's Compensation, Misc - Miscellaneous.

6.2.2 Calculating dependence the measures

The measures of dependence, Kendall's τ, Spearman's ρ and the Tail dependence are now calculated from their relationship with θ and are tabulated in Tables 6.2, Table 6.3 and Table 6.4 respectively. These are bounded at $+1$ and so the closer the quantity is to one, the more dependence there is between the two lines in question. In order to cluster the various LOB, with respect to their dependence, we need a criterion of telling how close any two lines are. This is done by calculating the distances between the dependence parameters.

6.2.3 Calculating distance matrices

Euclidean and the Manhattan distances for every pair of insurance classes were calculated for the measures of dependence in readiness of clustering process and are found in Tables 6.5, 6.6, 6.7, 6.8, 6.9 and 6.10. These distance matrices will now be used to cluster the LOB but not before comparing their performances in the clustering exercise.

Table 6.2: *Kendall's τ calculated from θ for each pair of general insurance classes*

	Av	Eng	F D	F I	Liab	M & T	M P	M C	P A	Theft	W C
Eng	0.828										
F D	0.834	0.851									
F I	0.838	0.855	0.854								
Liab	0.827	0.846	0.847	0.859							
M & T	0.821	0.838	0.841	0.850	0.848						
M P	0.856	0.861	0.878	0.876	0.869	0.859					
M C	0.876	0.879	0.886	0.888	0.881	0.877	0.901				
P A	0.837	0.853	0.860	0.859	0.854	0.850	0.878	0.887			
Theft	0.845	0.856	0.858	0.863	0.855	0.851	0.874	0.883	0.855		
W C	0.840	0.857	0.862	0.864	0.861	0.851	0.879	0.892	0.865	0.861	
Misc	0.818	0.844	0.848	0.844	0.850	0.831	0.866	0.873	0.850	0.848	0.850

KEY: Av - Aviation, Eng - Engineering, F D - Fire Domestic, F I - Fire Industrial, Liab - Liability, M & T - Marine and Transit, M P - Motor Private, M C - Motor Commercial, P A - Personal Accident, W C - Workman's Compensation, Misc - Miscellaneous.

Table 6.3: *Spearman's ρ calculated from θ for each pair of general insurance classes*

	Av	Eng	F D	F I	Liab	M & T	M P	M C	P A	Theft	W C
Eng	0.958										
F D	0.960	0.968									
F I	0.963	0.970	0.969								
Liab	0.957	0.966	0.966	0.971							
M & T	0.954	0.962	0.964	0.968	0.967						
M P	0.970	0.972	0.978	0.978	0.975	0.971					
M C	0.978	0.979	0.981	0.982	0.980	0.978	0.986				
P A	0.962	0.969	0.971	0.971	0.969	0.968	0.978	0.981			
Theft	0.965	0.970	0.971	0.973	0.970	0.968	0.977	0.980	0.970		
W C	0.963	0.970	0.973	0.973	0.972	0.968	0.979	0.983	0.974	0.972	
Misc	0.953	0.965	0.967	0.965	0.968	0.959	0.974	0.977	0.968	0.967	0.968

KEY: Av - Aviation, Eng - Engineering, F D - Fire Domestic, F I - Fire Industrial, Liab - Liability, M & T - Marine and Transit, M P - Motor Private, M C - Motor Commercial, P A - Personal Accident, W C - Workman's Compensation, Misc - Miscellaneous.

Table 6.4: *Tail index calculated from θ for each pair of general insurance classes*

	Av	Eng	F D	F I	Liab	M & T	M P	M C	P A	Theft	W C
Eng	0.873										
F D	0.878	0.891									
F I	0.881	0.894	0.894								
Liab	0.872	0.888	0.888	0.897							
M & T	0.868	0.881	0.884	0.890	0.889						
M P	0.895	0.899	0.912	0.910	0.905	0.897					
M C	0.910	0.913	0.918	0.919	0.914	0.911	0.929				
P A	0.881	0.893	0.898	0.897	0.893	0.890	0.912	0.918			
Theft	0.887	0.895	0.897	0.901	0.895	0.891	0.909	0.915	0.895		
W C	0.883	0.896	0.900	0.901	0.899	0.891	0.913	0.922	0.902	0.899	
Misc	0.866	0.886	0.873	0.886	0.890	0.876	0.903	0.908	0.890	0.889	0.891

KEY: Av - Aviation, Eng - Engineering, F D - Fire Domestic, F I - Fire Industrial, Liab - Liability, M & T - Marine and Transit, M P - Motor Private, M C - Motor Commercial, P A - Personal Accident, W C - Workman's Compensation, Misc - Miscellaneous.

Table 6.5: *Euclidean distances from Kendall's τ for each pair of general insurance classes*

	Av	Eng	F D	F I	Liab	M & T	Misc	M C	M P	P A	Theft
Eng	0.249										
F D	0.244	0.212									
F I	0.241	0.207	0.207								
Liab	0.253	0.218	0.217	0.200							
M & T	0.256	0.230	0.228	0.217	0.217						
Misc	0.260	0.222	0.218	0.225	0.214	0.239					
M C	0.237	0.205	0.189	0.181	0.198	0.217	0.221				
M P	0.237	0.209	0.181	0.183	0.197	0.219	0.209	0.146			
P A	0.242	0.210	0.199	0.200	0.208	0.217	0.216	0.183	0.179		
Theft	0.232	0.205	0.202	0.194	0.206	0.216	0.220	0.187	0.185	0.205	
W C	0.242	0.206	0.196	0.193	0.200	0.217	0.218	0.171	0.175	0.191	0.197

KEY: Av - Aviation, Eng - Engineering, F D - Fire Domestic, F I - Fire Industrial, Liab - Liability, M & T - Marine and Transit, M P - Motor Private, M C - Motor Commercial, P A - Personal Accident, W C - Workman's Compensation, Misc - Miscellaneous.

Table 6.6: *Manhattan distances from Kendall's τ for each pair of general insurance classes*

	Av	Eng	F D	F I	Liab	M & T	Misc	M C	M P	P A	Theft
Eng	0.492										
F D	0.531	0.351									
F I	0.554	0.371	0.337								
Liab	0.524	0.348	0.356	0.346							
M & T	0.455	0.380	0.423	0.434	0.385						
Misc	0.472	0.377	0.408	0.440	0.377	0.376					
M C	0.751	0.596	0.532	0.497	0.563	0.653	0.656				
M P	0.666	0.507	0.423	0.398	0.463	0.564	0.544	0.322			
P A	0.553	0.375	0.316	0.316	0.344	0.432	0.426	0.502	0.394		
Theft	0.540	0.369	0.341	0.307	0.347	0.432	0.433	0.507	0.400	0.323	
W C	0.582	0.400	0.338	0.310	0.364	0.463	0.460	0.457	0.357	0.304	0.321

KEY: Av - Aviation, Eng - Engineering, F D - Fire Domestic, F I - Fire Industrial, Liab - Liability, M & T - Marine and Transit, M P - Motor Private, M C - Motor Commercial, P A - Personal Accident, W C - Workman's Compensation, Misc - Miscellaneous.

Table 6.7: *Euclidean distances from Spearman's ρ for each pair of general insurance classes*

	Av	Eng	F D	F I	Liab	M & T	Misc	M C	M P	P A	Theft
Eng	0.064										
F D	0.063	0.046									
F I	0.062	0.044	0.044								
Liab	0.067	0.048	0.048	0.042							
M & T	0.067	0.054	0.054	0.049	0.048						
Misc	0.069	0.051	0.049	0.053	0.047	0.058					
M C	0.072	0.053	0.046	0.042	0.049	0.060	0.060				
M P	0.066	0.049	0.038	0.037	0.044	0.055	0.051	0.025			
P A	0.063	0.045	0.041	0.041	0.044	0.049	0.049	0.042	0.036		
Theft	0.059	0.044	0.042	0.038	0.044	0.049	0.051	0.043	0.038	0.043	
W C	0.064	0.045	0.040	0.038	0.042	0.050	0.050	0.037	0.033	0.037	0.040

KEY: Av - Aviation, Eng - Engineering, F D - Fire Domestic, F I - Fire Industrial, Liab - Liability, M & T - Marine and Transit, M P - Motor Private, M C - Motor Commercial, P A - Personal Accident, W C - Workman's Compensation, Misc - Miscellaneous.

Table 6.8: *Manhattan distances from Spearman's ρ for each pair of general insurance classes*

	Av	Eng	F D	F I	Liab	M & T	Misc	M C	M P	P A	Theft
Eng	0.151										
F D	0.166	0.085									
F I	0.175	0.093	0.081								
Liab	0.164	0.085	0.087	0.085							
M & T	0.136	0.099	0.116	0.120	0.101						
Misc	0.143	0.098	0.110	0.124	0.098	0.099					
M C	0.246	0.177	0.152	0.138	0.164	0.201	0.202				
M P	0.216	0.145	0.113	0.102	0.127	0.170	0.162	0.073			
P A	0.175	0.095	0.072	0.071	0.083	0.120	0.118	0.139	0.100		
Theft	0.170	0.094	0.082	0.067	0.084	0.120	0.121	0.141	0.102	0.074	
W C	0.186	0.106	0.081	0.069	0.091	0.132	0.131	0.122	0.086	0.066	0.072

KEY: Av - Aviation, Eng - Engineering, F D - Fire Domestic, F I - Fire Industrial, Liab - Liability, M & T - Marine and Transit, M P - Motor Private, M C - Motor Commercial, P A - Personal Accident, W C - Workman's Compensation, Misc - Miscellaneous.

Table 6.9: *Euclidean distances from Tail index for each pair of general insurance classes*

	Av	Eng	F D	F I	Liab	M & T	Misc	M C	M P	P A	Theft
Eng	0.183										
F D	0.178	0.156									
F I	0.178	0.151	0.151								
Liab	0.187	0.159	0.160	0.146							
M & T	0.189	0.169	0.167	0.159	0.159						
Misc	0.192	0.164	0.181	0.166	0.157	0.176					
M C	0.176	0.151	0.141	0.132	0.145	0.160	0.167				
M P	0.175	0.153	0.134	0.133	0.144	0.161	0.157	0.105			
P A	0.178	0.153	0.146	0.146	0.152	0.159	0.160	0.134	0.130		
Theft	0.171	0.150	0.148	0.141	0.150	0.158	0.163	0.137	0.135	0.149	
W C	0.179	0.151	0.144	0.141	0.145	0.159	0.161	0.125	0.127	0.139	0.144

KEY: Av - Aviation, Eng - Engineering, F D - Fire Domestic, F I - Fire Industrial, Liab - Liability, M & T - Marine and Transit, M P - Motor Private, M C - Motor Commercial, P A - Personal Accident, W C - Workman's Compensation, Misc - Miscellaneous.

Table 6.10: *Manhattan distances from Tail index for each pair of general insurance classes*

	Av	Eng	F D	F I	Liab	M & T	Misc	M C	M P	P A	Theft
Eng	0.368										
F D	0.382	0.267									
F I	0.415	0.273	0.257								
Liab	0.392	0.256	0.277	0.255							
M & T	0.339	0.281	0.303	0.322	0.284						
Misc	0.346	0.293	0.333	0.342	0.293	0.283					
M C	0.564	0.444	0.411	0.368	0.419	0.488	0.505				
M P	0.500	0.377	0.328	0.293	0.343	0.421	0.421	0.236			
P A	0.414	0.276	0.246	0.231	0.253	0.320	0.331	0.372	0.291		
Theft	0.404	0.272	0.266	0.225	0.255	0.320	0.337	0.376	0.295	0.237	
W C	0.437	0.296	0.264	0.227	0.268	0.344	0.357	0.338	0.263	0.222	0.235

KEY: Av - Aviation, Eng - Engineering, F D - Fire Domestic, F I - Fire Industrial, Liab - Liability, M & T - Marine and Transit, M P - Motor Private, M C - Motor Commercial, P A - Personal Accident, W C - Workman's Compensation, Misc - Miscellaneous.

Comparisons between the performance of the Manhattan and Euclidean distances, as a criterion to aid in clustering of risk classes is shown in Table 6.11, and the Euclidean distance performed better than the Manhattan due to the fact that it has greater values of the Mantel statistic.

Table 6.11: *Comparison between the performance of the Euclidean and the Manhattan distances on the classes of general insurance*

	Distance			
	Manhattan		Euclidean	
Measure	Mantel statistic	Significance	Mantel statistic	Significance
Kendall's tau	0.658	<0.01	0.9029	<0.01
Spearman's Rho	0.7041	<0.01	0.785	<0.01
Tail index	0.6511	<0.01	0.9067	<0.01

6.2.4 Clustering the insurance business classes

Dendrograms are the tools we use for the clustering exercise as found in Figures 6.3, through 6.8. Considering the Kendall's tau, Figure 6.3 used the Manhattan distance while Figure 6.4 employs the Euclidean distance. The Spearman's rho we have Figure 6.5 and Figure 6.6 while the Tail index is used in Figures 6.7 and 6.8. The relationships from any of the two distances are consistently similar for the three measures of dependence. The top panel of each of these figures contains the *'crude'* classification while the bottom one represents a case of five LOB. It is important to note that if we want fewer LOB, we only need to follow the dendrogram to a new level.

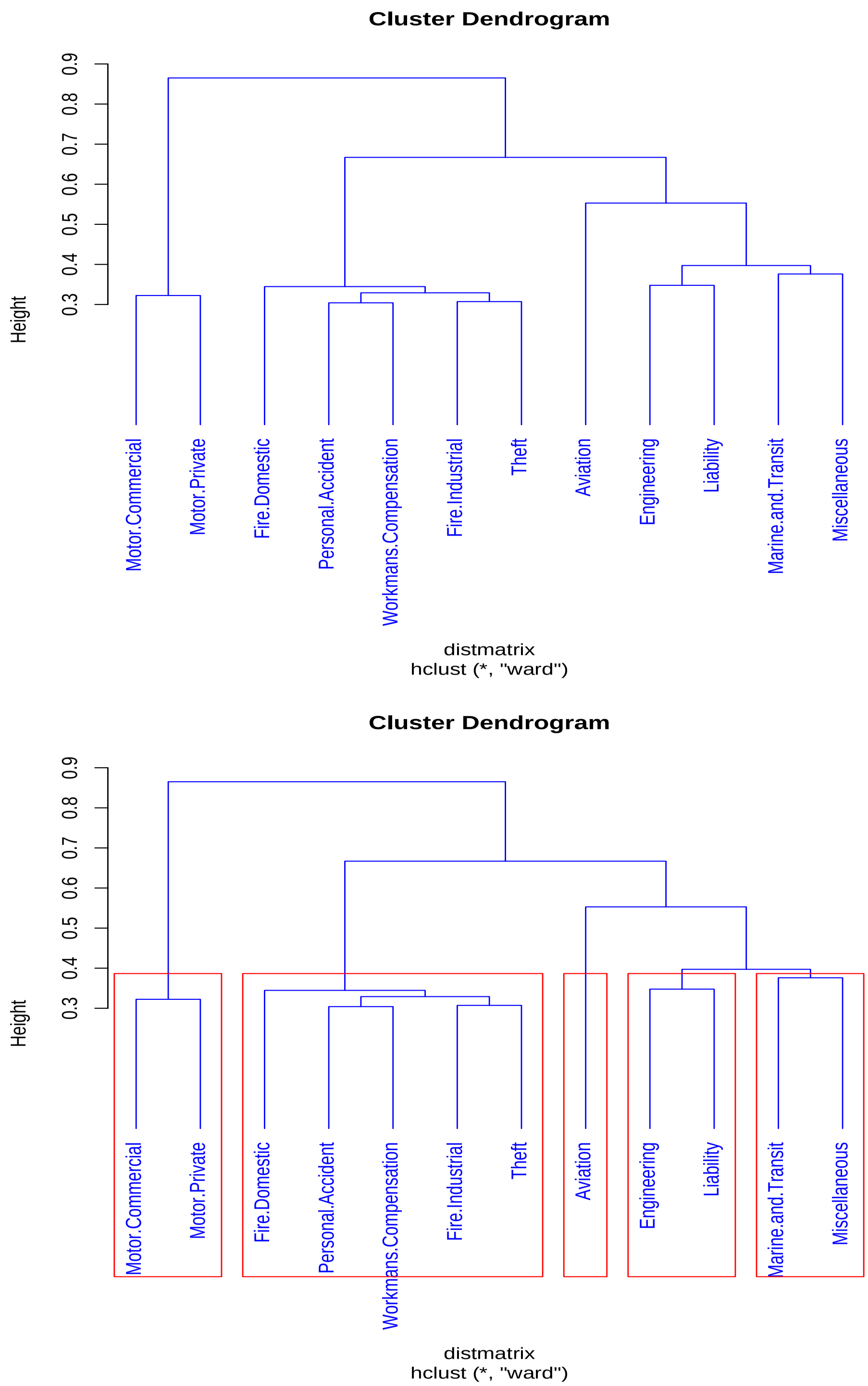
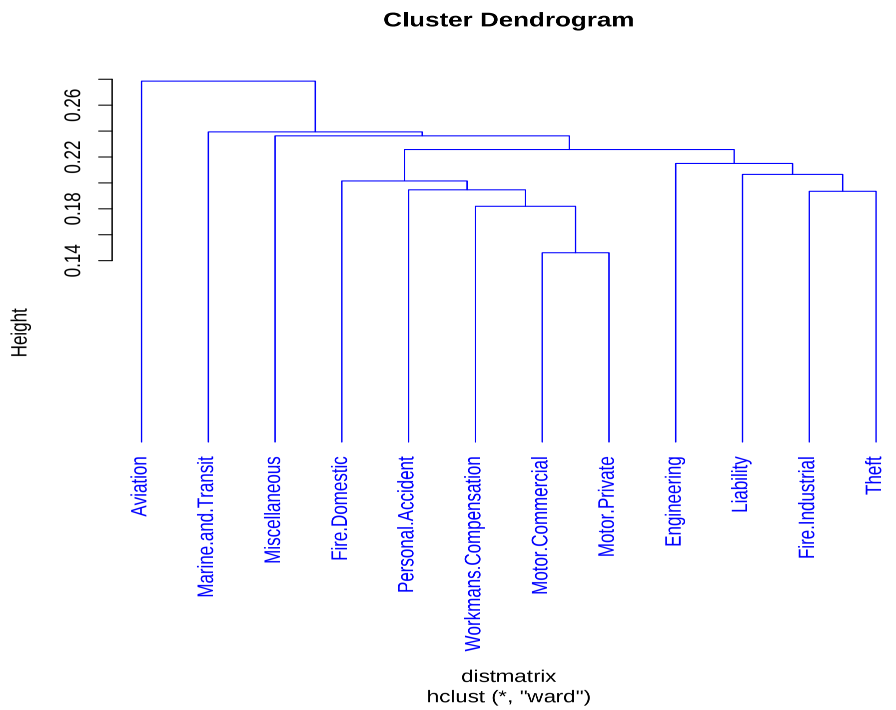


Figure 6.3: *The general insurance classes clustered by the Manhattan distances for the Kendall's* τ



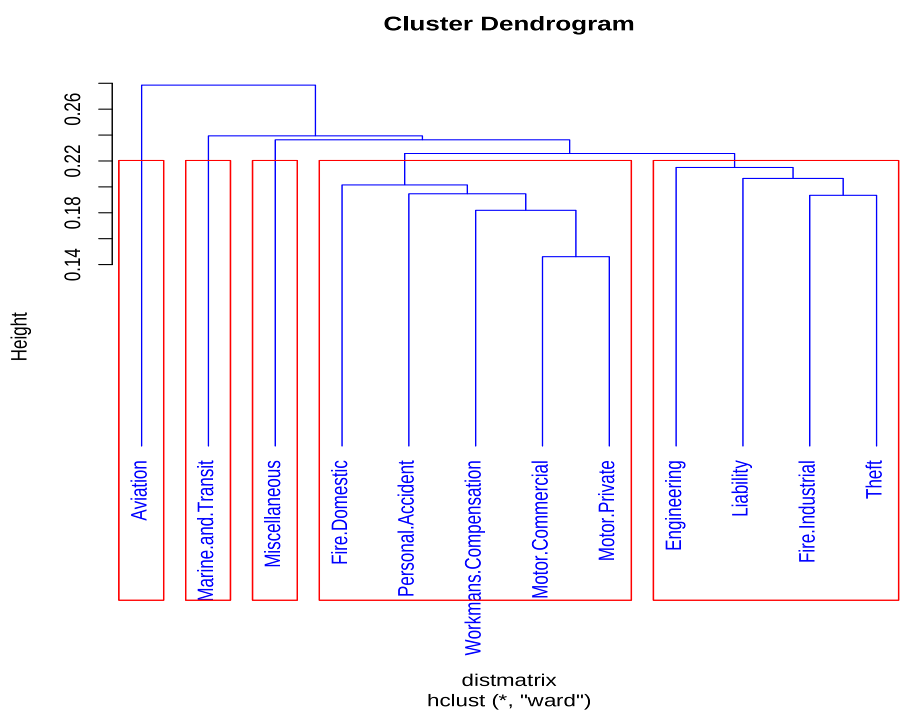
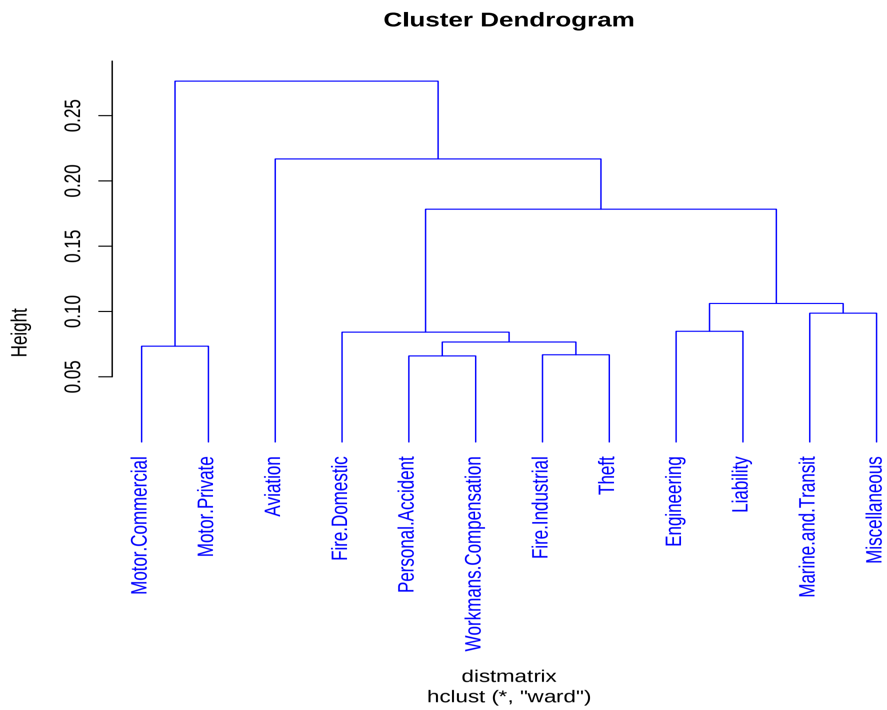


Figure 6.4: *The general insurance classes clustered by the Euclidean distances for the Kendall's* τ



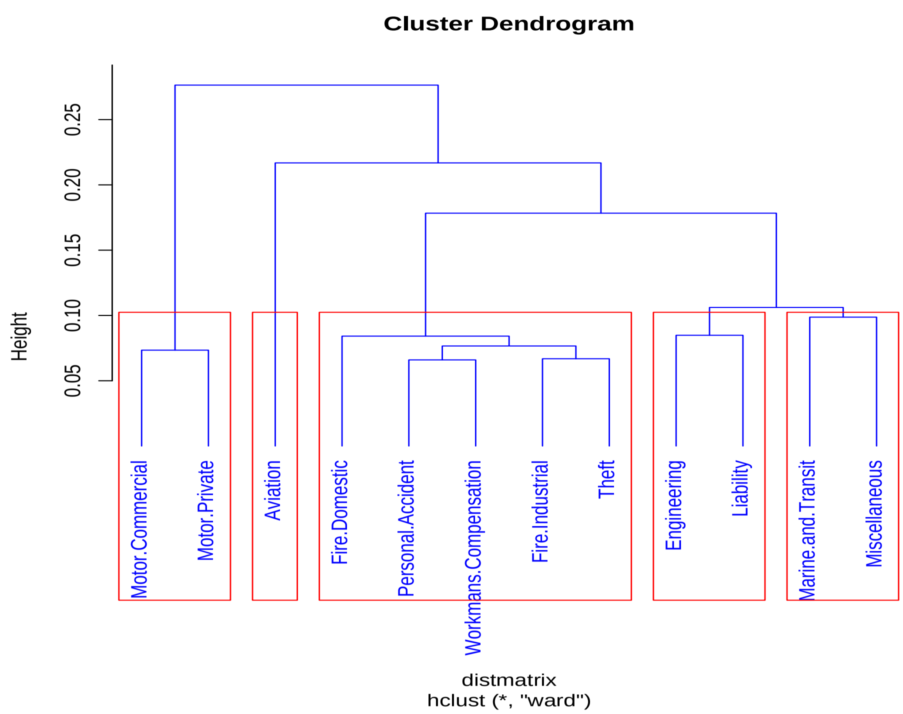


Figure 6.5: *The general insurance classes clustered by the Manhattan distances for the Spearman's* ρ

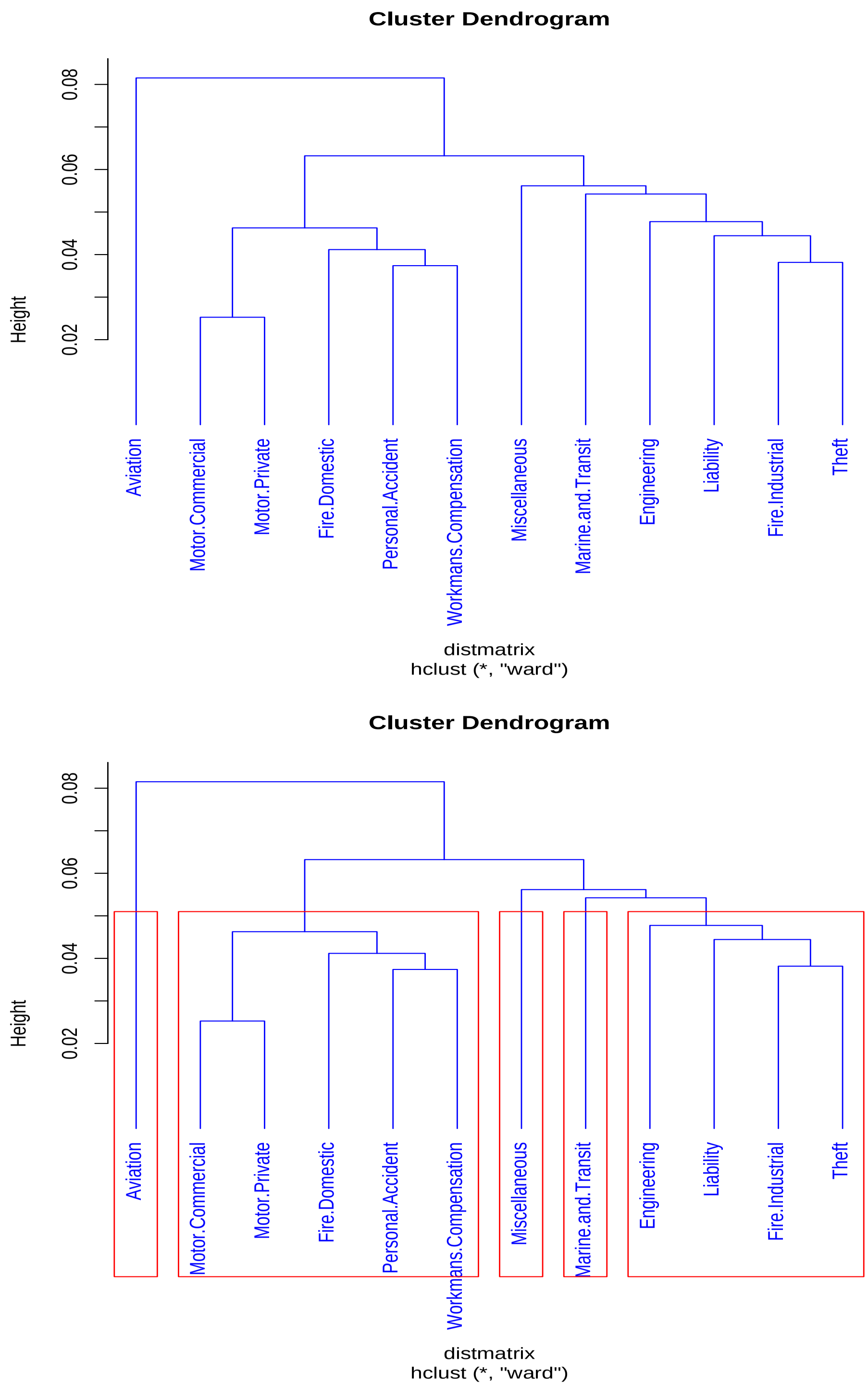


Figure 6.6: *The general insurance classes clustered by the Euclidean distances for the Spearman's ρ*

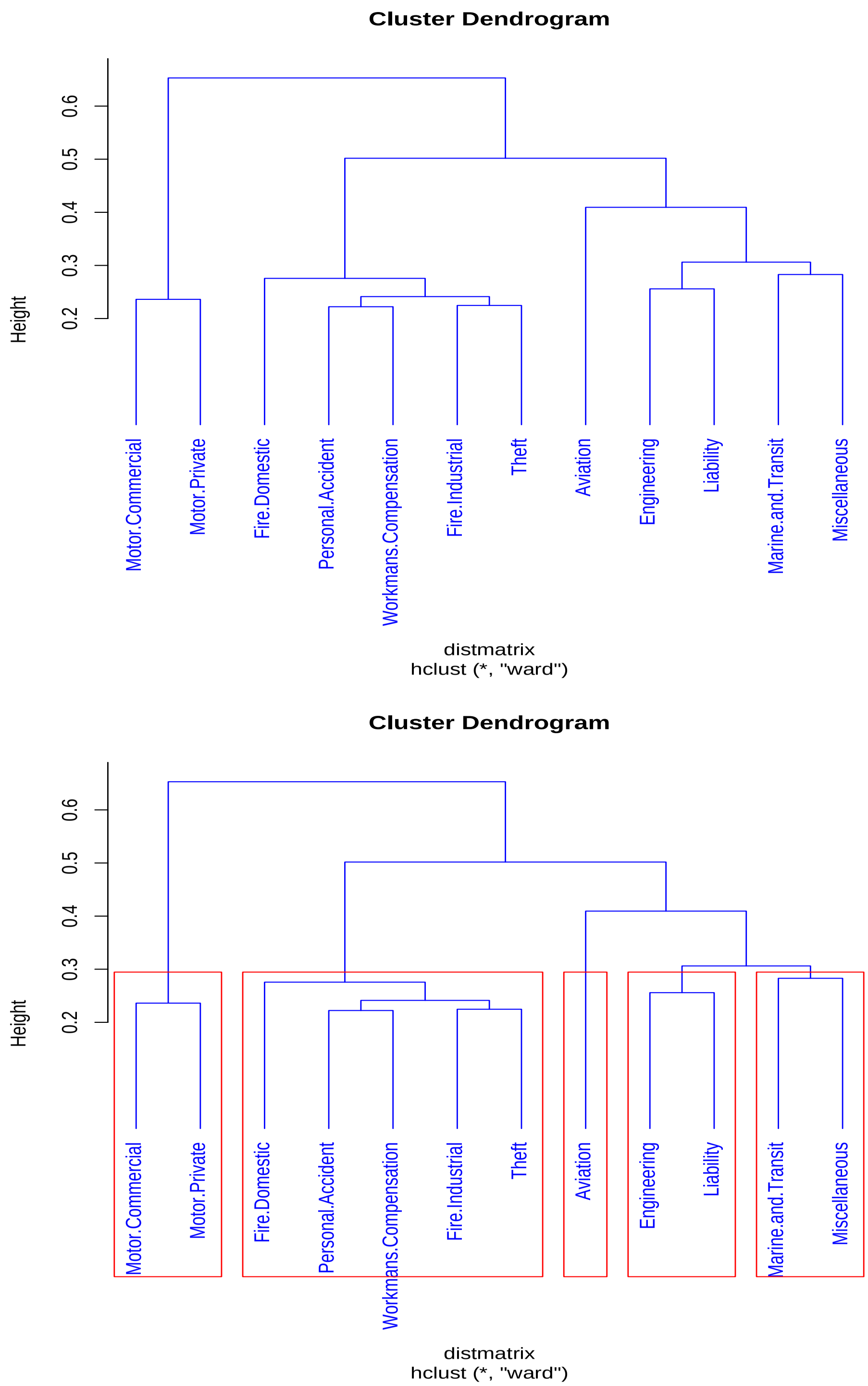


Figure 6.7: *The general insurance classes clustered by the Manhattan distances for the Tail dependence*

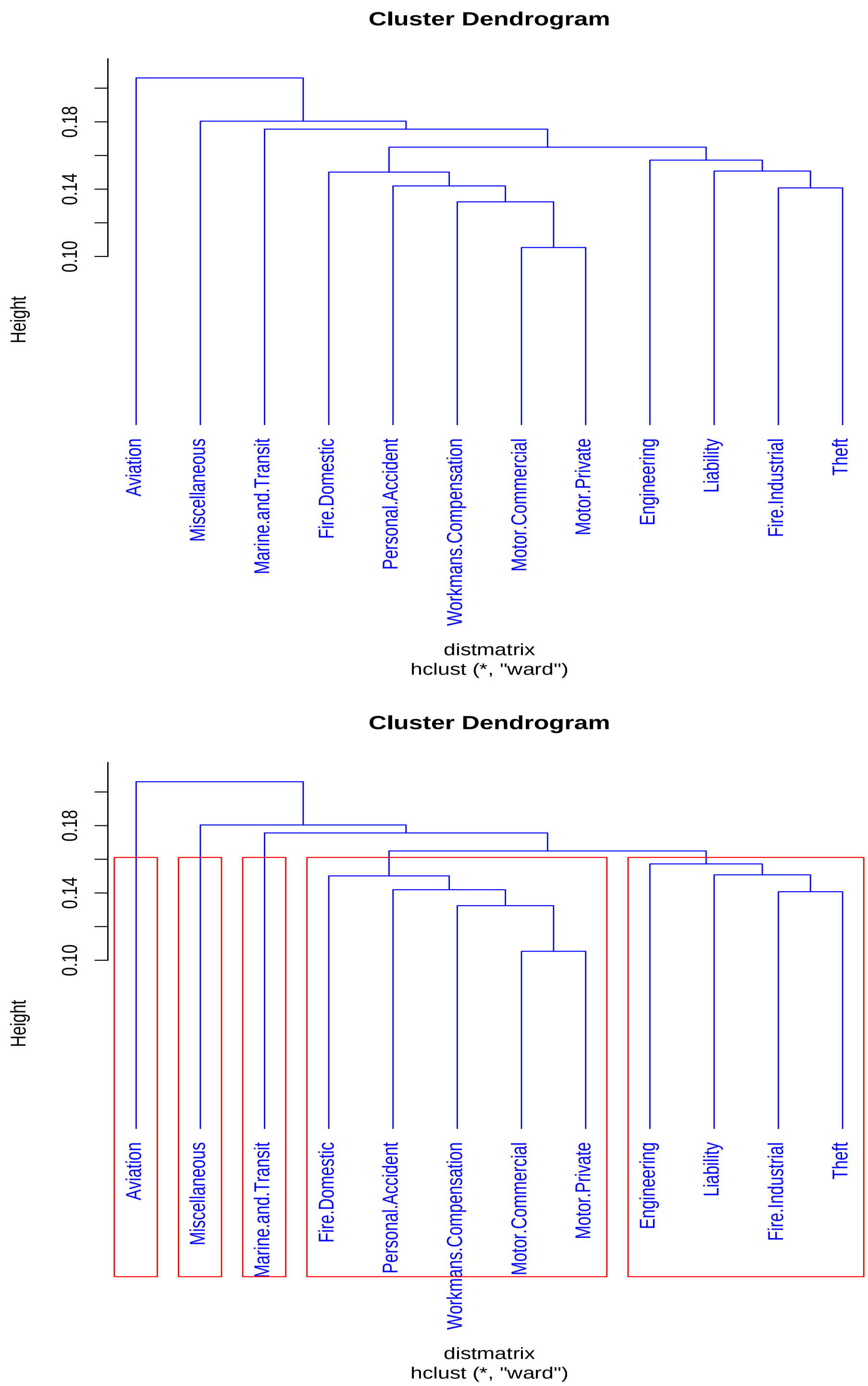


Figure 6.8: *The general insurance classes clustered by the Euclidean distances for the Tail dependence*

Discussion

The calculated Spearman's ρ quantities are consistently closer together than for comparable Kendall's τ. For instance, consider the Euclidean distance between the Aviation and the Engineering lines. The distance for the tau is 0.249 while that of the rho is 0.064. They are consistent as they produce the same clustering structure as evident from Figures 6.4 and 6.6. This is also true when you consider the Manhattan distances whereby for τ we have 0.492 and a ρ of 0.151 for the Aviation Vs Engineering. The cluster structures are the same as well.

This work proposed the use of the upper tail dependence derived from the dependence parameter in determining the retention limits for a re-insurance arrangement. Though the dependence is not the only factor to consider for such re-insurance treaties the forwarding proportions should be some where proportional to ${}^{1}\!/_{(1-\text{Tail index})}$. This will ensure that for highly dependent risks in the upper tail will forward higher proportions to the re-insurer and vice versa. The behaviour of this proposed quantity is found in Figure 6.9.

Five major classes stand out each with peculiar characteristics. The first cluster involves the rare but with a high probability of a huge claim amount lines: Engineering, Liability, Fire industrial and Theft. The second cluster contain lines with moderate

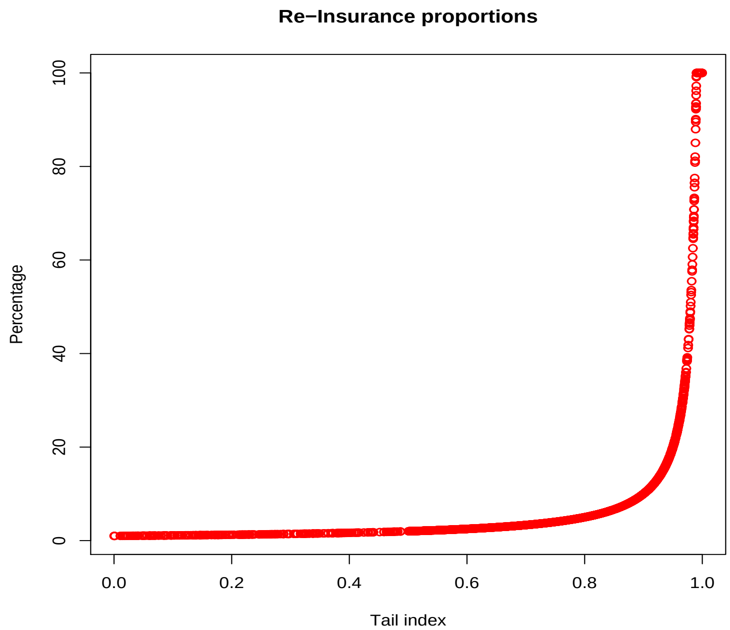


Figure 6.9: *The proposed re-insurance proportions in relation to the tail index*

claim amounts as compared to the previous cluster but rather slightly more frequent: Fire domestic, Personal accident, Workman's compensation, Motor commercial and Motor private. In the next cluster we have the less popular lines under the umbrella of the Miscellaneous class. Marine and Transit which is completely erratic clusters alone while the Aviation line whose main business is exported to foreign countries forming the last cluster. Finally, it can be remarked that the choice of distance to apply is crucial and that with the dendrograms one can choose the number of efficient divisions quite easily.

Chapter 7

CONCLUSION AND RECOMMENDATIONS

7.1 Conclusion

It can be concluded that the dependence structure of various lines of business cannot be ignored especially in rate making and other computations in the insurance industry. Copulas allow for the inclusion of features such as fat tails and skewness for nonelliptically distributed risks. Copulas model the dependence structure as well as isolating the marginal distributions' characteristics. The choice of distance, for use in the clustering of risks, is crucial and depends on the problem at hand. This means that different distances will perform differently for the same task. The Mantel statistic is the best choice to measure the strength of the relationship between the cophenetic matrix and the distance matrix and consequently aiding in choosing the appropriate distance to use. Cophenetic distance matrix contains pairwise distances among all entities. This is because the Mantel test evaluates correlation between distance (or similarity or correlation or dissimilarity) matrices. Therefore if the correlation between the cophenetic matrix and the distance matrix is high and significant, then the relationships obtained are not by mere chance. It can be seen that when the Mantel statistic is low, the clustering criterion is weaker and one may miss out on some important dependence information. The strength of the relationship, from the Mantel statistic, dictates the strength of the clustering criterion. Sufficient data exploration will reduce the uncertainty of choosing between the copula functions to fit. The Gumbel is the best choice in the family of the Archimedean copulas, owing to its strength in capturing the upper tail dependence, for loss data distributions have *'active'* upper tails since the other Archimedean copulas (that is, Frank and Clayton) would under estimate the dependence in the upper tail as they have no upper tail dependence. The business classes in one line are expected to have strong dependence than those in different lines if a proper market research precedes their establishment in order to enjoy diversification benefits. The calculated Spearman's rho from the dependence parameter are consistently closer together than for comparable Kendau's tau. We proposed the use of the upper tail dependence derived from the dependence parameter in determining the retention limits for a re-insurance arrangement. Though the dependence is not the only factor to consider for such re-insurance treaties the forwarding proportions should be somewhere proportional to $^1\!/_{(1\,-\,\text{Tail index})}$. This ensures that for highly dependent risks in the upper tail will forward higher proportions to the re-insurer and vice versa.

Considering the data for the Kenyan insurance sector, five major classes stand out each with unique characteristics. The first cluster involves the rare but with a high probability of a huge claim amount lines: Engineering, Liability, Fire industrial and Theft. The second cluster contain lines with moderate claim amounts as compared to the first cluster but rather slightly more frequent: Fire domestic, Personal accident, Workman's compensation, Motor commercial and Motor private. In the third cluster we have the less popular lines under the umbrella of the Miscellaneous class. Marine and Transit which is completely erratic clusters alone while the Aviation line whose main business is exported to foreign countries forming a cluster alone.

7.2 Recommendations

We recommend to investors seeking to establish general insurance business to ,first, consider the dependence structure so as to arrive at a diversified portfolio in order to benefit from diversification benefits. The business classes that form their own individual clusters like the Aviation, Miscellaneous, Marine and Transit should be given special attention when a company engages in them as they present peculiar characteristics within themselves. We do also recommend that the insurance regulator uses the methods outlined in this thesis in order to compute the dependencies between insurance classes for advisory purposes. This is due to the fact that there may be no single insurance company that operates all the insurance classes for it to have sufficient data. Finally, the proposed algorithm is long and tedious (see the Appendices) but this can be made easier by having dedicated computer software.

7.3 Further Research

There is need to come up with a further criterion, possibly which derives some *indicators*, that can help us answer the questions concerning: how to determine the need for split, at what point can you split a risk and when do you stop splitting risks.

REFERENCES

Alfonsi, A. and Brigo, D. (2005). New families of copulas based on periodic functions. *Comm. Stat. Theory and Methods*, **34(7)**, 1437–1447.

Bowers, N., Gerber, H.U.and Hickman, J., Jones, D. and Nesbitt, C. (1997). *Actuarialmathematics*. Society of Actuaries, Itasca, Illinois.

Burgi, R. B., Dacorogna, M. M. and Iles, R. (2008). *Risk Aggregation, Dependence Structure and Diversification Benefit*. Stress Testing for Financial Institutions, http://ssrn.com/=1468526.

Embrechts, P., Lindskog, F. and McNeil, A. (2003). Modelling dependence with copulas and applications to risk management. In Rachev, S., editor, *Handbook of Heavy Tailed Distributions in Finance*, pages 329–384. Elsevier, North-Holland, Amsterdan.

Embrechts, P., McNeil, A. and Straumann, D. (2002). Correlation and dependence in risk management: Properties and pitfalls. In H., D. M. A., editor, *Risk Management: Value at Risk and Beyond*, pages 176–223. Cambridge University Press, Cambridge.

Faivre, F. (2003). *Copula: A New Vision for Economic Capital and Application to a Four Line of Business Company*. Paper presented at ASTIN Colloquium, Berlin.

Filler, G., Odening, M., Okhrin, O. and Xu, W. (2009). *On the Systemic Nature of Weather Risk*. SFB 649 Discussion Papers, Humboldt University, Berlin, Germany.

Frees, E. W. and Valdez, E. A. (2008). Hierarchical insurance claims modeling. *Journal of the American Statistical Association*, **103(484)**, 1457–1469.

Frees, E. W. and Wang, P. (2004). *Credibility Using Copulas*. University of Wisconsin, Madison.

Frees, E. and Valdez, E. (1997). *Understanding Relationships using copulas*. 32nd Actuarial Research Conference, University of Calgary.

Gatzert, N., Schmeiser, H. and Schuckmann, S. (2007). *Enterprise Risk Management in Finacial Groups: Analysis of Risk Concentration and Default Risk*. Working Papers on Risk Management and Insurance N.o 35, University of St. Gallen, St. Gallen, Switzerland.

Genest, C. and Rivest, L. P. (1993). Statistical inference procedures for bivariate archimedean copulas. *Journal of the American Statistical Association*, **88**, 1034–1043.

Genest, C. and Rivest, L. P. (2001). On the multivariate probability integral transformation. *Statistics & Probability Letters*, **53**, 391–399.

Goorbergh, R., Genest, C. and Werker, B. J. (2005). Bivariate option pricing using dynamic copula models. *Insurance: Mathematics and Economics*, **37**, 101–114.

Hipp, C. (2007). *Dependence concepts in Finance and Insurance: Copulas.* Karlsruhe University.

Kang, L. (2007). *Modeling the Dependence Structure between Bonds and Stocks: A Multidimensional Copula Approach.* Indiana University, Bloomington.

Liu, W. (2006). *Currencies Portfolio Return: A Copula Methodology.* Department of Economics, University of Toronto, Toronto.

Nelsen, R. B. (2006). *An Introduction to Copulas.* Springer Verlag, New York.

Pettere, G. and Kollo, T. (2006). *Modelling claim size time via copulas.* Transactions of 28th International Congress of Actuaries, N 206, 2006.

R Development Core Team (2009). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. URL `http://www.R-project.org`. ISBN 3-900051-07-0.

Romano, C. (2002). *Calibrating and simulating copula functions: An application to the Italian stock market.* Working Paper of CIDEM, Universit Degli Studi di Roma 12.

Savu, C. and Trede, M. (2006). *Hierarchical Archimedean Copulas.* Working Paper, University of Mnster.

Sklar, A. (1959). Fonctions de rpartition n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Universite de Paris*, **8**, 229–231.

Straub, E. (1997). *Non-Life insurance Mathematics.* Association of Swiss Actuaries, Berlin.

Tasfack, G. (2006). *Dependence Structure and Extreme Comovements in International Equity and Bond Markets.* Working paper, Universite de Montreal, CIRANO and CIREQ.

Walters, M. A. (1981). Risk classification standards. *Proceedings*, **"LXVIII, Part 1"**, No. 129.

Whelan, N. (2004). Sampling from archimedean copulas. *Quantitative Finance*, **4**, 339–352.

Wu, F., Valdez, E. A. and Sherris, M. (2006). *Simulating exchangeable multivariate Archimedean copulas and its applications*. Working Paper, School of Actuarial Studies, university of New South Wales.

Wu, F., Valdez, E. and Sherris, M. (2007). Simulating from exchangeable archimedean copulas. *Communications in Statistics: Simulation and Computation*, **36(5)**, 1019–1034.

APPENDICES

We now present selected R programs used for simulations and real data analyses in this thesis.

A.1 Simulation of loss distributions' codes

The Loss Distributions

```
library(copula)
library(actuar)
set.seed(1)
n=1000
#logNormal distribution
x1 = (rlnorm(n, meanlog=1.9, sdlog=exp(-1)))*100000
x2 = (rlnorm(n, meanlog=2.4, sdlog=exp(-0.6)))*100000
x3 = (rlnorm(n, meanlog=2.4, sdlog=exp(-.8)))*100000
x4 = (rlnorm(n, meanlog=1.9, sdlog=exp(-0.6)))*100000
#the Burr distribution
x5=(rburr(n, 3, 1.5, rate =5 , scale = 2.5))*100000
x6=(rburr(n, 1.5, 3, rate =5 , scale = 2.5))*100000
x7=(rburr(n, 1.5, 2, rate =5 , scale = 2.5))*100000
x8=(rburr(n, 2, 1.5, rate =10 , scale = 5))*100000
#single parameter Pareto or the European Pareto
x9=rpareto1(n, 3.2, 100000)
x10=rpareto1(n, 2.6, 100000)
x11=rpareto1(n, 2.1, 100000)
x12=rpareto1(n, 1.8, 100000)
#inverse of the Weibull distribution
x13=(rinvweibull(n, 1.8, rate =5, scale = 2.5))*100000
x14=(rinvweibull(n, 1.5, rate =5, scale = 2.5))*100000
x15=(rinvweibull(n, 2.1, rate =10, scale = 5))*100000
x16=(rinvweibull(n, 1.5, rate =10, scale = 5))*100000
```

Exploring the distributions

```
library(MASS)

kx1=density(x1)
```

```
q_range<-range(0,x1,x2,x3,x4)
plot(kx1,col="black",lwd=2.5, xlab="Loss",
main="Log-Normal distribution densities",
xlim= q_range)
kx2=density(x2)
lines(kx2,col="black",lwd=2.5, type="l", lty=2)
kx3=density(x3)
lines(kx3,col="black",lwd=2.5, type="l", pch=23, lty=3)
kx4=density(x4)
lines(kx4,col="black",lwd=2.5, type="l", pch=24, lty=4)
legend(6500000, 1.0e-06,c("X1", "X2","X3", "X4" ),
lwd=2.5, cex=1, lty=1:4, title="LoB")
kx5=density(x5)
q1_range<-range (0,x5,x6,x7,x8)
plot(kx5,col="black",lwd=2.5, xlab="Loss",
main="Burr distribution densities", xlim= q1_range)
kx6=density(x6)
lines(kx6,col="black",lwd=2.5, type="l", lty=2)
kx7=density(x7)
lines(kx7,col="black",lwd=2.5, type="l", lty=3)
kx8=density(x8)
lines(kx8,col="black",lwd=2.5, type="l", lty=4)
legend(4500000, 3e-06, legend=c("X5", "X6","X7", "X8" ),
  lwd=2.5, cex=1, lty=1:4,
title="LoB")
kx9=density(x9)
q2_range<-range (0,x9,x10,x11,x12)
plot(kx9,col="black",lwd=2.5, xlab="Loss",
main="Single parameter Pareto distribution
densities", xlim= q2_range)
kx10=density(x10)
lines(kx10,col="black",lwd=2.5, type="l", lty=2)
kx11=density(x11)
lines(kx11,col="black",lwd=2.5, type="l", lty=3)
kx12=density(x12)
lines(kx12,col="black",lwd=2.5, type="l", lty=4)
legend(3200000, 1.0e-05, legend=c("X9", "X10","X11", "X12" ),
```

```
 lwd=2.5, cex=1, lty=1:4, title="LoB")
kx13=density(x13)
q3_range<-range (0,x13,x14,x15,x16)
plot(kx13,col="black",lwd=2.5, xlab="Loss",
main="Inverse Weibull distribution
densities", xlim=c(0,4.5e+7))
kx14=density(x14)
lines(kx14,col="black",lwd=2.5, type="l", lty=2)
kx15=density(x15)
lines(kx15,col="black",lwd=2.5, type="l", lty=3)
kx16=density(x16)
lines(kx16,col="black",lwd=2.5, type="l", lty=4)
legend(3.3e+7, 1.0e-06, legend=c("X13", "X14","X15", "X16" ),
lwd=2.5, cex=1, lty=1:4, title="LoB")
```

A.2 Copula Models fitting

```
> library(copula)
> library(actuar)

#first line of business
> gmb<-gumbelCopula(3, dim = 4)
> myCDF<- mvdc(gmb, c("exp","exp","exp","exp"),
list(list(rate=300),
list(rate=300),list(rate=300),list(rate=300)))
> x <- matrix(1/c(x1,x2,x3,x4),byrow=T,ncol=4)
> Fitted4<-fitMvdc(x, myCDF, c(3,3,3,3,3),method="SANN")
> par(mfrow = c(2, 2))
> K <- density(1/x[,1])
> #x[,1] gets risk 1 results
> plot(K, main="Marginal 1",xlab="Loss",lwd=4)
> K <- density(1/x[,2])
> #x[,2] gets risk 2 results
> plot(K, main="Marginal 2",xlab="Loss",lwd=4)
> K <- density(1/x[,3])
> #x[,3] gets risk 3 results
> plot(K, main="Marginal 3",xlab="Loss",lwd=4)
> K <- density(1/x[,4])
> #x[,1] gets risk 4 results
> plot(K, main="Marginal 4",xlab="Loss",lwd=4)

#second line of business
> gmb<-gumbelCopula(3, dim = 4)
> myCDF<- mvdc(gmb, c("exp","exp","exp","exp"),
list(list(rate=300),
list(rate=300),list(rate=300),list(rate=300)))
> x <- matrix(1/c(x5,x6,x7,x8),byrow=T,ncol=4)
> Fitted5<-fitMvdc(x, myCDF, c(3,20,20,20,20),
method="SANN")
> par(mfrow = c(2, 2))
> K <- density(1/x[,1])
> #x[,1] gets risk 1 results
```

```
> plot(K, main="Marginal 1",xlab="Loss",lwd=4)
> K <- density(1/x[,2])
> #x[,2] gets risk 2 results
> plot(K, main="Marginal 2",xlab="Loss",lwd=4)
> K <- density(1/x[,3])
> #x[,3] gets risk 3 results
plot(K, main="Marginal 3",xlab="Loss",lwd=4)
> K <- density(1/x[,4])
> #x[,1] gets risk 4 results
> plot(K, main="Marginal 4",xlab="Loss",lwd=4)

#third class of business
> gmb<-gumbelCopula(3, dim = 4)
> myCDF<- mvdc(gmb, c("exp","exp","exp","exp"),
list(list(rate=300),
list(rate=300),list(rate=300),list(rate=300)))
> x <- matrix(1/c(x9,x10,x11,x12),byrow=T,ncol=4)
> Fitted6<-fitMvdc(x, myCDF, c(3,20,20,20,20),
method="SANN")
#fourth class of business
gmb<-gumbelCopula(3, dim = 4)
> myCDF<- mvdc(gmb, c("exp","exp","exp","exp"),
list(list(rate=300),
list(rate=300),list(rate=300),list(rate=300)))
> x <- matrix(1/c(x13,x14,x15,x16),byrow=T,ncol=4)
> Fitted7<-fitMvdc(x, myCDF, c(3,20,20,20,20),
method="SANN")
> K <- density(1/x[,1])
> #x[,1] gets risk 1 results
> plot(K, main="Marginal 1")
> K <- density(1/x[,2])
> #x[,1] gets risk 1 results
> plot(K, main="Marginal 2")
> K <- density(1/x[,3])
> #x[,1] gets risk 1 results
> plot(K, main="Marginal 3")
> K <- density(1/x[,4])
```

```
> #x[,1] gets risk 1 results
> plot(K, main="Marginal 4")

#The second tire
fra<-frankCopula(2, dim = 4)
> myCDF<- mvdc(fra, c("norm","norm","norm","norm"),
list(list(mean=.5,sd=2),
list(mean=.5,sd=2),list(mean=.5,sd=2),
list(mean=.5,sd=2)))
> x <- cbind(x21,x22)
> Fitted11<-fitMvdc(x, myCDF, c(2,1,1,1,1,1,1,1,1),
method="SANN")
> E=rexp(10000000,3)
> E1=density(E)
> plot(E1, main="Marginal 1",col="red",
xlab="Variable X21",lty="dashed")
> K <- density(x[,1])
> #x[,1] gets risk 1 results
> lines(K, main="Marginal 1",lwd=4,col="black")
> N=rnorm(10000000,.49,.28)
> N1=density(N)
> lines(N1,lwd=2,col="blue",lty="dashed")
> #lines(K, main="Marginal 2",lwd=4,col="blue",lwd=2)
> legend(1.5, 1.15, legend=c("Exponential",
"Normal","Normal marginal"),
   lwd=2.2, col=c("red","blue", "black"))

fra<-frankCopula(2, dim = 4)
> myCDF<- mvdc(fra, c("norm","norm","norm","norm"),
list(list(mean=.5,sd=2),
list(mean=.5,sd=2),list(mean=.5,sd=2),
list(mean=.5,sd=2)))
> x <- cbind(x23,x24)
> Fitted12<-fitMvdc(x, myCDF, c(2,1,1,1,1,1,1,1,1),
method="SANN")
fra<-frankCopula(2, dim = 4)
> myCDF<- mvdc(fra, c("norm","norm","norm","norm"),
```

```
list(list(mean=.5,sd=2),
list(mean=.5,sd=2),list(mean=.5,sd=2),
list(mean=.5,sd=2)))
> x <- cbind(x31,x32)
> Fitted13<-fitMvdc(x, myCDF, c(2,1,1,1,1,1,1,1,1),
method="SANN")
```

A.3 Copula Graphics codes

Simulating and plotting the Copulas

```
library(copula)
myCop.norm <- ellipCopula(family = "normal", dim = 3,
 dispstr = "ex",param = 0.5)
myCop.t <- ellipCopula(family = "t", dim = 3,
dispstr = "toep",param = c(0.8, 0.5), df = 8)
#simulation and ploting
par(mfrow = c(1, 2), mar = c(2, 2, 1, 1),
oma = c(1, 1, 0, 0),mgp = c(2, 1, 0))
u <- rcopula(myCop.norm, 5000)
scatterplot3d(u, xlab="x", ylab="y", zlab="z",
main="Normal Copula")
v <- rcopula(myCop.t, 5000)
scatterplot3d(v,xlab="x", ylab="y", zlab="z",
main="t Copula")
myCop.norm1 <- ellipCopula(family = "normal",
dim = 2, dispstr ="ex",param = 0.5)
myCop.t1 <- ellipCopula(family = "t", dim = 2,
dispstr = "toep",param =c(0.5), df = 8)
par(mfrow = c(1, 2), mar = c(2, 2, 1, 1),
oma = c(1, 1, 0, 0),mgp = c(2, 1, 0))
persp(myCop.norm1, dcopula, xlab = "x",
ylab = "y", zlab = "Density",
main = "Normal Copula",theta = 30, phi = 35,
expand = 0.7, col = "lightblue",
ltheta = 10, shade = 0.5, ticktype = "detailed")
persp(myCop.t1, dcopula, xlab = "x", ylab = "y",
zlab = "Density",main = "t Copula",
theta = 30, phi = 35, expand = 0.7, col = "lightblue",
ltheta = 10, shade = 0.5,
ticktype = "detailed")
myCop.norm1 <- ellipCopula(family = "normal", dim = 2,
 dispstr = "ex",param = 0.5)
myCop.t1 <- ellipCopula(family = "t", dim = 2,
```

```
dispstr = "toep",param = c(0.5), df = 8)
par(mfrow = c(1, 2), mar = c(2, 2, 1, 1),
oma = c(1, 1, 0, 0),mgp = c(2, 1, 0))
persp(myCop.norm1, pcopula, xlab = "x", ylab = "y",
zlab = "cdf",main = "Normal
Copula",theta = 30, phi = 35, expand = 0.7,
col = "lightblue",ltheta = 10,
shade = 0.5, ticktype = "detailed")
persp(myCop.t1, pcopula, xlab = "x", ylab = "y",
 zlab = "cdf",main = "t Copula",
theta = 30, phi = 35, expand = 0.7, col =
"lightblue",ltheta = 10,
shade = 0.5, ticktype = "detailed")
par(mfrow = c(1, 2), mar = c(2, 2, 1, 1),
oma = c(1, 1, 0, 0),mgp = c(2, 1, 0))
contour(myCop.norm1, dcopula, xlab = "x",
ylab = "y",main = "Normal Copula",
 sub = NULL, col = "black")
contour(myCop.t1, dcopula, xlab = "x",
ylab = "y",main = "t Copula",
sub = NULL, col = "black")

#Archimedean copulas

myMvd1 <- mvdc(copula = archmCopula
(family = "clayton", param = 2),margins =
c("norm", "norm"), paramMargins = list(list
(mean = 0,sd = 1), list(mean = 0, sd = 1)))
myMvd2 <- mvdc(copula = archmCopula(family =
"frank", param = 5.736),
margins = c("norm",
"norm"), paramMargins = list(list(mean = 0,sd = 1),
list(mean = 0, sd = 1)))
myMvd3 <- mvdc(copula = archmCopula(family = "gumbel",
param = 2),margins = c("norm",
"norm"), paramMargins = list(list(mean = 0,sd = 1),
list(mean = 0, sd = 1)))
```

```
par(mfrow = c(1, 3), mar = c(2, 2, 1, 1),
oma = c(1, 1, 0, 0),mgp = c(2, 1, 0))
contour(myMvd1, dmvdc, xlim = c(-3, 3),
ylim = c(-3, 3), main="Clayton")
contour(myMvd2, dmvdc, xlim = c(-3, 3),
ylim = c(-3, 3),main="Frank")
contour(myMvd3, dmvdc, xlim = c(-3, 3),
ylim = c(-3, 3),main="Gumbel")
clayton.cop <- claytonCopula(sqrt(3), dim = 3)
frank.cop <- frankCopula(sqrt(3), dim = 3)
gumbel.cop <- archmCopula("gumbel", sqrt(3),dim = 3)
par(mfrow = c(1, 3), mar = c(2, 2, 1, 1),
oma = c(1, 1, 0, 0),mgp = c(2, 1, 0))
scatterplot3d(rcopula(clayton.cop, 5000),
main="Clayton", xlab="x", ylab="y", zlab="z")
scatterplot3d(rcopula(frank.cop, 5000),
main="Frank", xlab="x", ylab="y", zlab="z")
scatterplot3d(rcopula(gumbel.cop, 5000),
main="Gumbel", xlab="x", ylab="y", zlab="z")
clayton.cop1 <- claytonCopula(sqrt(3), dim = 2)
frank.cop1 <- frankCopula(sqrt(3), dim = 2)
gumbel.cop1 <- archmCopula("gumbel", sqrt(3),dim = 2)
par(mfrow = c(1, 3), mar = c(2, 2, 1, 1),
oma = c(1, 1, 0, 0),mgp = c(2, 1, 0))
persp(clayton.cop1, dcopula, xlab = "x", ylab = "y",
zlab = "Density",main = "Clayton",
 theta = 30, phi = 30, expand = 0.78, col = "lightblue",
 ltheta = 120, shade = 0.75,
  ticktype = "detailed")
persp(frank.cop1, dcopula, xlab = "x", ylab = "y",
zlab = "Density",main = "Frank",
theta = 30, phi = 30, expand = 0.78, col = "lightblue",
ltheta = 120, shade = 0.75,
ticktype = "detailed")
persp(gumbel.cop1, dcopula, xlab = "x", ylab = "y",
zlab = "Density",main = "Gumbel",
 theta = 30, phi = 30, expand = 0.78, col = "lightblue",
```

```
 ltheta = 120, shade = 0.75,
  ticktype = "detailed")
par(mfrow = c(1, 3), mar = c(2, 2, 1, 1),
oma = c(1, 1, 0, 0),mgp = c(2, 1, 0))
persp(clayton.cop1, pcopula, xlab = "x", ylab = "y",
zlab = "CDF",main = "Clayton",
theta = 30, phi = 30, expand = 0.78, col = "lightblue",
ltheta = 120, shade = 0.75,
ticktype = "detailed")
persp(frank.cop1, pcopula, xlab = "x", ylab = "y",
zlab = "CDF",main = "Frank",
theta = 30, phi = 30, expand = 0.78, col = "lightblue",
ltheta = 120, shade = 0.75,
ticktype = "detailed")
persp(gumbel.cop1, pcopula, xlab = "x", ylab = "y",
zlab = "CDF",main = "Gumbel",
theta = 30, phi = 30, expand = 0.78, col = "lightblue",
ltheta = 120, shade = 0.75,
 ticktype = "detailed")
```

A.4 Calculating measures of dependence from the dependence parameter

```
library(copula)
gumbel11.cop <- archmCopula("gumbel", 24.59428,dim = 2)
kendallsTau(gumbel11.cop)
spearmansRho(gumbel11.cop)
tailIndex(gumbel11.cop)
gumbel12.cop <- archmCopula("gumbel", 20.45,dim = 2)
kendallsTau(gumbel12.cop)
spearmansRho(gumbel12.cop)
tailIndex(gumbel12.cop)
gumbel13.cop <- archmCopula("gumbel", 23.43973,dim = 2)
kendallsTau(gumbel13.cop)
spearmansRho(gumbel13.cop)
tailIndex(gumbel13.cop)
gumbel14.cop <- archmCopula("gumbel", 19.80858,dim = 2)
kendallsTau(gumbel14.cop)
spearmansRho(gumbel14.cop)
tailIndex(gumbel14.cop)
gumbel15.cop <- archmCopula("gumbel", 11.3086,dim = 2)
kendallsTau(gumbel15.cop)
spearmansRho(gumbel15.cop)
tailIndex(gumbel15.cop)
gumbel16.cop <- archmCopula("gumbel", 17.95782,dim = 2)
kendallsTau(gumbel16.cop)
spearmansRho(gumbel16.cop)
tailIndex(gumbel16.cop)
gumbel17.cop <- archmCopula("gumbel", 14.35448,dim = 2)
kendallsTau(gumbel17.cop)
spearmansRho(gumbel17.cop)
tailIndex(gumbel17.cop)
gumbel18.cop <- archmCopula("gumbel", 11.8632,dim = 2)
kendallsTau(gumbel18.cop)
spearmansRho(gumbel18.cop)
tailIndex(gumbel18.cop)
gumbel19.cop <- archmCopula("gumbel", 25.86368,dim = 2)
```

```
kendallsTau(gumbel19.cop)
spearmansRho(gumbel19.cop)
tailIndex(gumbel19.cop)
gumbel110.cop <- archmCopula("gumbel", 22.24507,dim = 2)
kendallsTau(gumbel110.cop)
spearmansRho(gumbel110.cop)
tailIndex(gumbel110.cop)
gumbel111.cop <- archmCopula("gumbel", 21.4728,dim = 2)
kendallsTau(gumbel111.cop)
spearmansRho(gumbel111.cop)
tailIndex(gumbel111.cop)
gumbel112.cop <- archmCopula("gumbel", 19.98291,dim = 2)
kendallsTau(gumbel112.cop)
spearmansRho(gumbel112.cop)
tailIndex(gumbel112.cop)
gumbel113.cop <- archmCopula("gumbel", 17.46607,dim = 2)
kendallsTau(gumbel113.cop)
spearmansRho(gumbel113.cop)
tailIndex(gumbel113.cop)
gumbel114.cop <- archmCopula("gumbel", 14.86397,dim = 2)
kendallsTau(gumbel114.cop)
spearmansRho(gumbel114.cop)
tailIndex(gumbel114.cop)
gumbel115.cop <- archmCopula("gumbel", 19.41276,dim = 2)
kendallsTau(gumbel115.cop)
spearmansRho(gumbel115.cop)
tailIndex(gumbel115.cop)
gumbel116.cop <- archmCopula("gumbel", 12.79195,dim = 2)
kendallsTau(gumbel116.cop)
spearmansRho(gumbel116.cop)
tailIndex(gumbel116.cop)
gumbel22.cop <- archmCopula("gumbel", 18.35061,dim = 2)
kendallsTau(gumbel22.cop)
spearmansRho(gumbel22.cop)
tailIndex(gumbel22.cop)
gumbel23.cop <- archmCopula("gumbel", 19.83296,dim = 2)
kendallsTau(gumbel23.cop)
```

```
spearmansRho(gumbel23.cop)
tailIndex(gumbel23.cop)
gumbel24.cop <- archmCopula("gumbel", 16.89746,dim = 2)
kendallsTau(gumbel24.cop)
spearmansRho(gumbel24.cop)
tailIndex(gumbel24.cop)
gumbel25.cop <- archmCopula("gumbel", 10.59436,dim = 2)
kendallsTau(gumbel25.cop)
spearmansRho(gumbel25.cop)
tailIndex(gumbel25.cop)
gumbel26.cop <- archmCopula("gumbel", 16.37294,dim = 2)
kendallsTau(gumbel26.cop)
spearmansRho(gumbel26.cop)
tailIndex(gumbel26.cop)
gumbel27.cop <- archmCopula("gumbel", 13.1148,dim = 2)
kendallsTau(gumbel27.cop)
spearmansRho(gumbel27.cop)
tailIndex(gumbel27.cop)
gumbel28.cop <- archmCopula("gumbel", 11.28093,dim = 2)
kendallsTau(gumbel28.cop)
spearmansRho(gumbel28.cop)
tailIndex(gumbel28.cop)
gumbel29.cop <- archmCopula("gumbel", 21.92459,dim = 2)
kendallsTau(gumbel29.cop)
spearmansRho(gumbel29.cop)
tailIndex(gumbel29.cop)
gumbel210.cop <- archmCopula("gumbel", 19.89369,dim = 2)
kendallsTau(gumbel210.cop)
spearmansRho(gumbel210.cop)
tailIndex(gumbel210.cop)
gumbel211.cop <- archmCopula("gumbel", 19.17953,dim = 2)
kendallsTau(gumbel211.cop)
spearmansRho(gumbel211.cop)
tailIndex(gumbel211.cop)
gumbel212.cop <- archmCopula("gumbel", 17.70546,dim = 2)
kendallsTau(gumbel212.cop)
spearmansRho(gumbel212.cop)
```

```
tailIndex(gumbel212.cop)
gumbel213.cop <- archmCopula("gumbel", 15.04316,dim = 2)
kendallsTau(gumbel213.cop)
spearmansRho(gumbel213.cop)
tailIndex(gumbel213.cop)
gumbel214.cop <- archmCopula("gumbel", 13.12138,dim = 2)
kendallsTau(gumbel214.cop)
spearmansRho(gumbel214.cop)
tailIndex(gumbel214.cop)
gumbel215.cop <- archmCopula("gumbel", 17.28395,dim = 2)
kendallsTau(gumbel215.cop)
spearmansRho(gumbel215.cop)
tailIndex(gumbel215.cop)
gumbel216.cop <- archmCopula("gumbel", 12.10579,dim = 2)
kendallsTau(gumbel216.cop)
spearmansRho(gumbel216.cop)
tailIndex(gumbel216.cop)
gumbel33.cop <- archmCopula("gumbel", 21.02007,dim = 2)
kendallsTau(gumbel33.cop)
spearmansRho(gumbel33.cop)
tailIndex(gumbel33.cop)
gumbel34.cop <- archmCopula("gumbel", 18.39264,dim = 2)
kendallsTau(gumbel34.cop)
spearmansRho(gumbel34.cop)
tailIndex(gumbel34.cop)
gumbel35.cop <- archmCopula("gumbel", 10.37714,dim = 2)
kendallsTau(gumbel35.cop)
spearmansRho(gumbel35.cop)
tailIndex(gumbel35.cop)
gumbel36.cop <- archmCopula("gumbel", 16.88054,dim = 2)
kendallsTau(gumbel36.cop)
spearmansRho(gumbel36.cop)
tailIndex(gumbel36.cop)
gumbel37.cop <- archmCopula("gumbel", 13.76924,dim = 2)
kendallsTau(gumbel37.cop)
spearmansRho(gumbel37.cop)
tailIndex(gumbel37.cop)
```

```
gumbel38.cop <- archmCopula("gumbel", 11.45215,dim = 2)
kendallsTau(gumbel38.cop)
spearmansRho(gumbel38.cop)
tailIndex(gumbel38.cop)
gumbel39.cop <- archmCopula("gumbel", 23.88196,dim = 2)
kendallsTau(gumbel39.cop)
spearmansRho(gumbel39.cop)
tailIndex(gumbel39.cop)
gumbel310.cop <- archmCopula("gumbel", 22.06464,dim = 2)
kendallsTau(gumbel310.cop)
spearmansRho(gumbel310.cop)
tailIndex(gumbel310.cop)
gumbel311.cop <- archmCopula("gumbel", 19.6937,dim = 2)
kendallsTau(gumbel311.cop)
spearmansRho(gumbel311.cop)
tailIndex(gumbel311.cop)
gumbel312.cop <- archmCopula("gumbel", 18.49001,dim = 2)
kendallsTau(gumbel312.cop)
spearmansRho(gumbel312.cop)
tailIndex(gumbel12.cop)
gumbel313.cop <- archmCopula("gumbel", 16.08886,dim = 2)
kendallsTau(gumbel313.cop)
spearmansRho(gumbel313.cop)
tailIndex(gumbel313.cop)
gumbel314.cop <- archmCopula("gumbel", 13.58637,dim = 2)
kendallsTau(gumbel314.cop)
spearmansRho(gumbel314.cop)
tailIndex(gumbel314.cop)
gumbel315.cop <- archmCopula("gumbel", 19.40581,dim = 2)
kendallsTau(gumbel315.cop)
spearmansRho(gumbel315.cop)
tailIndex(gumbel315.cop)
gumbel316.cop <- archmCopula("gumbel", 12.88396,dim = 2)
kendallsTau(gumbel316.cop)
spearmansRho(gumbel316.cop)
tailIndex(gumbel316.cop)
gumbel44.cop <- archmCopula("gumbel", 17.23922,dim = 2)
```

```
kendallsTau(gumbel44.cop)
spearmansRho(gumbel44.cop)
tailIndex(gumbel44.cop)
gumbel45.cop <- archmCopula("gumbel", 10.09766,dim = 2)
kendallsTau(gumbel45.cop)
spearmansRho(gumbel45.cop)
tailIndex(gumbel45.cop)
gumbel46.cop <- archmCopula("gumbel", 16.08946,dim = 2)
kendallsTau(gumbel46.cop)
spearmansRho(gumbel46.cop)
tailIndex(gumbel46.cop)
gumbel47.cop <- archmCopula("gumbel", 12.43129,dim = 2)
kendallsTau(gumbel47.cop)
spearmansRho(gumbel47.cop)
tailIndex(gumbel47.cop)
gumbel48.cop <- archmCopula("gumbel", 10.7525,dim = 2)
kendallsTau(gumbel48.cop)
spearmansRho(gumbel48.cop)
tailIndex(gumbel48.cop)
gumbel49.cop <- archmCopula("gumbel", 20.72022,dim = 2)
kendallsTau(gumbel49.cop)
spearmansRho(gumbel49.cop)
tailIndex(gumbel49.cop)
gumbel410.cop <- archmCopula("gumbel", 18.59881,dim = 2)
kendallsTau(gumbel410.cop)
spearmansRho(gumbel410.cop)
tailIndex(gumbe410.cop)
gumbel411.cop <- archmCopula("gumbel", 17.82808,dim = 2)
kendallsTau(gumbel411.cop)
spearmansRho(gumbel411.cop)
tailIndex(gumbel411.cop)
gumbel412.cop <- archmCopula("gumbel", 16.75534,dim = 2)
kendallsTau(gumbel412.cop)
spearmansRho(gumbel412.cop)
tailIndex(gumbel412.cop)
gumbel413.cop <- archmCopula("gumbel", 14.84369,dim = 2)
kendallsTau(gumbel413.cop)
```

```
spearmansRho(gumbel413.cop)
tailIndex(gumbel413.cop)
gumbel414.cop <- archmCopula("gumbel", 12.29216,dim = 2)
kendallsTau(gumbel414.cop)
spearmansRho(gumbel414.cop)
tailIndex(gumbel414.cop)
gumbel415.cop <- archmCopula("gumbel", 16.45009,dim = 2)
kendallsTau(gumbel415.cop)
spearmansRho(gumbel415.cop)
tailIndex(gumbel415.cop)
gumbel416.cop <- archmCopula("gumbel", 11.97745,dim = 2)
kendallsTau(gumbel416.cop)
spearmansRho(gumbel416.cop)
tailIndex(gumbel416.cop)
gumbel55.cop <- archmCopula("gumbel", 7.521216,dim = 2)
kendallsTau(gumbel55.cop)
spearmansRho(gumbel55.cop)
tailIndex(gumbel55.cop)
gumbel56.cop <- archmCopula("gumbel", 9.691672,dim = 2)
kendallsTau(gumbel56.cop)
spearmansRho(gumbel56.cop)
tailIndex(gumbel56.cop)
gumbel57.cop <- archmCopula("gumbel", 8.54854,dim = 2)
kendallsTau(gumbel57.cop)
spearmansRho(gumbel57.cop)
tailIndex(gumbel57.cop)
gumbel58.cop <- archmCopula("gumbel", 7.830925,dim = 2)
kendallsTau(gumbel58.cop)
spearmansRho(gumbel58.cop)
tailIndex(gumbel58.cop)
gumbel59.cop <- archmCopula("gumbel", 10.86895,dim = 2)
kendallsTau(gumbel59.cop)
spearmansRho(gumbel59.cop)
tailIndex(gumbel59.cop)
gumbel510.cop <- archmCopula("gumbel", 19.29592,dim = 2)
kendallsTau(gumbel510.cop)
spearmansRho(gumbel510.cop)
```

```
tailIndex(gumbel510.cop)
gumbel511.cop <- archmCopula("gumbel",10.44113,dim = 2)
kendallsTau(gumbel511.cop)
spearmansRho(gumbel511.cop)
tailIndex(gumbel511.cop)
gumbel512.cop <- archmCopula("gumbel", 10.0494,dim = 2)
kendallsTau(gumbel512.cop)
spearmansRho(gumbel512.cop)
tailIndex(gumbel512.cop)
gumbel513.cop <- archmCopula("gumbel", 9.290266,dim = 2)
kendallsTau(gumbel513.cop)
spearmansRho(gumbel513.cop)
tailIndex(gumbel513.cop)
gumbel514.cop <- archmCopula("gumbel", 8.627145,dim = 2)
kendallsTau(gumbel514.cop)
spearmansRho(gumbel514.cop)
tailIndex(gumbel514.cop)
gumbel515.cop <- archmCopula("gumbel", 10.4105,dim = 2)
kendallsTau(gumbel515.cop)
spearmansRho(gumbel515.cop)
tailIndex(gumbel515.cop)
gumbel516.cop <- archmCopula("gumbel", 8.303656,dim = 2)
kendallsTau(gumbel516.cop)
spearmansRho(gumbel516.cop)
tailIndex(gumbel516.cop)
gumbel66.cop <- archmCopula("gumbel", 14.61707,dim = 2)
kendallsTau(gumbel66.cop)
spearmansRho(gumbel66.cop)
tailIndex(gumbel66.cop)
gumbel67.cop <- archmCopula("gumbel", 11.74468,dim = 2)
kendallsTau(gumbel67.cop)
spearmansRho(gumbel67.cop)
tailIndex(gumbel67.cop)
gumbel68.cop <- archmCopula("gumbel", 10.00142,dim = 2)
kendallsTau(gumbel68.cop)
spearmansRho(gumbel68.cop)
tailIndex(gumbel68.cop)
```

```
gumbel69.cop <- archmCopula("gumbel", 19.57513,dim = 2)
kendallsTau(gumbel69.cop)
spearmansRho(gumbel69.cop)
tailIndex(gumbel69.cop)
gumbel610.cop <- archmCopula("gumbel", 16.92689,dim = 2)
kendallsTau(gumbel610.cop)
spearmansRho(gumbel610.cop)
tailIndex(gumbel610.cop)
gumbel611.cop <- archmCopula("gumbel", 16.55551,dim = 2)
kendallsTau(gumbel611.cop)
spearmansRho(gumbel611.cop)
tailIndex(gumbel611.cop)
gumbel612.cop <- archmCopula("gumbel", 15.40211,dim = 2)
kendallsTau(gumbel612.cop)
spearmansRho(gumbel612.cop)
tailIndex(gumbel612.cop)
gumbel613.cop <- archmCopula("gumbel", 13.42136,dim = 2)
kendallsTau(gumbel613.cop)
spearmansRho(gumbel613.cop)
tailIndex(gumbel613.cop)
gumbel614.cop <- archmCopula("gumbel", 12.11641,dim = 2)
kendallsTau(gumbel614.cop)
spearmansRho(gumbel614.cop)
tailIndex(gumbel614.cop)
gumbel615.cop <- archmCopula("gumbel", 16.07287,dim = 2)
kendallsTau(gumbel615.cop)
spearmansRho(gumbel615.cop)
tailIndex(gumbel615.cop)
gumbel616.cop <- archmCopula("gumbel", 11.34403,dim = 2)
kendallsTau(gumbel616.cop)
spearmansRho(gumbel616.cop)
tailIndex(gumbel616.cop)
gumbel77.cop <- archmCopula("gumbel", 10.15075,dim = 2)
kendallsTau(gumbel77.cop)
spearmansRho(gumbel77.cop)
tailIndex(gumbel77.cop)
gumbel78.cop <- archmCopula("gumbel", 9.026152,dim = 2)
```

```
kendallsTau(gumbel78.cop)
spearmansRho(gumbel78.cop)
tailIndex(gumbel78.cop)
gumbel79.cop <- archmCopula("gumbel", 14.50697,dim = 2)
kendallsTau(gumbel79.cop)
spearmansRho(gumbel79.cop)
tailIndex(gumbel79.cop)
gumbel710.cop <- archmCopula("gumbel", 13.80126,dim = 2)
kendallsTau(gumbel710.cop)
spearmansRho(gumbel710.cop)
tailIndex(gumbel710.cop)
gumbel711.cop <- archmCopula("gumbel", 12.86187,dim = 2)
kendallsTau(gumbel711.cop)
spearmansRho(gumbel711.cop)
tailIndex(gumbel711.cop)
gumbel712.cop <- archmCopula("gumbel", 12.30449,dim = 2)
kendallsTau(gumbel712.cop)
spearmansRho(gumbel712.cop)
tailIndex(gumbel712.cop)
gumbel713.cop <- archmCopula("gumbel", 11.21817,dim = 2)
kendallsTau(gumbel713.cop)
spearmansRho(gumbel713.cop)
tailIndex(gumbel713.cop)
gumbel714.cop <- archmCopula("gumbel", 10.33549,dim = 2)
kendallsTau(gumbel714.cop)
spearmansRho(gumbel714.cop)
tailIndex(gumbel714.cop)
gumbel715.cop <- archmCopula("gumbel", 12.56657,dim = 2)
kendallsTau(gumbel715.cop)
spearmansRho(gumbel715.cop)
tailIndex(gumbel715.cop)
gumbel716.cop <- archmCopula("gumbel", 9.608888,dim = 2)
kendallsTau(gumbel716.cop)
spearmansRho(gumbel716.cop)
tailIndex(gumbel716.cop)
gumbel88.cop <- archmCopula("gumbel", 8.380593,dim = 2)
kendallsTau(gumbel88.cop)
```

```
spearmansRho(gumbel88.cop)
tailIndex(gumbel88.cop)
gumbel89.cop <- archmCopula("gumbel", 11.7845,dim = 2)
kendallsTau(gumbel89.cop)
spearmansRho(gumbel89.cop)
tailIndex(gumbel89.cop)
gumbel810.cop <- archmCopula("gumbel", 11.28214,dim = 2)
kendallsTau(gumbel810.cop)
spearmansRho(gumbel810.cop)
tailIndex(gumbel810.cop)
gumbel811.cop <- archmCopula("gumbel",10.7309,dim = 2)
kendallsTau(gumbel811.cop)
spearmansRho(gumbel811.cop)
tailIndex(gumbel811.cop)
gumbel812.cop <- archmCopula("gumbel", 10.94955,dim = 2)
kendallsTau(gumbel812.cop)
spearmansRho(gumbel812.cop)
tailIndex(gumbel812.cop)
gumbel813.cop <- archmCopula("gumbel", 9.773103,dim = 2)
kendallsTau(gumbel813.cop)
spearmansRho(gumbel813.cop)
tailIndex(gumbel813.cop)
gumbel814.cop <- archmCopula("gumbel", 8.775602,dim = 2)
kendallsTau(gumbel814.cop)
spearmansRho(gumbel814.cop)
tailIndex(gumbel814.cop)
gumbel815.cop <- archmCopula("gumbel", 10.95028,dim = 2)
kendallsTau(gumbel815.cop)
spearmansRho(gumbel815.cop)
tailIndex(gumbel815.cop)
gumbel816.cop <- archmCopula("gumbel", 8.63004,dim = 2)
kendallsTau(gumbel816.cop)
spearmansRho(gumbel816.cop)
tailIndex(gumbel816.cop)
gumbel99.cop <- archmCopula("gumbel", 27.39562,dim = 2)
kendallsTau(gumbel99.cop)
spearmansRho(gumbel99.cop)
```

```
tailIndex(gumbel99.cop)
gumbel910.cop <- archmCopula("gumbel", 25.0694,dim = 2)
kendallsTau(gumbel910.cop)
spearmansRho(gumbel910.cop)
tailIndex(gumbel910.cop)
gumbel911.cop <- archmCopula("gumbel", 22.16318,dim = 2)
kendallsTau(gumbel911.cop)
spearmansRho(gumbel911.cop)
tailIndex(gumbel911.cop)
gumbel912.cop <- archmCopula("gumbel", 20.70767,dim = 2)
kendallsTau(gumbel912.cop)
spearmansRho(gumbel912.cop)
tailIndex(gumbel912.cop)
gumbel913.cop <- archmCopula("gumbel",16.97585,dim = 2)
kendallsTau(gumbel913.cop)
spearmansRho(gumbel913.cop)
tailIndex(gumbel913.cop)
gumbel914.cop <- archmCopula("gumbel", 14.59269,dim = 2)
kendallsTau(gumbel914.cop)
spearmansRho(gumbel914.cop)
tailIndex(gumbel914.cop)
gumbel915.cop <- archmCopula("gumbel", 21.83363,dim = 2)
kendallsTau(gumbel915.cop)
spearmansRho(gumbel915.cop)
tailIndex(gumbel915.cop)
gumbel916.cop <- archmCopula("gumbel", 12.93417,dim = 2)
kendallsTau(gumbel916.cop)
spearmansRho(gumbel916.cop)
tailIndex(gumbel916.cop)
gumbel1010.cop <- archmCopula("gumbel", 21.67569,dim = 2)
kendallsTau(gumbel1010.cop)
spearmansRho(gumbel1010.cop)
tailIndex(gumbel1010.cop)
gumbel1011.cop <- archmCopula("gumbel", 20.20132,dim = 2)
kendallsTau(gumbel1011.cop)
spearmansRho(gumbel1011.cop)
tailIndex(gumbel1011.cop)
```

```
gumbel1012.cop <- archmCopula("gumbel", 19.9295,dim = 2)
kendallsTau(gumbel1012.cop)
spearmansRho(gumbel1012.cop)
tailIndex(gumbel1012.cop)
gumbel1013.cop <- archmCopula("gumbel", 15.84157,dim = 2)
kendallsTau(gumbel1013.cop)
spearmansRho(gumbel1013.cop)
tailIndex(gumbel1013.cop)
gumbel1014.cop <- archmCopula("gumbel", 14.06791,dim = 2)
kendallsTau(gumbel1014.cop)
spearmansRho(gumbel1014.cop)
tailIndex(gumbel1014.cop)
gumbel1015.cop <- archmCopula("gumbel", 19.45971,dim = 2)
kendallsTau(gumbel1015.cop)
spearmansRho(gumbel1015.cop)
tailIndex(gumbel1015.cop)
gumbel1016.cop <- archmCopula("gumbel", 12.58622,dim = 2)
kendallsTau(gumbel1016.cop)
spearmansRho(gumbel1016.cop)
tailIndex(gumbel1016.cop)
gumbel1111.cop <- archmCopula("gumbel", 18.77081,dim = 2)
kendallsTau(gumbel1111.cop)
spearmansRho(gumbel1111.cop)
tailIndex(gumbel1111.cop)
gumbel1112.cop <- archmCopula("gumbel", 17.60055,dim = 2)
kendallsTau(gumbel1112.cop)
spearmansRho(gumbel1112.cop)
tailIndex(gumbel1112.cop)
gumbel1113.cop <- archmCopula("gumbel", 14.93381,dim = 2)
kendallsTau(gumbel1113.cop)
spearmansRho(gumbel1113.cop)
tailIndex(gumbel1113.cop)
gumbel1114.cop <- archmCopula("gumbel", 13.79122,dim = 2)
kendallsTau(gumbel1114.cop)
spearmansRho(gumbel1114.cop)
tailIndex(gumbel1114.cop)
gumbel1115.cop <- archmCopula("gumbel", 18.14978,dim = 2)
```

```
kendallsTau(gumbel1115.cop)
spearmansRho(gumbel1115.cop)
tailIndex(gumbel1115.cop)
gumbel1116.cop <- archmCopula("gumbel", 11.85016,dim = 2)
kendallsTau(gumbel1116.cop)
spearmansRho(gumbel1116.cop)
tailIndex(gumbel1116.cop)
gumbel1212.cop <- archmCopula("gumbel", 17.49925,dim = 2)
kendallsTau(gumbel1212.cop)
spearmansRho(gumbel1212.cop)
tailIndex(gumbel1212.cop)
gumbel1213.cop <- archmCopula("gumbel", 14.87095,dim = 2)
kendallsTau(gumbel1213.cop)
spearmansRho(gumbel1213.cop)
tailIndex(gumbel1213.cop)
gumbel1214.cop <- archmCopula("gumbel", 12.80826,dim = 2)
kendallsTau(gumbel1214.cop)
spearmansRho(gumbel1214.cop)
tailIndex(gumbel1214.cop)
gumbel1215.cop <- archmCopula("gumbel", 16.68777,dim = 2)
kendallsTau(gumbel1215.cop)
spearmansRho(gumbel1215.cop)
tailIndex(gumbel1215.cop)
gumbel1216.cop <- archmCopula("gumbel", 11.64434,dim = 2)
kendallsTau(gumbel1216.cop)
spearmansRho(gumbel1216.cop)
tailIndex(gumbel1216.cop)
gumbel1313.cop <- archmCopula("gumbel", 12.54261,dim = 2)
kendallsTau(gumbel1313.cop)
spearmansRho(gumbel1313.cop)
tailIndex(gumbel1313.cop)
gumbel1314.cop <- archmCopula("gumbel", 11.68133,dim = 2)
kendallsTau(gumbel1314.cop)
spearmansRho(gumbel1314.cop)
tailIndex(gumbel1314.cop)
gumbel1315.cop <- archmCopula("gumbel", 14.37334,dim = 2)
kendallsTau(gumbel1315.cop)
```

```
spearmansRho(gumbel1315.cop)
tailIndex(gumbel1315.cop)
gumbel1316.cop <- archmCopula("gumbel", 10.31197,dim = 2)
kendallsTau(gumbel1316.cop)
spearmansRho(gumbel1316.cop)
tailIndex(gumbel1316.cop)
gumbel1414.cop <- archmCopula("gumbel", 10.44898,dim = 2)
kendallsTau(gumbel1414.cop)
spearmansRho(gumbel1414.cop)
tailIndex(gumbel1414.cop)
gumbel1415.cop <- archmCopula("gumbel", 12.87818,dim = 2)
kendallsTau(gumbel1415.cop)
spearmansRho(gumbel1415.cop)
tailIndex(gumbel1415.cop)
gumbel1416.cop <- archmCopula("gumbel", 9.94889,dim = 2)
kendallsTau(gumbel1416.cop)
spearmansRho(gumbel1416.cop)
tailIndex(gumbel1416.cop)
gumbel1515.cop <- archmCopula("gumbel", 17.60056,dim = 2)
kendallsTau(gumbel1515.cop)
spearmansRho(gumbel1515.cop)
tailIndex(gumbel1515.cop)
gumbel1516.cop <- archmCopula("gumbel", 11.60311,dim = 2)
kendallsTau(gumbel1516.cop)
spearmansRho(gumbel1516.cop)
tailIndex(gumbel1516.cop)
gumbel1616.cop <- archmCopula("gumbel", 9.09929,dim = 2)
kendallsTau(gumbel1616.cop)
spearmansRho(gumbel1616.cop)
tailIndex(gumbel1616.cop)
```

A.5 Clustering

Distance matrices

```
> Community.1 <- makecommunitydataset(kencom,row='LOB',
column='LOB1',value='kend')
#Euclidean
> Distmatrix.1 <- vegdist(Community.1,method='euclidean')
#Manhattan
> Distmatrix.1 <- vegdist(Community.1,method='manhattan')
```

Dendrogram

```
>dist.eval(Community.1,'euclidean')
> Cluster.1 <- hclust(distmatrix, method='ward')
> copheneticdist <- cophenetic(Cluster.1)
> mantel(distmatrix,copheneticdist,permutations=100)
```