UNIVERSITÉ DU QUÉBEC À MONTRÉAL

COUNTING PROCESSES AND COPULAS: APPLICATIONS IN INSURANCE

DISSERTATION

PRESENTED

AS PARTIAL REQUIREMENT

TO THE MASTER IN MATHEMATICS

BY

FRANK BARNING

MAY 2018

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

PROCESSUS DE COMPTAGE ET COPULES: APPLICATIONS EN ASSURANCE

MÉMOIRE

PRÉSENTÉ

COMME EXIGENCE PARTIELLE

DE LA MAÎTRISE EN MATHÉMATIQUES

PAR

FRANK BARNING

MAI 2018

REMERCIEMENTS

Tout d'abord, je remercie Dieu Tout Puissant d'avoir veillé sur moi durant toutes ces années d'étude. Ensuite, j'exprime mon admiration et profonde gratitude à mes surperviseurs, Dr. Matthieu Dufour et Dr. Michel Adès. Je leur dois beaucoup pour tout le support, l'inspiration, les conseils et surtout leur engagement à l'excellence. Ils ont été essentiels, non seulement, pour le succès de ce travail mais aussi pour mon progrès et bien-être. Je suis également reconnaissant à tous les membres du département de mathématiques de l'UQAM pour ce magnifique environnement propice à notre étude. J'ai apprécié tous les conseils, l'enseignement et les services inestimables que vous m'avez donnés afin d'assurer un environnement d'étude confortable et épanouissant.

Je remercie de façon particulière Mme Gisèle Legault, analyste au département de mathématiques pour sa disponibilité et son expertise informatique bien utile et fortement appréciée.

Je remercie ma chère épouse Naomi Armah, ma chère fille Elliette N. Barning, ma mère Gladys Dansowaa et mon frère Chris Arthur. Vous êtes toutes et tous très chers à mon cœur.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

$\perp$	unique
$\|\|$	common
MBP	Maximization by Parts
IBNR	Incurred But Not Reported Claims
MLE	Maximum Likelihood Estimation
iid	independent and identically distributed
ρ_s	Spearman's Rho
τ	Kendall's Tau
CPP	Compound Poisson Processes
$U[0,1]$	Uniform distribution on the interval $[0,1]$
GLM	Generalized Linear Regression Models
IFM	Inference Functions for Margins
LC1	Lévy Copula 1, as a notation for a type of Lévy Copula
$\bar{C}$	Survival copula
$1_A(x)$	indicator function (of a set A): $1_A(x) = 1$ if $x \in A$ and otherwise 0
F,G,F_i, etc.	cumulative distribution functions (cdf)
$\bar{F} = 1 - F$	univariate survival function
pdf	"probability density function"
C	copula
$(\Omega, \mathscr{F}, \mathbb{P})$	probability space
$\mathbb{R}$,$\mathbb{R}^+$	real line, non-negative real line
P_C	probability measure induced by a copula C
$U(a,b)$	uniform distribution on (a,b)

RÉSUMÉ

Les processus de comptage ont un rôle majeur et des applications variées dans plusieurs domaines telles que la tarification, la réserve de perte, l'allocation du capital en assurance. Avec toutes ces applications, il y a quelques risques ou des facteurs de risque qui dépendent sur un autre ensemble de risque ou des facteurs de risques, et cela constitue précisément un grand intérêt pour les compagnies d'assurance. Ces compagnies d'assurance veulent construire des modèles spécifiques pour capturer quelques, ou toutes les, structures de dépendance existantes entre les risques connus. Quelques-uns de ces risques connus sont associés avec les processus de comptage. La modélisation de la dépendance utilisant la théorie des copules et les processus de comptage a attiré l'attention de plusieurs chercheurs ces dernières années. Dans ce mémoire, nous étudions deux champs d'intérêt dans la modélisation de la dépendance avec applications en assurance et finance. Premièrement, nous étudions plusieurs méthodes de modélisation, les techniques d'estimation et l'implémentation des algorithmes qui sont utilisés dans la modélisation des copules autour des processus de comptage. Par exemple, dans le deuxième chapitre deux de ce mémoire, nous allons étudier comment la modélisation de la dépendance est utilisée pour un risque bivarié ou pour des facteurs de risque telle que la classe Bonus-Malus et les comptes de réclamations du passé, le compte de réclamations et la taille des réclamations, le compte de réclamations de deux processus de comptage différents qui se sont produits à partir du même événement,etc. Dans la deuxième partie du mémoire, nous scrutons et adressons quelques remarques autour du choix de quelques copules de la première partie 1 de ce mémoire, et nous présentons une discussion au sujet des approches utilisées. Cette deuxième partie du mémoire est motivée par le fait que différentes analyses vont choisir un ensemble différent de distributions univariées pour ajuster les mêmes données et choisir différents types de copules pour modéliser les structures de dépendance. En fait, on cherche dans la seconde partie à répondre à la question : Devons-nous dépendre sur les marginales même si nous avons un ensemble large de données disponibles ? En dernier lieu, nous discutons à propos des estimés des vrais paramètres de copule et nous analysons un ensemble de vraies données.

Mots-Clés : Processus de comptage, copule, structure de dépendance, paramètres de modèle, estimation, copule de Clayton, distribution conjointe, processus de Lévy.

ABSTRACT

Counting process has a major and several applications in different areas such as rate-making, loss reserving and capital allocation in insurance. Within all these applications, some risks or risk factors depend on the other set of risk or risk factors and this dependence is of high interest to insurance companies, particularly rate-making actuaries and loss reserving actuaries. These insurance companies would want to build specific models that captures some or all the dependence structures existing between the known risks. Some of these known risks are associated with counting process. Modeling dependence using copula has drawn the attention of several authors in recent years. In this thesis, we study two areas of interest in dependence modeling with application in insurance and finance. First we study several modeling methods, estimation techniques and implemented algorithms that are used in copula modeling surrounding counting process. For instance, in chapter two of this thesis, we will study how dependence modeling is carried for bivariate risk or risk factors such current Bonus-Malus class and past count of claims, count of claims and size of claims, count of claims of two different counting processes that occurred from the same event etc. In the second part of the thesis, we investigate and address some concerns surrounding the choice of some the copulas in the first part of the thesis and present a discussion to the approaches that were used. This second part of the thesis is motivated by the fact that different analysts will select different set of univariate distributions to fit the same data and choose from different types of copula to model the dependence structures. In fact, the second part seeks to answer the question: Should we depend on the fitting of marginals even if we have large set of data available?. Lastly we discuss about the estimates of a true copula parameter as we analyze a real dataset.

keyworkds: Counting process, Copula, dependence structure, model parameters, estimation, Clayton copula, joint distribution, Lévy process.

INTRODUCTION

Two events A and B are dependent if the occurrence of one event changes the probability of the occurrence of other event. For example, large values of event A always occur with large values of event B or large values of event A always brings about small values of event B. Linear dependence between two events is by far the most popular form of dependence in many disciplines especially in the financial community. It is mostly measured with linear correlation coefficient (see chapter 1). Lawless (2014), in his article explained that, this linear correlation coefficient measures how close a point cloud is to a straight line.

Though the linear correlation coefficient, among many dependence measures, for example Kendall's Tau and Spearman's Rho, is by far the most popular dependence measure used, it is also often misunderstood as a general measure of dependence. The popularity of this linear correlation coefficient started with the ease with which it can be calculated and it is a natural scalar measure of dependence in elliptical distributions (i.e. probability distributions that generalize the multivariate normal distribution there by forming an ellipse, for example the multivariate t-distribution).

However, most random variables are not jointly elliptically distributed, and using linear correlation as a measure of dependence in such situations might prove very misleading (Embrechts et al., 2001). This motivated the use of concordance measures (see chapter 1). Two random variables are concordant when large values of one go with large values of the other. The most obvious application of the concordance is to use them to measure the strength of dependence empirically observed in some set of data.

Recently in actuarial literature, the study of the impact of dependence among risks has become a major and flourishing topic: Even in traditional risk theory, individual risks have usually been assumed to be independent, this assumption is very convenient for tractability but it is not generally realistic. Think for example of the aggregate claim amount in which any random variable represents the individual claim size of an insurer's risk portfolio. When the risk is represented by residential dwellings exposed to danger of an earthquake in a given location or by adjoining buildings in fire insurance, it is unrealistic to state that individual risks are not correlated, because they are subject to the same event cause (Campana & Ferretti, 2005). Many more research work in actuarial science among others may be found in Frees et al. (1996), Free and Valdez (1998) and Frees and Wang (2005).

With the proliferation of large datasets from a variety of sources, perhaps the most pressing and ubiquitous challenge is posed by the need to "leverage/influence big data." This calls for ways to build dependence models involving hundreds, and even thousands of variables (Lawless, 2014).

Unfortunately the traditional measures of dependence: the linear correlation coefficient among others such as Kendall's Tau and the Spearman's correlation coefficient, just by themselves comes with some amount of limitations for bivariate distributions. An interesting concept that makes it possible to study dependence in broader terms was proposed by the American mathematician Abe Sklar in response to a question posed by his French colleague Maurice Fréchet is Copula. Copulas are useful tool to model dependent data as they allow to separate the dependence properties of the data from their marginal properties and to construct multivariate models with marginal distributions of arbitrary form (Aristidis K. Nikoloulopoulos, Dimitris Karlis, 2007). For instance, in finance, copula is used in modelling dependence structures in the analysis of credit risks, the insolvency of several debtors at the same time or for insurances the risk of appearance of different claims at the same time have to be modeled to insure solvency of the bank and insurance, respectively, all the time. Modeling dependence between asset returns and modelling the dependence between companies default times are the most common financial modeling tasks for which copulas are applied frequently (Mai & Scherer, 2014).

Count data occur in several areas in actuarial studies. In property insurance models, insurance claims count form a core part of risk theory. Also in health insurance, count data models have been widely used to estimate the predictors of health care demand. Many studies on copula published has revealed some dependence structures that exist between these count variables and other known variables in the insurance and finance setting. In recent times, modeling in insurance is moving towards a broader perspective on assets and liabilities where dependencies between processes (for example different lines of business, count of claims from an insurance company and its main reinsurer, IBNR problem where late claims arise from the same event, and the distribution of the next claim arrival) are taken into account.

The objective of my present work is to review some of the copulas used in models in the insurance and finance setting in order to explore some methods in constructing copulas, different estimation procedures and techniques used to select the best copula (as this task is no very easy in practice). This will encourage more of the copulas being applied and stimulate further developments of copulas in this area.

CHAPTER I

MODELLING DEPENDENCE WITH COPULAS

1.1 Overview of Modelling Dependence

Suppose there are d variables $y_1, ..., y_d$ and the data set consists of $y_i = (y_{i1}, ..., y_{id})$ for $i = 1, ..., n$ considered as a random sample of size n; that is, the y_i are independent and identically distributed (i.i.d.) realizations of a random vector $\mathbf{Y} = (Y_1, ..., Y_d)$. For example, we can consider the data matrix Y given by:

$$\begin{bmatrix} y_{11} & y_{12} & y_{13} & \cdots & y_{1d} \\ y_{21} & y_{22} & y_{23} & \cdots & y_{2d} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & y_{n3} & \cdots & y_{nd} \end{bmatrix}$$

Two goals are interesting here. First is to study the probabilistic behaviour of each of one of the component of $\mathbf{Y}$ and second, is to investigate the relationship between them. Similarly, in dependence modelling, the steps are:
(i) univariate models for each of the variables $Y_1, ..., Y_d$;
(ii) copula models for the dependence of the d variables.

In step (i), the choices for univariate parametric families with two or more parameters depends on the modality, tail-weight, scale of dispersion and asymmetry. After the univariate models are chosen, next in step (ii) copula models must then be considered to model the dependence.

1.2 What are copulas?

Informally, copulas are functions that join or "couple" multivariate distribution functions to their one-dimensional marginal distribution functions (Nelsen, 1999). The purpose of a copula is to “glue together the margins” or “couple the individual probabilities” (hence the Latin term “copulare”) in order to generate dependence between the variables (Genest & Nešlehová, 2005).

1.2.1 Historical Background on the Development of Copula Theory

The history of copulas may be said to begin with (Fréchet, 1951). Fréchet's problem: given the distribution functions F_j $(j = 1,2,...,d)$ of d random variables $X_1, X_2, ..., X_d$ defined on the same probability space $(\Omega, \mathscr{F}, \mathbb{P})$, what can be said about the set $\Gamma(F_1, F_2, ..., F_d)$ of d-dimensional distribution functions whose marginals are the given F_j? He added that, if the random variables $X_1, X_2, ..., X_d$ are independent then

$$\prod_{j=1}^{d} F_j$$

is a member of the set $\Gamma(F_1, F_2, ..., F_d)$. However the problem is what then are the other members of this set?

In 1959, Abe Sklar obtained the most important result in this respect, by introducing the notion, and the name, of a copula, and proving the theorem that now bears his name (Durante & Sempi, 2010).

In a more simplest possible terms, let us consider the Maurice Fréchet's problem by this way. Suppose that X and Y are two fire claim amounts, say, for which we know how to compute the probabilities $\mathbb{P}(X \leq x)$ and $\mathbb{P}(Y \leq y)$ for any values x and y. Viewed as functions of x and y, the probabilities $F(x) = \mathbb{P}(X \leq x)$ and $G(y) = \mathbb{P}(Y \leq y)$ are called the marginal distribution functions, or margins of X and Y . The question is then how to construct a model for the probability of the events $\{X \leq x\}$ and $\{Y \leq y\}$ occurring simultaneously, denoted $\mathbb{P}(X \leq x \text{ and } Y \leq y)$, while ensuring that X has distribution F and Y has distribution G. Sklar (1959) suggested the equation below as an idea to answer the problem:

$$\mathbb{P}(X \leq x \text{ and } Y \leq y) = C\{\mathbb{P}(X \leq x), \mathbb{P}(Y \leq y)\} \tag{1.1}$$

where C is a specific function of two variables called a copula. From the above equation, one can say that copulas are expressed in the form;

$$C\{u, v\} = \mathbb{P}(U \leq u \text{ and } V \leq v). \tag{1.2}$$

1.2.2 Why Do We Care About Copulas?

Copulas have been of interest to statisticians for two main reasons: firstly, as a way of studying scale-free measures of dependence; and secondly, as a starting point for constructing families of bivariate distributions, sometimes with a vew to simulation (Fisher, 1997). In other words, we care about copulas because, copulas reveal the true nature of dependence between variables and lead to flexible multivariate models (Genest, 2011).

Example 1.2.1

(The following example explains the intuition behind Sklar's work in a financial setting (credit to (Cherubini et al., 2004)).

Assume a product written on the Nikkei 225 and S&P 500 indexes which pays, at some exercise date T, one unit if both are lower than some given levels K_{NKY} and K_{SP}. The price of this digital put option is:

$$DP = e^{-r(T-t)}\mathbb{Q}(K_{NKY}, K_{SP})$$

Where $\mathbb{Q}(K_{NKY}, K_{SP})$ *is the joint risk-neutral probability that both the Japanese and US market indexes are below the corresponding strike prices. Also, r is the discounting rate and t is the time we at which we are pricing the option. A put option gives the holder an option to sell the underlying asset at a strike price K if the price of the underlying asset at the time of expiration is lower than K. How can someone going to buy a put option recover a price that are consistent with the market price or in order words, how can a put option buyer pay for a price which he or she will end up exercising the option at time of expiration?*

In modelling, our first goal will be to study the probabilistic behaviour of each of the marginals and next we investigate their relationship. So we will need some models for the risk-neutral probability $\mathbb{Q}_{NKY}$ *that the Japanese Nikkei Index at time T will be below the level* K_{NKY} *and also the risk-neutral probability* $\mathbb{Q}_{SP}$ *that the US S&P Index at time T will be below the level* K_{SP}*. In* ***financial terms****, we are asking what is the forward price of univariate digital options with strike prices* K_{NKY} *and* K_{SP}*; in* ***statistical terms****, we are estimating from the market data, the marginal risk-neutral distributions of Nikkei and S&P indexes. So our price can be written as;*

$$DP = e^{-r(T-t)}\mathbb{Q}(K_{NKY}, K_{SP}) = e^{-r(T-t)}C(\mathbb{Q}_{NKY}, \mathbb{Q}_{SP})$$

where $C(\cdot,\cdot)$ *is a bivariate function and has a basic requirement to be in the unit interval in order to be able to represent a joint probability distribution. Other three requirements also comes into mind. First, if one of the two events has zero probability, the joint probability that both events occur must also be zero and Secondly, if one event will occur for sure, the joint probability that both the events will take place is the same as the probability that the second event will be observed. Lastly, if the probabilities of both the events increase, then the joint probability should also increase.*

Below are some associated definitions and theorems. H-Volume is a volume contained by a rectangle $[x_1, x_2] \times [y_1, y_2]$ of a $3-$*dimensional* function.

Definition 1.2.1 (2-Increasing)

Let $\emptyset \neq S_1, S_2 \subset \bar{R} =$ extended real line on $\{-\infty, +\infty\}$ and let H be a $S_1 \times S_2 \rightarrow R$ function . The $H-volume$ of $B = [x_1, x_2] \times [y_1, y_2]$ *is defined to be:*

$$V_H([B]) = H(x_2, y_2) - H(x_2, y_1) - H(x_1, y_2) + H(x_1, y_1), \tag{1.3}$$

H *is* $2-$ *increasing function if* $V_H(B) \geq 0$ *for all* $B \subset S_1 \times S_2$.

Definition 1.2.2 (Grounded)

Suppose $b_1 = maxS_1$ *and* $b_2 = maxS_2$ *exist. Then the margins F and G of H are given by*

$$F \ : S_1 \rightarrow R, \ F(x) = H(x, b_2),$$
$$G \ : S_2 \rightarrow R, \ G(y) = H(b_1, y).$$

Suppose also $a_1 = minS_1$ *and* $a_2 = minS_2$ *exist, then H is called grounded if:*

$$H(a_1, y) = H(x, a_2) = 0, \textit{ for all } (x, y) \in S_1 \times S_2$$

Definition 1.2.3 (Bivariate Copula)

A bivariate copula function is a function C, whose domain is $[0,1]^2$ *and whose range is* $[0,1]$ *with the following properties:*

(BC1): $C(x) = 0$ *for all* $x \in [0,1]^2$ *when at least one element of x is* 0;

(BC2): $C(x,1) = C(1,x) = x$ *for all* $x \in [0,1]^2$;

(BC3): *for all* $(a_1, a_2), (b_1, b_2) \in [0,1]^2$ *with* $a_1 \leq a_2$ *and* $b_1 \leq b_2$, *we have :*
$$V_C([a,b]) = C(a_2, b_2) - C(a_1, b_2) - C(a_2, b_1) + C(a_1, b_1) \geq 0.$$

Where (BC1) denotes Bivariate Copula or Axiom or property 1.

The function V_C is called the $C-volume$ of the rectangle $[a,b] \times [c,d]$.

Theorem 1 (Sklar's Theorem)
Let H be a bivariate distribution function with marginal distributions F and G. Then there exists a copula C such that:

$$H(x,y) = C(F(x), G(y)) \tag{1.4}$$

Conversely, for any distribution functions F and G and any copula C, the function H defined above is bivariate distribution function with marginal distributions F and G.Furthermore, if F and G are continuous, then C is unique.

Example 1.2.2
Consider the function $\pi(u,v) = uv$. *This function satisfies conditions* $(BC1)$, $(BC2)$ *and* $(BC3)$, *and hence the function* $\pi(u,v)$ *is a copula.*

1.3 Multivariate Copulas

Inference for multivariate models and in particular higher dimensional copulas is a far less developed area of statistics than univariate applications. One reason for this is, that the likelihood usually is less tractable (Schepsmeier & Stöber, 2014).

This section presents a brief extension of the bivariate copula theory. A multivariate copula can be used to specify a multivariate distribution and every multivariate distribution provides a multivariate copula. We extend the basic properties of the bivariate copulas to the multivariate case.

Definition 1.3.1 (Multivariate Copula)
A p-dimensional copula is a function $C : [0,1]^p \to [0,1]$ *that satisfies:*

(MC1): $C(u_1, ..., u_{i-1}, 0, u_{i+1}, ..., u_p) = 0$ *for all* $1 \leq i \leq p$,

(MC2): $C(1, ..., 1, u, 1, ..., 1) = u$ *for all u in each of the p arguments ,*

(MC3): For $a_i \leq b_i, a_i, b_i \in [0,1], i = 1,...,p,$

$$\sum_{i_1=1}^{2} \cdots \sum_{i_p=1}^{2} (-1)^{i_1+\ldots+i_p} C(u_{1,i1},...,u_{p,ip}) \geq 0$$

where $u_{j,1} = a_j$ *and* $u_{j,2} = b_j$ *for* $j = 1,...,p.$

Theorem 2 (Sklar's Theorem - Multivariate Copulas)

Let H be a p-dimensional distribution function with margins $F_1,...,F_p$*. Then there exists a p-copula C such that for all* $x_i \in \mathbb{R}$,

$$H(x_1,...,x_p) = C(F(x_1),...,F(x_p)) \tag{1.5}$$

If $F_1,...,F_p$ *are continuous, then C is unique.*

Conversely, if $F_1,...,F_p$ *are distribution functions and C is a copula, then H defined by 1.5, is a joint distribution function with margins* $F_1,...,F_p$.

1.4 Fundamental Computational Techniques Used In Copula Theory

1.4.1 The Fréchet-Hoeffding Bounds

This is the original work of Hoeffding (1994) and Fréchet (1951) . Similar to the correlation coefficient and other numeric dependence measures, there exist lower and upper bounds for all copulas. This computational technique is mostly relevant when one is studying the most extreme negative and positive dependence within a family of copulas.

Theorem 3 (Fréchet-Hoeffding Bounds)

Let C be a copula. Then for every (u,v) *in* $[0,1]^2$,

$$max(u+v-1,0) \leq C(u,v) \leq min(u,v) \tag{1.6}$$

Proof. Let (u,v) be an arbitrary point in $[0,1]^2$. Since:

$$\begin{aligned} C(u,v) &\leq C(u,1)=u \\ C(u,v) &\leq C(1,v)=v \\ \Rightarrow C(u,v) &\leq min(u,v) \end{aligned}$$

Furthermore from $(BC3)$ in Definition 1.2.3;

$$\begin{aligned} V_C([a,b]) = C(a_2,b_2)-C(a_1,b_2)-C(a_2,b_1)+C(a_1,b_1) \geq 0 \\ V_C([u,1]\times[v,1]) = 1-u-v+C(u,v) \geq 0 \\ C(u,v) \geq u+v-1 \\ \Rightarrow C(u,v) \geq max(u+v-1,0) \end{aligned}$$

□

Theorem 4 (Multivariate Copula - Bounds)
For every copula C and any u in $[0,1]^p$,

$$max(u_1+u_2+...+u_p-p+1,0) \leq C(u_1,...,u_p) \leq min(u_1,...,u_p) \tag{1.7}$$

In the multidimensional case of the Fréchet-Hoeffding bounds, the upper bound is still a copula but the lower bound is not.

1.4.2 Switching from Distribution to Survival Functions

In some applications, for example in the context of portfolio credit-risk modelling, it is very natural to consider survival functions rather than distribution functions. For instance the lifetime of a company is a random variable X taking only positive values. Modelling with exponential distribution (the most popular distribution on the positive half-axis), we have its probability distribution function as: $f(x) = \lambda e^{-\lambda x}$. The above distribution has many nice and useful analytical properties such as: $\overline{F}(x) = e^{-\lambda x}$ and

$\overline{F}(x)\overline{F}(y) = \overline{F}(x+y)$, where $\overline{F}(x)$ is the survival function of x. Multivariate concepts of the exponential distribution rely on a treatment of multivariate survival functions. In order to apply copula theory for these concepts, the survival analog of Sklar's Theorem is necessary.

Theorem 5 (Sklar's Theorem for Survival Functions)
A function $\overline{F}(x) : R^d \to [0,1]$ is the survival function of some random vector $(X_1, X_2, ..., X_d)$ if and only if there are, a copulas $\hat{C} : [0,1]^d \to [0,1]$ and univariate functions $\overline{F}_1, ..., \overline{F}_d$: $R \to [0,1]$ such that

$$\hat{C}(\overline{F}_1(x_1), ..., \overline{F}_d(x_d)) = \overline{F}(x_1, x_2, ..., x_d), \; x_1, x_2, ..., x_d \in \mathbb{R}$$

The correspondence between $\overline{F}$ and $\hat{C}$ is one-to-one if all survival functions $\overline{F}_1, ..., \overline{F}_d$ are continuous.

Example 1.4.1
Assume that (X_1, X_2) is a random vector on a probability space $(\Omega, \mathbb{F}, \mathbb{P})$ with distribution function $F(x_1, x_2) := \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2)$, which admits the representation $F(x_1, x_2) = C(F_1(x_1), F_2(x_2))$, for $x_1, x_2 \in \mathbb{R}$, with a bivariate copula C and two univariate distribution functions F_1, F_2. Our goal will be to switch by computing the survival function of (X_1, X_2) and the survival copula $\hat{C}$ of C.

$$\begin{aligned}
\overline{F}(x_1, x_2) &:= \mathbb{P}(X_1 > x_1, X_2 > x_2) \\
&= 1 - (\mathbb{P}(X_1 \leq x_1) \cup \mathbb{P}(X_2 \leq x_2)) \\
&= 1 - \mathbb{P}(X_1 \leq x_1) - \mathbb{P}(X_2 \leq x_2) + \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2) \\
&= 1 - (1 - \overline{F}_1(x_1)) - (1 - \overline{F}_2(x_2)) + C(1 - \overline{F}_1(x_1), 1 - \overline{F}_2(x_2)) \\
&= \overline{F}_1(x_1) + \overline{F}_2(x_2) - 1 + C(1 - \overline{F}_1(x_1), 1 - \overline{F}_2(x_2)).
\end{aligned}$$

Now from the above theorem, it implies that the survival copula can also be expressed as;

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

Mai and Scherer (2014), provided a lemma (on page 24) for the multivariate case where the dimension of the random vector is more than two (2).

1.4.3 Invariance Under Strictly Monotone Transformations

In finance, as an example, if dependence between the values of two stock prices at some future time point is modeled in terms of copula, their logarithmic values have the same copula. Also the conversion into other currencies by multiplication with the respective exchange rates or scale changes of credit spreads from percent into basis points have no effect on the dependence structure. This is what we call the invariance of the copula and this happens only when one applies a strictly monotone transformation to the random variables.

Theorem 6 (Invariance of Copulas)
Let $X \sim F$ and $Y \sim G$ be random variables with copula C. If $\alpha(\cdot)$, $\beta(\cdot)$ are increasing functions on RanX and RanY, then $\alpha(X) \sim F_\alpha$ and $\beta(Y) \sim G_\beta$, have copula $C_{\alpha\beta} = C$. Hence C is invariant under increasing transformation X and Y.Only the marginal laws changes.

Proof.

$$
\begin{aligned}
C_{\alpha\beta}(F_\alpha(x), G_\beta(y)) &= \mathbb{P}[\alpha(X) \leq x, \beta(Y) \leq y] \\
&= \mathbb{P}[X \leq \alpha^{-1}(x), Y \leq \beta^{-1}(y)] \\
&= C(F(\alpha^{-1}(x)), G(\beta^{-1}(y))) \\
&= C(\mathbb{P}[X < \alpha^{-1}(x)], \mathbb{P}[Y < \beta^{-1}(y)]) \\
&= C(\mathbb{P}[\alpha(X) < x], \mathbb{P}[\beta(Y) < y]) \\
&= C(F_\alpha(x), G_\beta(y)).
\end{aligned}
$$

□

1.4.4 Copula Derivatives

In finance applications, most of the copulas used are absolutely continuous (Mai & Scherer, 2014). As a result, differentiating these copula functions is simple and further usage of these derivatives comes in handy. In practice, market participants are interested in knowing the risk of their portfolios. More often than not, they require the derivatives of copulas in order to calculate the observed Fisher information in multivariate models (Schepsmeier & Stöber, 2014). Also, most methods of parameter estimation require the use of score functions. And as result, we would need the non-negative copula density $c : (0,1)^2 \to [0,\infty)$ associated to the above bivariate copula $C(u_1,u_2)$ computed from a successive partial differentiation which is given by:

$$c(u_1,u_2) = C(u_1,u_2)u_1u_2 \tag{1.8}$$

It is obvious that, applying the chain rule to Sklar's theorem would yield the joint density function. Bouyé et al (2000) added that, this reduces to the joint density of the random vector (X_1,X_2), given in the relation below:

$$f(x_1,x_2) = c(F_1(x_1),F_2(x_2))f_1(x_1)f_2(x_2) \tag{1.9}$$

where f_1 and f_2 are the density functions of $F_1(x_1)$ and $F_2(x_2)$ respectively. One main usefulness of copula density is that, with the marginal distributions a random vector, one can generate the joint distribution of that random vector and vice versa.

Example 1.4.2 (Copula Density Function)
Assume that the random vector (X_1,X_2) follows the joint normal standard density and that X_1 and X_2 obeys the univariate standard normal density. One can simply derive

the copula density function by:

$$\begin{aligned} c(F_1(x_1),F_2(x_2)) &= \frac{f(x_1,x_2)}{f_1(x_1)f_2(x_2)} \\ &= \frac{\frac{1}{2\pi(\sqrt{1-\rho^2})}e^{-\frac{x_1^2-2\rho x_1x_2+x_2^2}{2(1-\rho^2)}}}{[\frac{1}{\sqrt{2\pi}}e^{-\frac{x_1^2}{2}}][\frac{1}{\sqrt{2\pi}}e^{-\frac{x_2^2}{2}}]} \\ &= \frac{1}{\sqrt{1-\rho^2}}e^{-\frac{(2\rho x_1x_2-x_1^2\rho^2-x_2^2\rho^2)}{2(1-\rho^2)}}. \end{aligned}$$

Example 1.4.3 (Finding Conditional Distribution Functions from Copulas)
Another applied area of copula derivatives is finding a conditional distribution functions from a copula. This can be seen in the partial derivative below:

$$F(x_1|x_2) = \frac{\partial}{\partial x_1}F(x_1,x_2) = \frac{\partial}{\partial x_1}C(F_1(x_1),F_2(x_2)) \tag{1.10}$$

1.5 How to Measure Dependence Structure In Copula Theory

Dependence structure (for example positive and negative dependence, independence, etc) between random variables is completely described by their joint distribution function. Since the notion of dependence between two (or more) random variables is not a simple mathematical concept, it is quite challenging to communicate information like the 'degree','level' or 'type' of dependence. We are able to achieve a simplified version if the information is compressed into a single number that quantifies the degree of dependence (Mai & Scherer, 2014). Most of the definitions and theorems relating to this chapter (with proofs) may be found in Nelson (1998), de Kort (2007) and Joe (2015). Below are some classical dependence measures that are used to quantify certain aspect (such as the strength or type of any dependence structure).

1.5.1 Classical Linear Correlation

The basis of linear correlation is to tell us how well two random variables cluster around a linear function. The linear correlation coefficient measures the degree to which a linear relation succeeds to describe the dependency between random variables.

Definition 1.5.1

For non-degenerate, square integrable random variables X and Y the linear correlation coefficient ρ is

$$\rho = \frac{Cov(X,Y)}{(Var(X)Var(Y))^{\frac{1}{2}}} \tag{1.11}$$

If two random variables are linearly and perfectly dependent, then $\rho = 1$ or $\rho = -1$. Unfortunately, linear correlation is not invariant under non-linear monotonic transformation of random variables.

Proof. Let X be a uniformly distributed random variable on the interval $(0,1)$ and set $Y = X^n, n \geq 1$. The random variables X and Y are perfectly positive dependent.
The n-th moment of X is

$$\mathbb{E}(X^n) = \int_0^1 x^n dx = \frac{1}{1+n} \tag{1.12}$$

The linear correlation between X and Y is

$$\begin{aligned}
\rho &= \frac{\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]}{(\mathbb{E}[X^2] - (\mathbb{E}[X])^2)^{\frac{1}{2}}(\mathbb{E}[Y^2] - (\mathbb{E}[Y])^2)^{\frac{1}{2}}} \\
&= \frac{\mathbb{E}[X^{n+1}] - \mathbb{E}[X]\mathbb{E}[X^n]}{(\mathbb{E}[X^2] - (\mathbb{E}[X])^2)^{\frac{1}{2}}(\mathbb{E}[X^{2n}] - (\mathbb{E}[X^n])^2)^{\frac{1}{2}}} \\
&= \frac{\sqrt{3+6n}}{2+n}
\end{aligned}$$

For $n = 1$, the correlation coefficient equals 1, for $n > 1$ it is less than 1. Hence linear correlation coefficient is not invariant under increasing, non-linear transformation. □

1.5.2 Measures of Concordance And Expressing them as a Function of Copulas

The notion 'concordance measure' was introduced by Scarsini (1984), who aimed to make the following intuition mathematically precise: Two random variables X_1 and X_2

are concordant when large values of X_1 go with large values of X_2. Concordance and its measures are introduced in this section to reflect the strength to which random variables cluster around a monotone function.

Definition 1.5.2

1. *Two observations* (x_1, y_1) *and* (x_2, y_2) *are concordant if* $x_1 < x_2$ *and* $y_1 < y_2$ *or if* $x_1 > x_2$ *and* $y_1 > y_2$. *An equivalent characterization is* $(x_1 - x_2)(y_1 - y_2) > 0$. *The observations* (x_1, y_1) *and* (x_2, y_2) *are said to be discordant if* $(x_1 - x_2)(y_1 - y_2) < 0$.

2. *if* C_1 *and* C_2 *are copulas, we say that* C_1 *is less concordant than* C_2 *(or* C_2 *is more concordant than* C_1*) and write* $C_1 \prec C_2 (C_2 \succ C_1)$ *if*

$$C_1(u) \leq C_2(u) \quad for \quad all \quad u \in \mathbb{I}^d \tag{1.13}$$

Definition 1.5.3

A measure of association $K_C = K_{X,Y}$ *is called a measure of concordance if:*

1. $K_{X,Y}$ *is defined for every pair* X, Y *of random variables,*

2. $-1 \leq K_{X,Y} \leq 1$, $K_{X,X} = 1$, $K_{-X,X} = -1$,

3. $K_{X,Y} = K_{Y,X}$,

4. *if* X *and* Y *are independent then* $K_{X,Y} = K_{C^{\perp}}$,

5. $K_{-X,Y} = K_{X,-Y} = -K_{X,Y}$,

6. *if* C_1 *and* C_2 *are copulas such that* $C_1 \prec C_2$ *then* $K_{C_1} = K_{C_2}$,

7. *if* $\{(X_n, Y_n)\}$ *is a sequence of continuous random variables with copulas* C_n *and if* C_n *converges pointwise to* C, *then* $lim_{n \to \infty} K_{X_n, Y_n} = K_C$.

Lemma 1.5.1. Measures of concordance are invariant under strictly monotone transformation of the random variables. Proof is shown in de Kort (2007).

1.5.3 Examples of Measure of Concordance

Kendall′s tau and Spearman′s rho are two examples of the measure of Concordance. They are also the two standard non-parametric dependence measures that may be expressed in copula forms.

Kendall′s Tau

Let Q be the difference between the probability of concordance and discordance of two independent random vectors (X_1, Y_1) and (X_2, Y_2):

$$Q = \mathbb{P}[(X_1 - X_2) - (Y_1 - Y_2) > 0] - \mathbb{P}[(X_1 - X_2) - (Y_1 - Y_2) < 0] \tag{1.14}$$

In case (X_1, Y_1) and (X_2, Y_2) are iid. random vectors, the quantity Q is called Kendall′s Tau τ.

Given a sample of $\{(x_1, y_1), (x_2, y_2), (x_3, y_3), ..., (x_n, y_n)\}$ of n observations from random vector (X, Y), an unbiased estimator (t) of τ is

$$t := \frac{c-d}{c+d}$$

where d is the number of discordants pairs and c is the number of concordants pairs. For the n observations, we can also express that;

$$c + d = \frac{n(n-1)}{2}.$$

Kendall′s Tau and Spearman's rho (which is defined in this section) may be expressed in a copula form by the following theorems. For a proof, see Embrechets et al. (2001):

Theorem 7 (Kendall′s Tau)
Let $(X, Y)^T$ be a vector of continuous random variables with copula C. The Kendall's Tau of $(X, Y)^T$ is given by:

$$\tau = Q(C,C) = 4\iint\limits_{[0,1]^2} C(u,v)dC(u,v) - 1, \tag{1.15}$$

where the integral above is the expected value of the random variable $C(U,V)$, *with* $U,V \sim U(0,1)$ *has a joint distribution function* C.

Spearman's Rho

Let (X_1,Y_1), (X_2,Y_2) and (X_3,Y_3) be iid from the random vector $(X,Y)^T$ with common joint distribution H, margins F, G and copula C. Spearman's rho is defined to be proportional to the probability of concordance minus the probability of discordance of the pairs (X_1,Y_1) and (X_2,Y_3):

$$\rho_s(X,Y) = 3(\mathbb{P}[(X_1 - X_2)(Y_1 - Y_3) > 0] - \mathbb{P}[(X_1 - X_2)(Y_1 - Y_3) < 0]) \tag{1.16}$$

Theorem 8 (Spearman's Rho)
Let $(X,Y)^T$ *be a vector of continuous random variables with copula* C. *Then the Spearman's Rho for* $(X,Y)^T$ *is given by:*

$$\rho_s(X,Y) = 12\iint\limits_{[0,1]^2} uv dC(u,v) - 3 = 12\iint\limits_{[0,1]^2} C(u,v)dudv - 3 \tag{1.17}$$

Hence, if $X \sim F$ *and* $Y \sim G$, *and we let* $U = F(X)$ *and* $V = G(Y)$, *then*

$$\begin{aligned}
\rho_s(X,Y) &= 12\iint\limits_{[0,1]^2} C(u,v)dudv - 3 = 12\mathbb{E}(UV) - 3 \\
&= \frac{\mathbb{E}(UV) - \frac{1}{4}}{\frac{1}{12}} \\
&= \frac{COV(U,V)}{\sqrt{Var(U)}\sqrt{Var(V)}} \\
&= \rho(F(X),G(Y))
\end{aligned}$$

1.6 Popular Families of Copulas In Insurance and Finance

In insurance and finance, it is common to come across certain popular families of bivariate copulas. These families are mostly presented by their distribution copula functions. Aas (2004) helped in providing a summary to the most common families applicable to finance.

$$C(u,v) = \mathbb{P}(U \leq u, V \leq v) = \int_{-\infty}^{u}\int_{-\infty}^{v} c(s,t)dsdt \tag{1.18}$$

where $c(s,t)$ is the density of the copula. We will consider two parametric families of copulas; the copulas of normal mixture distributions and Archimedean copulas. The first are so-called *implicit* copulas, for which the double integral at the right-hand side of Eq.(1.18) is implied by a well-known bivariate distribution function, while the latter are *explicit* copulas, for which this integral has a simple closed form.

1.6.1 Implicit Copulas

Let us consider two implicit copulas: the Gaussian and the Students t-copulas. Both of them belong to the elliptical family of copulas. They do not come with a simple closed form.

Gaussian copula The Gausian copula is the copula generated by random variables that have a bivariate normal distribution, each with mean 0, variance 1, and correlation ρ. The Gaussian copula is given by

$$C_\rho(u,v) = \int_{-\infty}^{\Phi^{-1}(u)}\int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi(1-\rho^2)^{\frac{1}{2}}} exp\{-\frac{x^2-2\rho xy+y^2}{2(1-\rho^2)}\}dxdy, \tag{1.19}$$

where ρ is the parameter of the copula, and $\Phi^{-1}(\cdot)$ is the inverse of the standard univariate Gaussian distribution function. Due to the popularity of the multivariate normal distribution and the lack of knowledge about other multivariate distributions in the pre-copula days, Gaussian copulas were naturally the first candidates to be applied by financial engineers when copula modeling became popular, Mai and Scherer (2014).

Student's t-copula This copula allows for join fat tails and an increased probability of joint extreme events compared with the Gaussian copula. It is expressed as:

$$C_{\rho,v}(u,v) = \int_{-\infty}^{t_v^{-1}(u)} \int_{-\infty}^{t_v^{-1}(v)} \frac{1}{2\pi(1-\rho^2)^{\frac{1}{2}}} \{1+\frac{x^2-2\rho xy+y^2}{v(1-\rho^2)}\}^{-(v+2)/2} dxdy, \quad (1.20)$$

where ρ and v are the parameters of the copula, and ${t_v}^{-1}(v)$ is the inverse of the standard univariate student-t-distribution with v degrees of freedom, expectation 0 and variance $\frac{v}{v-2}$. In finance, the Students-t dependence structure supports joint extreme movements regardless of the marginal behaviour of the individual assets.

1.6.2 Explicit Copulas

Implicit copulas are noted to have a drawback of complicated algebraic expressions and a great level of symmetry. These drawbacks motivated the such of many other families of copulas. Let us consider two explicit copulas: the Clayton and Gumbel copulas. Both of them belong to the Archimedean family of copulas.

Clayton copula The Clayton copula is an asymmetric copula, exhibiting greater dependence in the negative tail than in the positive. It is given by:

$$C_\delta(u,v) = (u^{-\delta}+v^{-\delta}-1)^{-\frac{1}{\delta}}, \quad (1.21)$$

where $0 < \delta < \infty$ is a parameter controlling the dependence. Perfect dependence is obtained if $\delta \to \infty$, while $\delta \to 0$ implies independence.

Gumbel copula The Gumbel copula is also an asymmetric copula, but it is exhibiting greater dependence in the positive tail than in the negative. This copula is given by:

$$C_\delta(u,v) = exp(-[(-log(u))^\delta+(-log(v))^\delta]^{\frac{1}{\delta}}), \quad (1.22)$$

where $\delta \geq 1$ is a parameter controlling the dependence. Perfect dependence is obtained if $\delta \rightarrow \infty$, while $\delta \rightarrow 1$ implies independence.

1.7 Counting Processes

Definition 1.7.1 (Stochastic Processes)
A stochastic process is a collection of random variables $\{X_t(w), t > 0\}$ where t is a time parameter and w is a path parameter. The process may be continuous (t takes on values on an interval) or discrete ($t = 0, 1, 2, 3, ...$).

Definition 1.7.2 (Counting Processes)
This is a continuous time stochastic process $\{N(t), t \geq 0\}$ (with $N(t)$ representing the total number of "events" that occur by time t) such that:

1. *$N(0) = 0$,*
2. *$N(t)$ is a non-negative integer number for each $t \geq 0$,*
3. *$N(t)$ is increasing for $0 \leq s \leq t$, then $N(s) \leq N(t)$,*
4. *For $s < t$, $N(t) - N(s)$ equals the number of events that occur in the interval $(s, t]$.*

Instances that portray counting processes are situations where words (among many others) such as arriving, entering, exiting and immigrating happen to be the keywords. In insurance environment, the word occurring turns up most to be associated with counting processes. For instance, claims occuring in the time interval $(s, t]$.

Properties of Counting Process

Some counting processes may posses these properties:

1. Independent Increments: This means that the process from any point is independent of that, which has already or previously occurred.

For all $m \geq 1$ and time parts $0 < t_0 < t_1 < ... < t_m$, the random variables $N(t_0)$, $N(t_1) - N(t_0)$, ..., $N(t_m) - N(t_{m-1})$ are mutually independent.

2. Stationary Increments: This also means that the process from any point on has the same distribution as the original process. For $0 \leq s \leq t$, the distribution of $N(s) \leq N(t)$ depends only on the length of the interval $[s,t]$ and not on the time points s and t.

Intuitively, independent and stationary increments properties simply means that the counting process can start all over again at any point in time (credit to Ross (2014)). In modelling the number of insurance claim number, there are two main types of counting process associated with it. The Renewal Process and the Poisson Process. Ross (2014) added that, the Poisson process is a counting process for which the times between successive events are independent and identically distributed exponential variables whilst the possible generalization of this Poisson process to have the times between successive events to be independent and identically distributed with an arbitrary distribution creates a Renewal Process.
Let $\{N(t), t \geq 0\}$ be a counting process and let W_n denote the time between the $(n-1)st$ and the nth event of this process, $n \geq 1$.

Definition 1.7.3 (Renewal Process)
If the sequence of nonnegative random variables $\{X_1, X_2, ...\}$ *is independent and identically distributed, then the counting process* $\{N(t), t \geq 0\}$ *is called a renewal process.*

Remark 1
In general, renewal process do not have independent and stationary increments.

Variables Definition, Relations and Fundamental properties

1. $N(t)$: Total number of insurance claims by time t.

2. Claim Arrival Times $\{T_1, T_2, ... T_m\}$: Time of first arrival of a claim, Time of second arrival of a claim,..., Time of mth arrival of a claim. Also $0 < T_0 < T_1 < ... < T_m$.

3. It is assumed that there is a finite number of claims in each finite interval. $\mathbb{P}(N(t) < \infty) = 1$.

4. $\mathbb{P}(N(t) = 0) = \mathbb{P}(T_1 > t)$.

5. For $n \geq 1$, $\mathbb{P}(N(t) = n) = \mathbb{P}(T_n \leq t < T_{n+1}) = \mathbb{P}(T_n \leq t) - \mathbb{P}(T_{n+1} \leq t)$.

6. $\mathbb{P}(N(t) \geq n) = \mathbb{P}(N_{(0,t]}) = \mathbb{P}(T_n \leq t)$.

7. $\mathbb{P}(N(t) < n) = \mathbb{P}(T_n > t)$.

8. The claim inter-arrival times $(W_1, W_2, ..., W_n)$ are positive random variables and that $T_n = T_1 + T_2 + ... + T_n$.

9. $\mathbb{P}(W_1 > t) = \mathbb{P}(T_1 > t) = \mathbb{P}(N(t) = 0)$.

Proposition 1 *1. The renewal process has finite values for each $t > 0$,*

$\mathbb{P}(N(t) < \infty) = 1,$

2. $\lim_{t \to +\infty} \frac{\mathbb{E}(N(t))}{t} = \lambda$, *where* $\lambda = \mathbb{E}(W_n)$.

Definition 1.7.4 (Non-Homogeneous Poisson Process)

The non-homogeneous poisson process with intensity function $\lambda(t), t > 0$ is a counting process $\{N(t), t \geq 0\}$ such that:

1. $\{N(t), t \geq 0\}$ *has independent increments,*

2. $\lim_{h \to +0} \frac{1}{h}[\mathbb{P}(N(t+h) - N(t) = 1)] = \lambda(t),$

3. $\lim_{h \to +0} \frac{1}{h}[\mathbb{P}(N(t+h) - N(t) > 1)] = 0.$

The above definition means that, this counting process has the independent increments property and that, for the smallest time inter-arrival times, the probability of recording an insurance claim is equivalent to the intensity function $\lambda(t)$, also, there is no possibility to record more than one insurance claim within this same smallest inter-arrival times.

Remark 2

The cumulative intensity function on $[0,t]$ *is given by:*

$$\Lambda(t) = \int_0^t \lambda(u)du \tag{1.23}$$

Similary, the cumulative intensity function on $\Delta = [s,t]$ *is given by:*

$$\begin{aligned} \Lambda(t) &= \int_s^t \lambda(u)du \\ &= \int_0^t \lambda(u)du - \int_0^s \lambda(u)du \\ &= \Lambda(t) - \Lambda(s) \end{aligned}$$

Theorem 9

Let $\{N(t), t \geq 0\}$ *be a Non-Homogeneous Poisson Process with intensity* $\lambda(t)$*. Then the number of claims in the interval* $(s,t]$ *follows a Poisson distribution with mean* $\Lambda(t) - \Lambda(s)$*. That is:*

$$N(t) - N(s) \sim Poisson(\Lambda(t) - \Lambda(s)) \tag{1.24}$$

In particular, on $(0,t]$, $N(t) \sim Poisson(\Lambda(t))$ with $\mathbb{E}(N(t)) = Var(N(t))$.

1.7.1 Application of Compound Processes in Insurance

In insurance, the insurer's income consists of the annual premiums collected from the policyholder while the loss depends on the policyholder's behavior and the cost of each reported claim, making the profit of each insurance contract stochastic. Because of this, it is of great interest for the insurance company both to be able to set suitable annual premiums based on the risk profile of the policyholder and to keep the policyholders with low risk profile that give higher profits. These concerns make it necessary for insurance companies to have good models for the number of insurance claims a policyholder will make, as well as the dependence between the number of claims in different products.

In other words, rate-making forms a core part of insurance process that allows insurers to be able to know their expected loss, expenses and make adequate provision for contingencies. In actuarial studies, the first step in ratemaking is to model the claim frequency distribution. That is the number of claims occurring over a particular period. Traditionally, the claim count distribution in general insurance is assumed to follow the Poisson or the negative binomial distributions, Samson and Thomas (1987) and Yip and Yau (2005).

Definition 1.7.5 (Compound Poisson Process)
Let $\{N(t), t \geq 0\}$ be a Non-Homogeneous Poisson Process, and let $\{X_i, i \geq 1\}$ be a family of independent and identical random variable that is independent of the random variable $N(t)$, then

$$S(t) = \sum_{i=1}^{N(t)} X_i \tag{1.25}$$

$\{S(t), t \geq 0\}$ is called a Compound Poisson Process.

Remark 3 (Basic Properties of Compound Poisson Process)

1. $S(t) = 0$, *if* $N(t) = 0$

2. $E(S(t)) = \Lambda(t)\mathbb{E}(X)$

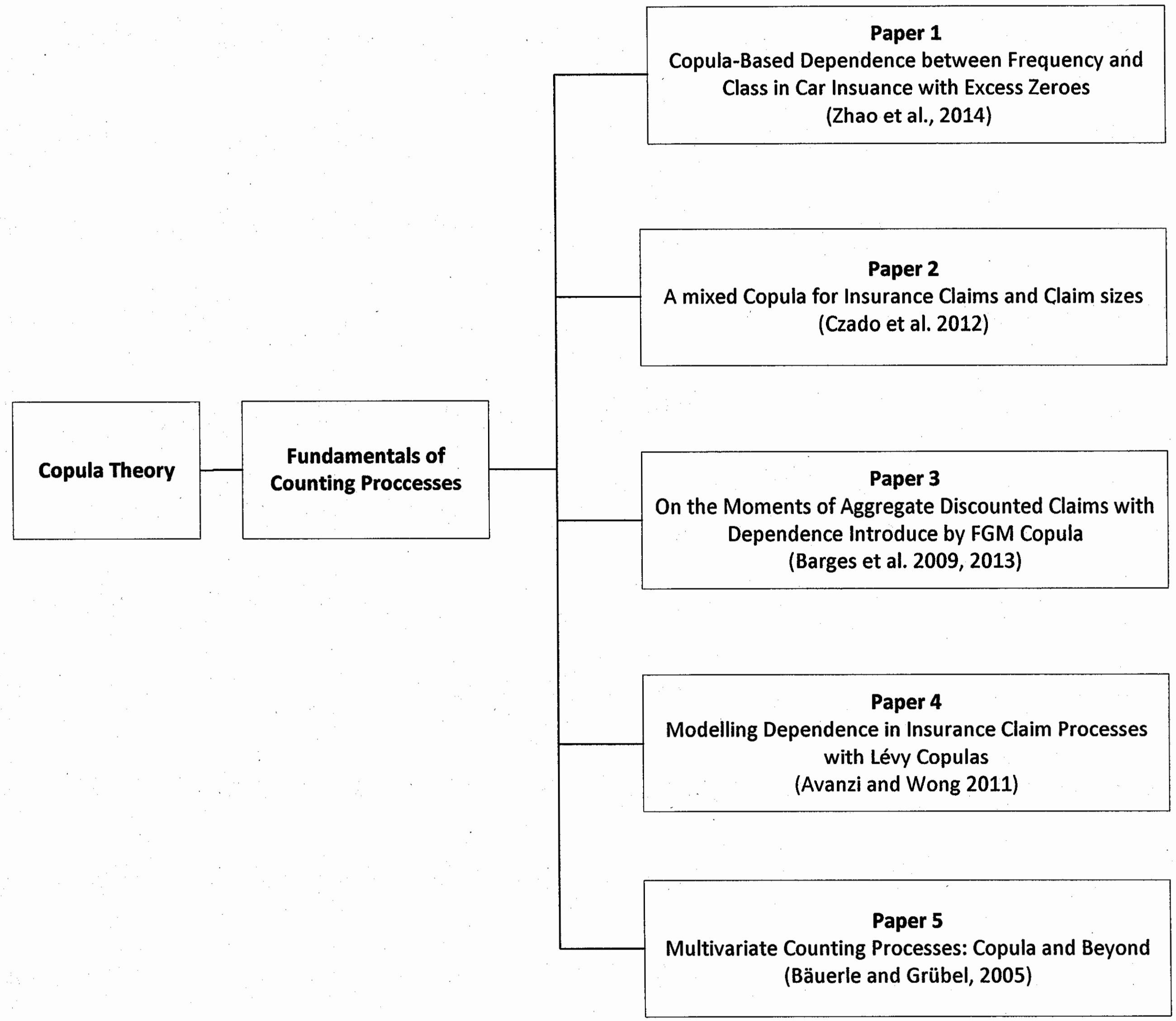
3. $Var(S(t)) = \Lambda(t)\mathbb{E}(X^2)$

In insurance, the random variable X_i represents the ith insurance claim size. $S(t)$ denotes to total claims as at the time t.

1.8 Organization for the rest of this Research Work

The figures below provides a pictorial view of the organization of this work. This current chapter presented the Copula Theory and the Counting Processes. The subsequent chapter, Chapter 2 will provide the multiple versions of the theories in Chapter 1 in the mentioned area of this thesis as we review five(5) published papers relating to this area.

ABSTRACTS:



Copula Theory
Fundamentals of Counting Proccesses
Paper 1
Copula-Based Dependence between Frequency and Class in Car Insuance with Excess Zeroes
(Zhao et al., 2014)
Paper 2
A mixed Copula for Insurance Claims and Claim sizes
(Czado et al. 2012)
Paper 3
On the Moments of Aggregate Discounted Claims with Dependence Introduce by FGM Copula
(Barges et al. 2009, 2013)
Paper 4
Modelling Dependence in Insurance Claim Processes with Lévy Copulas
(Avanzi and Wong 2011)
Paper 5
Multivariate Counting Processes: Copula and Beyond
(Bäuerle and Grübel, 2005)

	Kind of Dependence in this study	Counting Processes	The Form of the Copula Selected
Paper 1 Copula-Based Dependence between Frequency and Class in Car Insuance with Excess Zeroes (Zhao et al., 2014)	Dependence study between Number of Claims at time t, **N(t)** and the Class of a Policyholder **C(t)**	**N(t)**	Clayton Copula $C(u,v)=\left(u^{-\theta}+v^{-\theta}-1\right)^{-1/\theta}$
Paper 2 A mixed Copula for Insurance Claims and Claim sizes (Czado et al. 2012)	Dependence study between Number of Claims at time t, **N(t)** and the Average Claim Size $\mathbf{X_i}$	**N(t)**	Gaussian Copula $C_\Sigma^\Phi(u_1,\ldots,u_n)=\Phi_\Sigma^n\left(\Phi^{-1}(u_1),\ldots,\Phi^{-1}(u_n)\right)$
Paper 3 On the Moments of Aggregate Discounted Claims with Dependence Introduce by FGM Copula (Barges et al. 2009, 2013)	Dependence study between Discounted Claim Size, $\mathbf{X_i}$ and Inter-Arrival Claim Time **W**	**W**	FGM Copula $C_\theta^{FGM}(u,v)=uv+\theta u(1-u)v(1-v)$
Paper 4 Modelling Dependence in Insurance Claim Processes with Lévy Copulas (Avanzi and Wong 2011)	Dependence study between **a group of Lévy Processes** with a common shock affecting both claim frequency **Ni(t)** and Claim Severity $\mathbf{X_i}$ **(t)** in a Compound Poisson Process	$\mathbf{N_1(t)}$ **and** $\mathbf{N_2(t)}$ $\mathbf{X_1(t)}$ **and** $\mathbf{X_2(t)}$	Clayton Copula $C(u,v)=\left(u^{-\theta}+v^{-\theta}-1\right)^{-1/\theta}$
Paper 5 Multivariate Counting Processes: Copula and Beyond (Bäuerle and Grübel, 2005)	Dependence modeling with Lévy Copulas and Introduction of TaS Models	$\mathbf{N_1(t)}$ **and** $\mathbf{N_2(t)}$ $\mathbf{X_1(t)}$ **and** $\mathbf{X_2(t)}$ Across Time and Components	

CHAPTER II

REVIEW OF DIFFERENT COPULAS RELATING TO COUNTING PROCESSES IN INSURANCE TOPICS

2.1 Copula-Based Dependence Between Frequency and Class in Car Insurance with Excess Zeros

2.1.1 Introduction

Although insurance has traditionally been built on the assumption of independence between variables (for example, claim counts are assumed to be independent on the size of claims in several literature among other variables in the insurance industry) and the law of large numbers has governed the determination of premiums, the increasing complexity of insurance and reinsurance products has led recently to increased actuarial interest in the modelling of dependent risks, Wang and Dhaene (1998) and Embrechets et al. (2002).

The dependence between the claim frequency and the class occupied by an insured has been mentioned by many authors. For instance, Denuit et al. (2007) assumed that the distribution of the number of claims is related to the risk classes possessed in multi-event Bonus-Malus scales. They also mentioned the dependence between the bonus class and annual expected claim frequency.

In this paper, Zhao and Zhou (2014) were of the view that the current class occupied by a policyholder depends on his or her claims history and therefore proposed a model for the dependence between the current bonus class occupied by the policyholder and the claim numbers in an insurance period using a bivariate copula function. In the next sub-sections under this topic, we will review this topic and conclude on how this topic has informed the research area of this thesis.

2.1.2 Model Specification

Variables Definition

Consider the following variable definitions:

1. n: the number of policies,
2. i=1,2,...,n : the observed insureds,
3. $C_{i,t}$: the bonus class occupied by ith insured at the beginning of period t,
4. $N_{i,t}$: the number of claims reported by ith insured for the t-th time period,
5. S: the total number of classes,
6. b: the level premium for the classs i.e. $b = (b_1, b_2, ..., b_s)'$,
7. $d_{i,t}$: the length of period that the ith policyholder stayed within a specific policy characteristics (risk exposure) at time t. For example $d_{i,t}$ is usually 1 when marital status of the insured i remains unchanged.

Marginal Distributions

We model claim counts for each period t as $N_{i,t} \sim Poisson(\lambda_{i,t})$ where the model parameter is a function given by; $\lambda_{i,t} = d_{i,t}exp(\alpha x'_{i,t} + \beta)$ with the necessary information on the observed insured i summarized in x which is called a covariate. The model is given by:

$$P(N_{i,t} = n_{i,t}) = \frac{\lambda_{i,t}^{n_{i,t}}}{n_{i,t}!} exp(-\lambda_{i,t}) \tag{2.1}$$

Next, we require the distribution of the current class. To do this, we consider first a reference from Denuit et al. (2007) and Zhao and Zhou (2014). They considered a bonus experience rating systems with six bonus-malus classes (i.e. $S = s = 6$), of which, level 5 is the starting class. A higher class number indicates a higher premium. For a policyholder i, the class $C_{i,t+1}$ in year $t+1$ is a function of class $C_{i,t}$. Hence, recursive relation between subsequent classes is given below:

$$C_{i,t+1} = \begin{cases} max[1, C_{i,t} - 1], & n_{i,t} = 0 \\ \\ min[6, C_{i,t} + 2], & n_{i,t} \geq 1 \end{cases}$$

Since policyholders move from one class to the other class over time, the marginal distribution of $C_{i,t}$ will have an inherent nature according to a transitional probability matrix (Ross, 2014), that is assumed to be associated with claim count. Denote t_{mn} as claim for a policy to be transferred from class m to class n (m,n=1,2,...,6). Then the transitional rule denoted by $T = (t_{mn})_{6\times6}$ for claim count is given by:

$$T = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} \begin{array}{c} \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\ \begin{pmatrix} \{0\} & - & \{1\} & - & \{2\} & \geq\{3\} \\ \{0\} & - & - & \{1\} & - & \geq\{2\} \\ - & \{0\} & - & - & \{1\} & \geq\{2\} \\ - & - & \{0\} & - & - & \geq\{1\} \\ - & - & - & \{0\} & - & \geq\{1\} \\ - & - & - & - & \{0\} & \geq\{1\} \end{pmatrix} \end{array}$$

Next we define the transitional probabilities associated with T above. Denoting $P(N_{i,j} =$

k) by $p_{i,j,k}$ for k=0,1,2 and $1-P(N_{i,j} \leq k-1)$ by $q_{i,j,k}$ for k=1,2,3 where the duration $j = 1,2,...,t-1$ and the period $t = 2,..., T_i$ means that the above transitional rule has a corresponding transitional probability $P_{i,j}$, written as;

$$P_{i,j} = \begin{array}{c c} & \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} & \begin{pmatrix} p_{i,j,0} & - & p_{i,j,1} & - & q_{i,j,2} & q_{i,j,3} \\ p_{i,j,0} & - & - & p_{i,j,1} & - & q_{i,j,2} \\ - & p_{i,j,0} & - & - & p_{i,j,1} & q_{i,j,2} \\ - & - & p_{i,j,0} & - & - & q_{i,j,1} \\ - & - & - & p_{i,j,0} & - & q_{i,j,1} \\ - & - & - & - & p_{i,j,0} & q_{i,j,1} \end{pmatrix} \end{array}$$

Hence, the marginal distribution of $C_{i,t}$ is written as:

$$P(C_{i,t} = c_{i,t}) = (0,0,0,0,1,0)\prod_{j=1}^{t-1} P_{i,j} \tag{2.2}$$

where $(0,0,0,0,1,0)$ is the starting class vector.

Modeling Dependence

Considering a time long period of time, T_i, with the objective to compute limiting probabilities and also, the bonus classes for the insured i which would create a discrete stochastic process $\{C_{i,1}, C_{i,2}, ..., C_{i,T_i}\}$. Intuitively, each subsequent class only takes into account the most recent past class, we can arguably state that, the Markov Property is satisfied and hence, this discrete stochastic process creates a Markov Chain.

$$\begin{aligned} P(C_{i,1}, C_{i,2}, ..., C_{i,T_i}) &= P(C_{i,1} = c_{i,1})\prod_{t=1}^{T_i} P(C_{i,t+1} = c_{i,t+1} \mid C_{i,t} = c_{i,t}) && (2.3) \\ &= P(C_{i,1} = c_{i,1})\prod_{t=1}^{T_i} P(N_{i,t} = n_{i,t} \mid C_{i,t} = c_{i,t}) && (2.4) \end{aligned}$$

In this paper, Zhao and Zhou (2014) mentioned that, most literature at this point, will show that, $P(N_{i,t} = n_{i,t} \mid C_{i,t} = c_{i,t}) = P(N_{i,t} = n_{i,t})$ in (2.3) . This means that, in a

Bonus-Malus system, the distribution of the number of claims is independent of the current risk class. Conversely, this is hard to justify in practice. Indeed, Denuit et al. (2007) justified that the distribution of the number of claims related to the risk class in multi-event bonus-malus scales, Zhao and Zhou (2014) assumed such reverse dependency in their work.

On the one hand, by considering the joint probability distribution function between $N_{i,t}$ and $C_{i,t}$. From (1.2.1), this joint probability distribution can be expressed by:

$$\begin{aligned} P(N_{i,t} = n_{i,t}, C_{i,t} = c_{i,t}) &= P(N_{i,t} \leq n_{i,t}, C_{i,t} \leq c_{i,t}) \\ &- P(N_{i,t} \leq n_{i,t} - 1, C_{i,t} \leq c_{i,t}) \\ &- P(N_{i,t} \leq n_{i,t}, C_{i,t} \leq c_{i,t} - 1) \\ &+ P(N_{i,t} \leq n_{i,t} - 1, C_{i,t} \leq c_{i,t} - 1) \end{aligned}$$

On the other hand, Zhao and Zhou (2014) proposed a bivariate copula to model the dependence between $N_{i,t}$ and $C_{i,t}$ through the cumulative distribution component (first term) of the above probability distribution.

$$P(N_{i,t} \leq n_{i,t}, C_{i,t} \leq c_{i,t}) = C_\delta(F(n_{i,t}), G(c_{i,t})) \tag{2.5}$$

where:

$C_\delta(\cdot,\cdot)$: bivariate copula function with copula parameter δ,

$F(\cdot)$: marginal cumulative distribution function of $N_{i,t}$,

$G(\cdot)$: marginal cumulative distribution function of $C_{i,t}$.

In addition, practically, zero claims has the highest frequency in insurance claims dataset and as a result, the usual Poisson distribution alone cannot model efficiently the number of claims for an insured. In their paper, Zhao and Zhou (2014) captured this in the model by adding an extra parameter $\phi_{i,t}$ to the Poisson distribution making it a Zero-Inflated Poisson distribution. The parameter $\phi_{i,t}$ represents the probability of no claim in the insurance period t for insured i.

Remark 4

$\phi_{i,t}$ depends on a time-varying covariate $z_{i,t}$ and it is given by:

$$\phi_{i,t} = \frac{exp(z'_{i,t}\gamma)}{1+exp(z'_{i,t}\gamma)}.$$

Our Zero-Inflated Poisson distribution becomes:

$$P(N_{i,t} = n_{i,t}) = \begin{cases} \phi_{i,t} + (1-\phi_{i,t})exp(-\lambda_{i,t}), & n_{i,t} = 0 \\ \\ (1-\phi_{i,t})\frac{\lambda_{i,t}^{n_{i,t}}}{n_{i,t}!}exp(-\lambda_{i,t}), & n_{i,t} \geq 1 \end{cases} \tag{2.6}$$

Lastly, from Eq.(2.5), the copula function glues the two distributions so using (2.6) for a cummulative probability distribution in a case where there is no claim and a case when there is at least one claim, the cummulative probability distribution becomes:

$$C_\delta(F(n_{i,t}), G(c_{i,t})) = \begin{cases} \phi_{i,t} + (1-\phi_{i,t})C_\delta(F(0), G(c_{i,t})), & n_{i,t} = 0 \\ \\ (1-\phi_{i,t})C_\delta(F(n_{i,t}), G(c_{i,t})), & n_{i,t} \geq 1 \end{cases} \tag{2.7}$$

2.1.3 Parameter Estimation

Though there is a large literature on how to select a copula for a given dataset, making an appropriate choice of copula that best fit for a given dataset is not an easy task. Zhao and Zhou (2014) in their paper selected the Clayton copula (which has a form belonging to the Archimedian family of copulas) as the bivariate copula function C_δ, since it has been shown to be the only absolutely continuous copula with time-dependent association under a measure in Oakes Cross-Ratio function provided by Oakes (1989). The approach with the Clayton copula proposed in this paper can be similarly applied to other copulas.

The form of a Clayton copula is expressed as:

$$C_\delta(u,v) = (u^{-\delta} + v^{-\delta} - 1)^{-\frac{1}{\delta}}, \delta > 0 \tag{2.8}$$

Given $\{x_{i,t}, z_{i,t}, n_{i,t}, c_{i,t}; i = 1, ..., T_i\}$ together with Eq.(2.8), the model parameters α, β, γ and the copula parameter δ can be estimated by maximizing the likelihood function. Intuitively, the full likelihood function for insured i is given by

$$\begin{aligned} l_i(.) &= P(C_{i,1} = c_{i,1}) \prod_{t=1}^{T_i} P(C_{i,t+1} = c_{i,t+1} \mid C_{i,t} = c_{i,t}) && (2.9) \\ &= P(C_{i,1} = c_{i,1}) \prod_{t=1}^{T_i} P(N_{i,t} = n_{i,t} \mid C_{i,t} = c_{i,t}) && (2.10) \\ &= P(C_{i,1} = c_{i,1}) \prod_{t=1}^{T_i} \frac{(L^1_{i,t})^{\Delta_{i,t}} (L^2_{i,t})^{1-\Delta_{i,t}}}{P(C_{i,t} = c_{i,t})} && (2.11) \end{aligned}$$

where $\Delta_{i,t}$ is an indicator function with:

$$\Delta_{i,t} = \begin{cases} I_{(N_{i,t}=0)} = 1 & if N_{i,t} = 0 \\ \\ 0 & otherwise \end{cases} \qquad (2.12)$$

$$\begin{aligned} L^1_{i,t} &= \phi_{i,t} + (1 - \phi_{i,t})\{C_\delta(F(0), G(c_{i,t})) - C_\delta(F(0), G(c_{i,t} - 1))\} && (2.13) \\ L^2_{i,t} &= (1 - \phi_{i,t})\{C_\delta(F(n_{i,t}), G(c_{i,t})) - C_\delta(F(n_{i,t} - 1), G(c_{i,t}))\} \\ &- C_\delta(F(n_{i,t}), G(c_{i,t} - 1)) + C_\delta(F(n_{i,t} - 1), G(c_{i,t} - 1)). && (2.14) \end{aligned}$$

For example, at time t, when a policyholder is in the lowest class,$C_{i,t} = 1$ and it means that:

$L^1_{i,t} = \phi_{i,t} + (1 - \phi_{i,t})C_\delta(F(0), G(1))$ and

$L^2_{i,t} = (1 - \phi_{i,t})C_\delta(F(n_{i,t}), G(c_{i,t})) - C_\delta(F(n_{i,t} - 1), G(c_{i,t}))$.

For us to complete the estimation of parameters, we will need the first derivative of the Clayton copula with respect to one marginal distribution and also find its corresponding inverse function. Example(1.4.3) in chapter 1 already shows that, finding the first derivative of the Clayton copula means we are looking for the conditional distribution function from a joint distribution. The first derivative of the Clayton copula is given by:

$$C_{u|v}(u|v;\delta) = v_{i,t}^{-(\delta+1)}[v_{i,t}^{\delta} + u_{i,t}^{\delta} - 1]^{(\delta^{-1}+1)}. \tag{2.15}$$

Its corresponding inverse function is given by:

$$C_{y|v}^{-1}(y_{i,t}|v_{i,t};\delta) = [(y_{i,t}^{\frac{-\delta}{(1+\delta)}} - 1)v_{i,t}^{-\delta} + 1]^{\frac{-1}{\delta}}. \tag{2.16}$$

2.1.4 Algorithm for Implementation

In this section, our objective will be to draw paired samples of the random variables $(C_{i,t}, N_{i,t})$ from the dependence models above. Zhao and Zhou (2014), in their paper further implemented (see simulation studies in section 4 of (Zhao & Zhou, 2014)) the algorithm in this section to proof the accuracy of the dependence models above. Below are the algorithm to draw paired samples of $(C_{i,t}, N_{i,t})$:

- Step 1: Draw $h_{i,t}$ from $U[0,1]$.
  If $h_{i,t} < \phi_{i,t}$ set $N_{i,t} \longleftarrow 0$; otherwise continue from step 2 to the end,
- Step 2: Draw T independent $U[0,1]$ random variables $\{Y_{i,t}\}_{t=1}^{T}$ (where T is the total observed time periods) ,
- Step 3: Draw T independent $U[0,1]$ random variables $\{V_{i,t}\}_{t=1}^{T}$ (set $V_{i,1} \longleftarrow 1$, for class five since it the starting class),
- Step 4: Find $U_{i,t} = [(Y_{i,t}^{\frac{-\delta}{(1+\delta)}} - 1)V_{i,t}^{-\delta} + 1]^{\frac{-1}{\delta}}$,
- Step 5: Draw the required paired samples by generating $C_{i,t}$ with probability $V_{i,t}$ from the marginal distribution in Eq.(2.2) and also by generating $N_{i,t}$ from the Poisson distribution with mean λ with probability $U_{i,t}$.

In this work, we implemented the above algorithm to the point of generating the paired samples of $(C_{i,t}, N_{i,t})$ and observed how the dependence models reflected in it over a given period of time. See appendix for the R code to this work.

2.1.5 Summary and Discussion

In this work, we have seen that, in insurance the dependence in claim counting process and Bonus-Malus rate-making class are two risks a rate-making actuary should account for during pricing. Among many other copulas, Clayton copula is used to model such dependence. It it noted that, techniques such as copula derivatives (copula theory and counting process from chapter 1) play an important role in measuring the dependence and sampling from a copula (see R code provided in appendix). The maximum likelihood approach in estimating both model parameters and copula parameters was employed. Justification to the use of these models were done by the authors Zhao and Zhou (2014) through a simulation study.

Whereas in the United Kingdom, each insurer is free to design its own Bonus-Malus System (BMS), the regulatory environments in a country like Switzerland have a government-imposed BMS. Some other countries over time, also move from an old Bonus-Malus system to a new one. For instance Belgium started with a system that had 18 number of classes and later moved to a new one with 23 number of classes (among several other changes). These geographical differences and internal transitions of Bonus-Malus systems among nations informed my research that designs of BMS differ. Furthermore, there are varieties of copulas to select from and choosing the appropriate copula for a given dataset makes it not an easy task. Despite these complexities, rate-making actuaries should account for dependence of these type in their work.

2.2 A Mixed Copula Model for Insurance Claims and Claim Sizes

2.2.1 Introduction

Estimating total loss incurred plays a very important role in pricing non-life insurance contracts and as a result, there has being the need for loss reserving actuaries to first build models that relates to this for any insurance portfolio. A very common approach used by many loss reserving actuaries, basing on the compound Poisson model suggested by Lundberg (1903), is the approach which models the average claim size and the number of claims independently and then defines the loss as the product of these two quantities.

However, the assumption of independence can be too restrictive and lead to a systematic over or under estimation of the policy loss. This makes the independence assumption by Lundberg (1903) not always true. One example to substantiate this point is that, Gschlößl and Czado (2007) in their work, detected a significant dependency between average claim size and number of claims when they analyzed a comprehensive car insurance dataset using a full Bayesian approach.

In their paper, Czado et al.,(2012) proposed a joint model that explicitly allows a dependency between average claim sizes and number of claims. This is achieved by combining marginal distributions for claim frequency and severity with families of bivariate copula. Furthermore, they allowed for more flexibility and generality in the type of dependency by extending the copula-based model to regression model with the help of Generalized Linear Regression Models (GLM). In the next sub-sections under this topic, you will see how this topic is reviewed under dependence modelling with a bivariate copula and how it informs the research area of this thesis.

2.2.2 Model Specification

Variables Definition

Consider the following variable definitions:

1. Y_{i1}: the continuous claims size random variables for the ith $(i = 1, 2, ..., n)$ policyholder,

2. Y_{i2}: the discrete claims count random variables for the ith $(i = 1, 2, ..., n)$ policyholder.

Marginal Distributions

In this sub-section, our goal will be to define the marginal distributions needed for the dependence modeling. With reference to their paper, Czado et al.,(2012) considered a bivariate model where the margins follow generalized linear regression model (GLM) were built. Firstly, we begin by making $Y_{i1} \in \mathbb{R}^+, i = 1, 2, ..., n$ depend on a covariate $x_i \in \mathbb{R}^p$ and assume that the claims size is a Gamma distributed variable. Under GLM (with also the link function) we will have:

$$Y_{i1} \sim Gamma(\mu_{i1}, v^2) \qquad with\ \ ln(\mu_{i1}) = x_i^{'}\alpha \tag{2.17}$$

The density of $Gamma(\mu_{i1}, v^2)$ is specified by:

$$g_1(y_{i1}|\mu_{i1}, v^2) := \frac{1}{\Gamma(\frac{1}{v^2})} y_{i1}^{1/v^2 - 1} e^{-\frac{y_{i1}}{\mu_{i1} v^2}}$$

with $\mu_{i1} := E[Y_{i1}]$ and $Var[Y_{i1}] = \mu_{i1}^2 v^2$. $G_1(.|\mu_{i1}, v)$ denotes the cumulative distribution function of Y_{i1}. We further assumed that, the parameter v is known and does not need to be estimated in the joint regression model. Also, we assume that the parameter v to be a dispersion parameter.

Secondly we let $Y_{i2} \in \mathbb{N}_0, i = 1, 2, ..., n$ be independent count random variables with covariate $z_i \in \mathbb{R}^q$, the Poisson GLM (with also the link function) is given by:

$$Y_{i2} \sim Poisson(\mu_{i2}) \qquad with \;\; ln(\mu_{i2}) = ln(e_i) + z_i^{'}\beta \tag{2.18}$$

The density of $Poisson(\mu_{i2})$ is specified by:

$$g_2(y_{i2}|\mu_{i2}) = \begin{cases} 0 & y_{i2} < 0 \\ \\ \frac{1}{y_{i2}!}\mu_{i2}^{y_{i2}}e^{-\mu_{i2}} & y_{i2} = 0,1,2,... \end{cases}$$

and the corresponding cumulative distribution function is $G_2(\cdot|\mu_{i2})$.

Modeling Dependence

At this point, either we have the form of the joint distribution already available or we have a specific, deemed to be approapriate and chosen to model the joint behaviour (by Sklar's theorem) of the two variables mentioned in the previous sub-section. From Eq.(1.4) in chapter 1, a joint distribution function of Y_{i1} and Y_{i2}, $F(y_{i1}, y_{i2})$ may be expressed as:

$$F(y_{i1}, y_{i2}|\mu_{i1}, v, \mu_{i2}) = C(G_1(y_{i1}|\mu_{i1}, v), G_2(y_{i2}|\mu_{i2})). \tag{2.19}$$

Czado et al.,(2012) confirmed that, because the bivariate Gaussian copula $C(\cdot,\cdot|\rho)$ is a well investigated and directly interpretable function in terms of correlation parameter, they selected this copula. This joint distribution function is a mix of copula and regression.

$$F(y_{i1}, y_{i2}|\mu_{i1}, v, \mu_{i2}, \rho) = C(u_{i1}, u_{i2}|\rho) = \phi_2\{\phi^{-1}(u_{i1}), \phi^{-1}(u_{i2})|\Gamma\} \tag{2.20}$$

where:

1. $u_{i1} := G_1(\cdot|\mu_{i1}, v^2)$,

2. $u_{i2} := G_2(\cdot|\mu_{i2})$,

3. $\phi(\cdot)$:= univariate standard normal cumulative distribution function,

4. $\phi_2(\cdot,\cdot|\Gamma)$:= bivariate normal cumulative distribution function with covariance Γ,

5. $\Gamma := \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ where ρ is the Pearson correlation between the two normal scores above which are denoted by $q_{i1} := \phi^{-1}(u_{i1})$ and $q_{i2} := \phi^{-1}(u_{i2})$.

Next, since we will need a likelihood function to estimate the model parameters, we first have to find the corresponding joint density function of Y_{i1} and Y_{i2}, $f(y_{i1}, y_{i2}|\mu_{i1}, v, \mu_{i2}, \rho)$. This can be achieved by applying the copula derivatives techniques to Eg.(2.20). Also, Eq.(6.9) in Song and Song (2007) presented a form of the required density as:

$$f(y_{i1}, y_{i2}|\mu_{i1}, v, \mu_{i2}, \rho) = g_1(y_{i1}|\mu_{i1}, v^2)[C_1'(u_{i1}, u_{i2}|\rho) - C_1'(u_{i1}, u_{i2}^-|\rho)] \tag{2.21}$$

where:

1. $C_1'(u_{i1}, u_{i2}|\rho) := \frac{\partial}{\partial u_1} C(u_1, u_{i2}|\rho)|_{u_1=u_{i1}}$,

2. $C_1'(u_{i1}, u_{i2}^-|\rho) := \frac{\partial}{\partial u_1} C(u_1, u_{i2}^-|\rho)|_{u_1=u_{i1}}$,

3. $u_{i2}^- := G_2(y_{i2} - 1|\mu_{i2})$.

At this point, the authors Czado et al.,(2012) recommended we consider the following notations and model representations for easiness. By denoting $q_1 := \phi^{-1}(u_1)$ and the vector of q_{i1} and q_{i2} partitioned as $x \equiv \left[\begin{array}{c} q_1 \\ \hline x_2 \end{array}\right]$ for the ith variable we have:

$$\begin{aligned}
C_1'(u_{i1}, u_{i2}|\rho) &= \frac{\partial}{\partial u_1} C(u_1, u_{i1}|\rho) \\
&= \frac{\partial}{\partial u_1} \frac{1}{2\pi\sqrt{|det(\Gamma)|}} \int_{-\infty}^{q_1} \int_{-\infty}^{q_{i2}} exp\{-\frac{1}{2} x' \Gamma^{-1} x\} dx \\
&= \frac{1}{2\pi\sqrt{|det(\Gamma)|}} \int_{-\infty}^{q_{i2}} exp\{-\frac{1}{2}(q_1, x_2)\Gamma^{-1} \begin{pmatrix} q_1 \\ x_2 \end{pmatrix}\} dx_2 (\frac{\partial}{\partial u_1} q_1).
\end{aligned}$$

Now since: $\frac{\partial}{\partial u_1} q_1 = \frac{\partial}{\partial u_1} \phi^{-1}(u_1) = \frac{1}{\phi(\phi^{-1}(u_1))} = \frac{1}{\phi(q_1)} = \frac{1}{\frac{1}{\sqrt{2\pi}} e^{-q_1^2/2}}$,

$$C_1^{'}(u_{i1}, u_{i2}|\rho) = \frac{\sqrt{2\pi} e^{-q_1^2/2}}{2\pi\sqrt{|det(\Gamma)|}} \int_{-\infty}^{q_{i2}} exp\{-\frac{1}{2}(q_1, x_2)\Gamma^{-1}\begin{pmatrix} q_1 \\ x_2 \end{pmatrix}\} dx_2.$$

Since $det(\Gamma) = 1 - \rho^2$ we have:

$$C_1^{'}(u_{i1}, u_{i2}|\rho) = \frac{1}{\sqrt{2\pi|(1-\rho^2)|}} \int_{-\infty}^{q_{i2}} exp\{-\frac{1}{2(1-\rho^2)}(q_1\rho - x_2)^2\} dx_2. \quad (2.22)$$

By transforming $x_2 = z\sqrt{1-\rho^2} + \rho q_1$, it is clear that the above (Eq. 2.22) reduces to a cumulative distribution function of a standard normal distribution given by:

$$\begin{aligned} C_1^{'}(u_{i1}, u_{i2}|\rho) &= \frac{1}{2\pi} \int_{-\infty}^{\frac{q_{i2} - q_1\rho}{\sqrt{1-\rho^2}}} exp\{-\frac{1}{2}z^2\} dz \\ &= \phi(\frac{\phi^{-1}(u_{i2}) - \rho\phi^{-1}(u_1)}{\sqrt{1-\rho^2}}) \\ &=: D_\rho(u_1, u_{i2}). \end{aligned}$$

Equivalently we get:

$$C_1^{'}(u_{i1}, u_{\bar{i2}}|\rho) = \phi(\frac{\phi^{-1}(u_{\bar{i2}}) - \rho\phi^{-1}(u_1)}{\sqrt{1-\rho^2}}) = D_\rho(u_1, u_{\bar{i2}}).$$

Now the model in Eq.(2.21) can be represented as:

$$\begin{aligned} f(y_{i1}, y_{i2}|\mu_{i1}, \nu, \mu_{i2}, \rho) &= g_1(y_{i1}|\mu_{i1}, \nu^2)[D_\rho(G_1(y_{i1}|\mu_{i1}, \nu^2), G_2(y_{i2}|\mu_{i2})) \\ &- D_\rho(G_1(y_{i1}|\mu_{i1}, \nu^2), G_2(y_{i2} - 1|\mu_{i2}))]. \end{aligned} \quad (2.23)$$

Remark 5

Czado et al.,(2012) added that, Eq.(2.23) may further be expressed as:

$$f(y_{i1}, y_{i2}|\mu_{i1}, \nu, \mu_{i2}, \rho) = g_1(y_{i1}|\mu_{i1}, \nu^2) f_{Y_{i2}|Y_{i1}}(y_{i2}|y_{i1}, \mu_{i1}, \nu, \mu_{i2}, \rho)$$

where $f_{Y_{i2}|Y_{i1}}(y_{i2}|y_{i1}, \mu_{i1}, \nu, \mu_{i2}, \rho)$ *is the conditional density of* Y_{i2} *given* Y_{i1}.

Remark 6

The last term in Eq.(2.23) reduces to zero when there is no claim (thus when $Y_{i2} = 0$). Proof to this is shown in the appendix of the original paper.

Thus more specifically, $Eq.(2.23)$ may be rewritten as:

$$f(y_{i1}, y_{i2}|\mu_{i1}, v, \mu_{i2}, \rho) = \begin{cases} g_1(y_{i1}|\mu_{i1}, v^2) D_\rho(G_1(y_{i1}|\mu_{i1}, v^2), G_2(y_{i2}|\mu_{i2})) \\ \qquad if \quad y_{i2} = 0 \\ \\ g_1(y_{i1}|\mu_{i1}, v^2) \Big\{ D_\rho(G_1(y_{i1}|\mu_{i1}, v^2), G_2(y_{i2}|\mu_{i2})) \\ -D_\rho(G_1(y_{i1}|\mu_{i1}, v^2), G_2(y_{i2} - 1|\mu_{i2}) \Big\} \\ \qquad if \quad y_{i2} \geq 1 \end{cases} \tag{2.24}$$

Remark 7

Eq.(2.24) can be used to find the conditional probability mass of Y_{i2} given Y_{i1} by dividing the joint distribution by $g_1(y_{i1}|\mu_{i1}, v, \mu_{i2})$.

2.2.3 Parameter Estimation

Method of Maximum Likelihood Estimation

Consider the set of unknown parameter at this point is defined by $\theta := (\alpha', \beta', \gamma')$, where the regression parameters, α in Eq. (2.17) and β in Eq.(2.18) whilst γ is given by the Fisher's z-transformation $\gamma = \frac{1}{2} ln(\frac{1+\rho}{1-\rho})$ for $\gamma \in \mathbb{R}$.

To ascertain the log-likelihood function, we begin by defining the design matrices associated with the covariate vectors to y_{i1} and y_{i2} including the intercepts respectively. Thus:

$$X := (x_1, ..., x_n)^T$$
$$Z := (z_1, ..., z_n)^T$$

Next we build an indexed set, $J := (i|i = 1,2,...,n; y_{i2} \geq 1)$ which intend to help us make a subset all observations with $y_{i2} \geq 1$. Correspondingly, X_J and Z_J are the design matrices restricted to the set J. Now the likelihood function conditional on $y_{i2} \geq 1 \;\; ,\forall i \in J$ can be given as:

$$L^c(\theta|y,X_J,Z_J) = \prod_{i\in J} \frac{f(y_{i1},y_{i2}|\mu_{i1},\nu,\mu_{i2},\rho)}{[1-g_2(0,\mu_{i2})]} \tag{2.25}$$

The conditional log likelihood is given by:

$$l^c(\theta|y,X_J,Z_J) = logL^c(\theta|y,X_J,Z_J)$$

$$\begin{aligned} = &\; -\sum_{i\in J} ln\Big\{1-g_2(0,\mu_{i2})\Big\} + \sum_{i\in J} ln\Big\{g_1(y_{i1}|\mu_{i1},\nu^2)\Big\} \\ + &\; \sum_{i\in J} ln\Big\{D_\rho(G_1(y_{i1}|\mu_{i1},\nu^2),G_2(y_{i2}|\mu_{i2})) \\ - &\; D_\rho(G_1(y_{i1}|\mu_{i1},\nu^2),G_2(y_{i2}-1|\mu_{i2}))\Big\} \end{aligned} \tag{2.26}$$

Method of Maximization by Parts

In the maximum likelihood estimation method, there is the need to solve the score equation at the extreme points and verify if those points gives the maximum likelihood. This verification requires the need to compute second order derivatives of the full likelihood function. When the full likelihood is so complicated, in other words, in a case of a high-dimensional full likelihood function, obtaining and verifying maximum points becomes unmanageable.

Song et al.(2005) after mentioning some approaches that bypasses the above problem and their challenges, proposed a new algorithm called the Maximization by Parts (MBP). This algorithm strategically separate the part of the full likelihood function with easily computed second-order derivatives from the remaining part which is more difficult when finding the second-order derivative. The remaining part serves as a residual part of the score equation to correct and improve the efficiency of estimation. The second-order derivative is not needed in this remaining part .

Similarly, since the full likelihood in Eq.(2.23) under our high-dimensional mixed copula regression model above would be a complex task, we first decompose the function into two parts. The first part which is straightforward in deriving the second-order derivative and the remaining part is used to update the solution of the first part to arrive at an efficient estimator of θ.

To begin with:

1. decompose the unkown parameter $\theta := (\alpha', \beta', \gamma') \in \mathbb{R}^{p+q+1}$ into $\theta := (\theta_1', \theta_2')$ with $\theta_1' = (\alpha', \beta') \in \mathbb{R}^{p+q}$ and $\theta_2' = \gamma \in \mathbb{R}$ and letting,
2. $l_m^c(\theta_1')$ must contain the marginal part of the conditional log-likelihood and is independent of γ,
3. $l_d^c(\theta_1', \gamma)$ must contain the copula part of the conditional log-likelihood and depends on γ.

From the log-likelihood Eq.(2.26), we can simply define:

$$l_m^c(\theta_1) := ln(L_m^c(\theta_1')) := -\sum_{i \in J} ln(1 - e^{-\mu_{i2}}) + \sum_{i \in J} ln(g_1(y_{i1}|\mu_{i1}, v^2))$$

$$\begin{aligned} l_d^c(\theta_1, \gamma) := ln(L_d^c(\theta_1, \gamma)) \quad &:= \quad \sum_{i \in J} ln\{D_\rho(G_1(y_{i1}|\mu_{i1}, v^2), G_2(y_{i2}|\mu_{i2})) \\ &- \quad D_\rho(G_1(y_{i1}|\mu_{i1}, v^2), G_2(y_{i2} - 1|\mu_{i2}))\} \end{aligned}$$

Next we apply the MBP algorithm by Song et al.(2005) to determine the MLE of θ.

$$l^c(\theta) = ln(L^c(\theta)) = l_m^c(\theta_1) + l_d^c(\theta_1, \gamma) \tag{2.27}$$

For the MBP algorithm, we need the score functions of $l_m^c(\theta_1)$ and $l_d^c(\theta_1,\gamma)$. By differentiation, it follows that;

$$
\begin{aligned}
\frac{\partial}{\partial\alpha} l_m^c(\theta_1) &= \sum_{i\in J} \frac{\partial ln(g_1(y_{i1}|\mu_{i1},v^2))}{\partial\alpha} = \sum_{i\in J} \frac{1}{g_1(y_{i1}|\mu_{i1},v^2)} \frac{\partial g_1(y_{i1}|\mu_{i1},v^2)}{\partial\mu_{i1}} \frac{\partial\mu_{i1}}{\partial\alpha} \\
&= \sum_{i\in J} \frac{1}{g_1(y_{i1}|\mu_{i1},v^2)} \frac{1}{(\mu_{i1})^2 v^2} (y_{i1}-\mu_{i1})\mu_{i1} x_i \\
&= \frac{1}{v^2} \sum_{i\in J} x_i \mu_{i1})^{-1} (y_{i1}-\mu_{i1})
\end{aligned}
$$

Similarly:

$$
\frac{\partial}{\partial\beta} l_m^c(\theta_1) = -\sum_{i\in J} z_i \frac{\mu_{i1}}{e^{\mu_{i1}}-1}
$$

Using the second order derivative, we arrived at a variance function by the Fisher Information given by:

$$
\begin{aligned}
\zeta_m^c(\theta_1) &:= -m^{-1} E[\frac{l_m^c(\theta_1)}{\partial\theta_1\theta_1'}] \\
&= m^{-1} \begin{pmatrix} \frac{1}{v^2}\sum_{i\in J} x_i x_i' & 0_{p\times q} \\ 0_{q\times p} & \sum_{i\in J} z_i \frac{\mu_{i2}(e^{\mu_{i2}}-1-\mu_{i2}e^{\mu_{i2}}}{(e^{\mu_{i1}}-1)^2} z_i' \end{pmatrix}.
\end{aligned}
$$

Where m represents the number of elements in J. For us, to compute the score function of the dependency part $l_m^c(\theta_1,\gamma)$ we compute:

$$
\frac{\partial}{\partial\theta} l_d^c(\theta_1,\gamma) = (\frac{\partial}{\partial\alpha} l_d^c(\theta_1,\gamma), \frac{\partial}{\partial\beta} l_d^c(\theta_1,\gamma), \frac{\partial}{\partial\gamma} l_d^c(\theta_1,\gamma))' \tag{2.28}
$$

The above can be achieved when we find each partial derivative (See lemma in Appendix of the original article). Let us first consider the following definitions subsequently used in the partial derivatives.

1. $G_{i1} := G_1(y_{i1}|\mu_{i1}, v)$,

2. $G_{i2} := G_2(y_{i2}|\mu_{i2})$,

3. $G_{i2}^{-} := G_2(y_{i2} - 1|\mu_{i2})$,

4. $d_\rho(u_1, u_{i2}) := \phi(\frac{\phi^{-1}(u_2) - \rho\phi^{-1}(u_2)}{\sqrt{1-\rho^2}})$ where $\phi(\cdot)$ denotes the density of the standard normal distribution.

Below are the partial derivatives which forms the components of Eq.(2.28)

$$\frac{\partial}{\partial\alpha} l_d^c(\theta_1, \gamma) = \sum_{i \in J} \frac{d_\rho(G_{i1}, G_{i2}) - d_\rho(G_{i1}, G_{i2}^{-})}{D_\rho(G_{i1}, G_{i2}) - D_\rho(G_{i1}, G_{i2}^{-})} \frac{G_{i1}^{*} - G_{i1}}{\phi(\phi^{-1}(G_{i1}))} \frac{-\rho}{\sqrt{(1-\rho^2}} x_i.$$

$$\begin{aligned} \frac{\partial}{\partial\beta} l_d^c(\theta_1, \gamma) &= \sum_{i \in J} \frac{1}{D_\rho(G_{i1}, G_{i2}) - D_\rho(G_{i1}, G_{i2}^{-})} \left\{ d_\rho(G_{i1}, G_{i2}^{-}) \frac{g_2(y_{i2} - 1|\mu_{i2})}{\phi^{-1}(G_{i2}^{-})} \right. \\ &- \left. d_\rho(G_{i1}, G_{i2}^{-}) \frac{g_2(y_{i2}|\mu_{i2}}{\phi^{-1}(G_{i2}^{-})} \right\} \frac{\mu_{i2}}{\sqrt{(1-\rho^2}} z_i \end{aligned}$$

$$\begin{aligned} \frac{\partial}{\partial\gamma} l_d^c(\theta_1, \gamma) &= \sum_{i \in J} \frac{1}{D_\rho(G_{i1}, G_{i2}) - D_\rho(G_{i1}, G_{i2}^{-})} \left\{ d_\rho(G_{i1}, G_{i2}) [\rho\phi^{-1}(G_{i2} - \phi^{-1}(G_{i1})] \right. \\ &- \left. d_\rho(G_{i1}, G_{i2}^{-}) [\rho\phi^{-1} G_{i2}^{-} - \phi^{-1}(G_{i1})] \right\} \frac{1}{\sqrt{(1-\rho^2}} \end{aligned}$$

However, in their paper, Song et al.(2005) imposed some common regularity conditions as well as information dominance (see condition (B) on page 1148 of Song et al.(2005)) to achieve convergence of an MBP-algorithm. Empirical evidence showed that an initial MBP algorithm based on Eq.(2.27) does not satisfy information dominance, so there is a need to modify the initial decompositon (Czado et al., 2012). For this we expand both components in our conditional likelihood Eq.(2.25).

To complete the expansion, first define this newly expanded likelihood $L^*(\theta)$ with its new decomposition as:

$$L^*(\theta) \;:=\; L^c(\theta)\frac{L_w(\theta_1)\prod_{i\in J} g_2(y_{i2}|\mu_{i2})}{L_w(\theta_1)\prod_{i\in J} g_2(y_{i2}|\mu_{i2})} \tag{2.29}$$

$$= L^c_m(\theta_1)L^c_w(\theta_1)\prod_{i\in J} g_2(y_{i2}|\mu_{i2})\frac{L^c_d(\theta_1,\gamma)}{L^c_w(\theta_1)\prod_{i\in J} g_2(y_{i2}|\mu_{i2})} \tag{2.30}$$

Next, define the component $L_w(\theta_1)$, that is the component we are using to expand Eq.(2.25) as a bivariate normal log-likelihood:

$$L_w(\theta_1) := \prod_{i\in J} |det(\textstyle\sum_w)|^{-\frac{1}{2}} exp\left\{-\frac{1}{2}(y_i-\mu)'(\textstyle\sum_w)^{-1}(y_i-\mu)\right\}$$

with $y_i = (y_{i1}, y_{i2})'$, $\mu = (\mu_{i1}, \mu_{i2})'$ and $\sum_w = \begin{pmatrix} 1 & \rho w \\ \rho w & 1 \end{pmatrix}$. The correlation ρ_w, is different from the underlying parameter of the copula ρ. Czado et al.,(2012) in their work added that, this can be a pre-specified value estimated from a preliminary analysis of the data.

Denoting the components in Eq.(2.29) by:

1. $L^*_m(\theta_1) := L^c_m(\theta_1)L^c_w(\theta_1)\prod_{i\in J} g_2(y_{i2}|\mu_{i2})$,
2. $L^*_d(\theta_1,\gamma) := \frac{L^c_d(\theta_1,\gamma)}{L^c_w(\theta_1)\prod_{i\in J} g_2(y_{i2}|\mu_{i2})}$.

The expanded log-likelihood of Eq.(2.29), $l^*(\theta)$ takes a decomposition of a form similar to Eq.(2.27). This is shown below:

$$l^*(\theta) := ln(L^*_m(\theta_1)) + ln(L^*_d(\theta_1,\gamma)). \tag{2.31}$$

Considering further the following notations:

1. $l^*_m(\theta_1) := ln(L^*_m(\theta_1))$,

2. $l_d^*(\theta_1,\gamma) := ln(L_d^*(\theta_1,\gamma))$.

we have:

$$\begin{aligned} l_m^*(\theta_1) &:= l_m^c(\theta_1) + ln(\prod_{i\in J}) + ln(L_w(\theta_1)) \\ l_d^*(\theta_1,\gamma) &:= l_d^c(\theta_1) - ln(\prod_{i\in J}) - ln(L_w(\theta_1)) \end{aligned}$$

At this point, finding the first and second order derivatives of $l_m^*(\theta_1)$ and $l_d^*(\theta_1,\gamma)$ are now easy. The Fisher Information corresponding to $l_m^*(\theta_1)$ is given by:

$$\begin{aligned} I_m^*(\theta_1) &:= -m^{-1}E[\frac{\partial^2}{\partial\theta_1\partial\theta_1'}l_m^*(\theta_1)] \\ &= I_m^c(\theta_1) + m^{-1}\begin{pmatrix} 0_{p\times q} & 0_{p\times q} \\ 0_{p\times q} & \sum_{i\in J}\mu_{i2}z_iz_i' \end{pmatrix} \\ &+ m^{-1}\sum_{i\in J}\begin{pmatrix} x_i & 0_p \\ 0_q & \mu_{i2} \end{pmatrix}\begin{pmatrix} \mu_{i1} & 0 \\ 0 & z_i \end{pmatrix}(\sum_w)^{-1}\begin{pmatrix} \mu_{i1} & 0 \\ 0 & z_i \end{pmatrix}\begin{pmatrix} x_i & 0_p \\ 0_q & \mu_{i2} \end{pmatrix}' \end{aligned}$$

From Eq.(2.27):

$$\begin{aligned} l^c(\theta) &= l_m^c(\theta_1) + l_d^c(\theta_1,\gamma) \\ &= l_m^*(\theta_1) + l_d^*(\theta_1,\gamma) \end{aligned}$$

and hence we will have the same score functions:

$$\begin{aligned} \frac{\partial l_m^c(\theta)}{\partial\theta_1} &= \frac{\partial l_m^c(\theta_1)}{\partial\theta_1} + \frac{\partial l_d^c(\theta_1,\gamma)}{\partial\theta_1} \\ &= \frac{\partial l_m^*(\theta_1)}{\partial\theta_1} + \frac{\partial l_d^*(\theta_1,\gamma)}{\partial\theta_1}. \end{aligned}$$

2.2.4 Algorithm for Implementation: Poisson-Gamma Regression Model

Part 1: Algorithm to sample correlated pairs from a Poisson-Gamma Regression Model

- Step 1: Generate the vector of means given by $\mu_{i1} = exp(\alpha_1 + x_{i1}\alpha_2)$ and $\mu_{i2} = exp(\beta_1 + z_{i2}\beta_2)$,

- Step 2: Draw y_{i1} from a Gamma(μ_{i1}, ν) distribution,

- Step 3: Calculate $p_k = f_{Y_{i2}|Y_{i1}}(y_{i2} = k|y_{i1}, \mu_{i1}, \nu, \mu_{i2}, \rho)$ for $k = 0, 1, ..., k^*$, where $p_{k^*} \geq \varepsilon$ and $p_{k^*} < \varepsilon$, $\varepsilon \in (0,1)$, here the copula parameter ρ, is pre-specified and will be updating in the MBP part of the algorithm,

- Step 4: Draw y_{i2} from {0,1,...,k*} with $P(Y_{i1} = k) = p_k$ for $k \in$ {0,1,...,k*}.

Part 2: The MBP part of the algorithm to estimate model parameters $\theta := (\alpha', \beta', \gamma')$

- Step 1: Using the equations in (2.17) and (2.18), estimate the regression coefficients α and β by the method of likelihood estimation (thus we find $[\hat{\alpha}_1', \hat{\beta}_1']$),

- Step 2: Set the initial value $\theta_1^0 \longleftarrow [\hat{\alpha}_1', \hat{\beta}_1']$,

- Step 3: Using the method of bisection, obtain γ from $\frac{\partial l_d^c(\theta_1^0,\gamma)}{\partial \gamma} = 0$ and set the intial value $\gamma^0 \longleftarrow \gamma$,

- Step 4: Set $\rho_w \longleftarrow \frac{e^{2\gamma^0}-1}{e^{2\gamma^0}+1}$,

- Step 5: For $(k = 1, 2, 3, ...)$ be updating the estimates of the regression coefficients by a step of Fisher scoring, thus:

$$\theta_1^k = \theta_1^{k-1} + \{I_m^*(\theta_1^{k-1})\}^{-1}\left(\frac{\partial l^c(\theta)}{\partial \theta_1}|_{\theta_1=\theta_1^{k-1},\gamma=\gamma^{k-1}}\right).$$

  Next, we solve the equation $\frac{\partial l_d^c(\theta^k,\gamma)}{\partial \gamma} = 0$ using bisection to obtain new γ^k

- Step 6: Set $\rho_w \longleftarrow \frac{e^{2\gamma^0}-1}{e^{2\gamma^0}+1}$,

- Step 7: Stop when $||\theta^k - \theta^{k-1}||_\infty < 10^{-6}$ is met. Output $\theta = [\theta_1', \gamma]'$.

2.2.5 Summary and Discussion

In their work, Czado et al.,(2012) implemented the above algorithm in a simulation studies. They further modeled similar dependence on a dataset which contained information on full comprehenesive car insurance policies in Germany in the year 2000.

During which they focused on allowing for dependence between the joint distribution of the number of claims and the average claim size of each policy. This was to help them estimate expected total loss. For the marginal distributions, the average claim size was modeled with a Gamma distribution and the number of claims was modeled with a Zero-truncated Poisson GLM. They ended up concluding that, the estimated expected total loss using a mixed copula model is about 2% smaller than the estimated expected total loss using the independent regression model.

This paper also followed similar procedure outlined in section 2.1. It confirmed also that, in dependence modeling under counting procresses, one has to find first the univariate models for each variable of interest and next introduce the copula model for the dependence of the variable of interest. This recent review in connection to how my research has been informed shows a new estimation technique (Maximization by Parts) for model parameters and copula parameters when the full likelihood function turns out to be extremely difficult tot estimate parameters. However, the numerical estimation under the lessons learnt in this paper may sometimes appear to numberically intensive if we consider other models other than the Poisson-Gamma Model.

2.3 Moments of the Aggregate Discounted Claims with Dependence Introduced by a FGM Copula

2.3.1 Introduction

In the area of insurance, Boudreault (2003) suggested that, in modeling natural catastrophic events, the total claim amount (or the intensity of the catastrophe) and the time elapsed since the previous catastrophe are dependent. Also in risk theory, Albrecher and Boxma (2004) added that, if claim amounts exceeds a certain threshold, then the parameters of the distribution of the next inter-claim time is modified. Among many other literatures, Biard et al. (2011) also supposed that, the distribution of a claim amount has its parameters modified when several preceding inter-claim times are all greater or all lower than a certain threshold.

In their work, Barges et al. (2011) model the dependence mentioned above between an inter-claim time and its subsequent claim amount with a specific copula and further went to present the first moment, second moment and a generalized mth moment of the aggregate discounted claims with dependence. In the next sub-sections under this topic, we will review again a dependence modelling with a bivariate copula and how it informs the research area of this thesis.

2.3.2 Model Specification

Variables Definition

Consider the following variable definitions:

1. X_i: as the continuous claims size random variable for the ith ($i = 1, 2, \ldots, N$) claim,
2. T_i: the random variable for the time of claim X_i,
3. W_i: the inter-claim time random variable.

Marginal Distributions

In actuarial risk theory, it has been assumed that the claim amount X_i, i= 1,2,... are independent and identically distributed (i.i.d.) random variable (r.v's) and the interclaim times $W_1 = T_1$ and $W_j = T_j - T_{j-1}, j = 2,3,...$ are also i.i.d. r.v's. The r.v's X_i and W_i, i= 1,2,... are classically supposed independent.

Also, we consider a continuous-time compound renewal risk model for an insurance portfolio and we define the compound process of the discounted claims $e^{-\delta T_i}X_i$, i=1,2,... occuring at time T_i, i= 1,2,... by $\underline{Z} = \{Z(t), \quad t \geq 0\}$ with

$$Z(t) = \begin{cases} \sum_i^{N(t)} e^{-\delta T_i} X_i, & N(t) > 0 \\ \\ 0, & N(t) = 0 \end{cases} \tag{2.32}$$

where $\underline{N} = \{N(t), t \geq 0\}$ is an homogeneous Poisson counting process and δ is the instantaneous rate of net interest. This last assumption also implies that X_i, i= 1,2,... are independent from $\underline{N}$.

The assumption of independence between the claim amount X_j and the inter-claim time W_j is relaxed to allow $\{(X_j, W_j), j \in \mathbb{N}^+\}$ to form a sequence of i.i.d. random vectors distributed as the canonical random vectors **(X,W)** in which the components may be dependent.The components of **(X,W)** have marginals f_X and f_W respectively.

Now recalling from Chapter 1 that $\mathbb{P}(W_1 > t) = \mathbb{P}(N(t) = 0)$, also $N(t)$ is a homogeneous Poisson process which means that $N(t) \sim Poisson(\beta t)$

$$\begin{aligned} \mathbb{P}(W_1 > t | W_1 = s) &= \mathbb{P}(N(t) = 0) \\ &= e^{-\beta t} \end{aligned}$$

Since the process from any point in time has the same distribution as the original process (stationary increment property), we have:

$$f_W(s) = \beta e^{-\beta s} \tag{2.33}$$

Modeling Dependence

In their paper, Barges et al. (2011) agreed to have restricted themselves with a specific structure of dependence between an inter-claim time and its subsequent claim amount. By this, it means that, they first focused on a set of copulas that captures light dependence, admits positive as well as negative dependence and also independence between a set of random variables. Next, among all the copulas that have the above specific structure of dependence, the Farlie-Gumbel-Morgenstern (FGM) copula was selected to model dependence between the *i*th claim amount and the *i*th interclaim time was used. This is because it offers the advantage of being mathematically tractable (Cossette et al., 2010). It is also known that the FGM copula is a Taylor approximation of order one of the Frank copula. The FGM copula is defined by:

$$C_\theta^{FGM}(u,v) = uv + \theta uv(1-u)(1-v) \tag{2.34}$$

for $(u,v) \in [0,1]^2$ and where the dependence parameter θ takes value in [-1,1].

From Theorem 1 and Eq.(2.34) , the cumulative distribution function (cdf) for the canonical random vector (X,W) is given as:

$$\begin{aligned} F_{X,W}(x,t) &= C_\theta^{FGM}(F_X(x), F_W(t)) \\ &= F_X(x)F_W(t) + \theta F_X(x)F_W(t)(1-F_X(x))(1-F_W(t)) \end{aligned}$$

It follows from Eq.(1.8) and Eq.(1.9) that the corresponding joint probability density function (pdf) is given by:

$$\begin{aligned} f_{X,W}(x,t) &= c_\theta^{FGM}(F_X(x), F_W(t)) f_X(x) f_W(t) \\ &= (1+\theta(1-2F_X(x)))(1-2F_W(t)) f_X(x) f_W(t) \end{aligned}$$

Estimation of the FGM copula parameter θ may be achieved through the method of maximum likelihood estimation. In their paper, Barges et al. (2011) further focused on computing the moments of $\underline{Z} = \{Z(t), t \geq 0\}$. This is because, if we know how to compute such moments, we can further calculate the total discounted cost of all claims made by time t, the expected value of this discounted cost and variances associated with it. In application, one will be able to find premium related to the risk of an insurance portfolio, easily use the moment matching methods in methods of estimation, determine solvency capital requirements among many others. The subsequent sections in this paper shows how the first moment and the second moment respectively are denoted by $\mu_Z(t)$ and $\mu_Z^{(2)}(t)$ of $\underline{Z} = \{Z(t), t \geq 0\}$ are derived.

First Moment

Our goal here is to derive an explicit fomula for the first moment $\mu_Z(t)$ of $Z(t)$. We assume that $E[X] < \infty$. Conditioning upon arrival for first claim X_1 and considering $0 \leq s \leq t$:

$$\begin{aligned}
\mu_Z(t) &= E_Z[Z(t)] \\
&= E_W[E_Z[e^{-\delta s}X_1 + e^{-\delta s}Z(t-s)]|W_1] \\
&= E_W[e^{-\delta s}E_Z[X_1|W_1 = s] + E_W[e^{-\delta s}E_Z[Z(t-s)|W_1 = s]] \\
&= \int_0^t e^{-\delta s}E[X|W = s]f_W(s)ds + \int_0^t e^{-\delta s}\mu_Z(t-s)f_W(s)ds. \qquad (2.35)
\end{aligned}$$

To be able to simplify Eq.(2.35), we need to simplify some terms. Taking the term $E[X|W=s]$ for now, we have:

$$\begin{aligned}
E[X|W=s] &= \int_0^\infty x f_{X|W=s}(x)dx \\
&= \int_0^\infty x\frac{f_{X,W}(x,w)}{f_W(w)}dx \\
&= \int_0^\infty x c_\theta^{FGM}(F_X(x),F_W(s))f_X(x)dx \\
&= E_X[X]+\theta\int_0^\infty x(2-2F_X(x))(1-2F_W(s))f_X(x)dx \\
&- \theta\int_0^\infty x(1-2F_W(s))f_X(x)dx \\
&= E_X[X][1-\theta(1-2F_W(s))]+\theta[1-2F_W(s)]\int_0^\infty [1-F_X(x)]^2dx \\
&= E_X[X][1-\theta(1-2F_W(s))]+\theta[1-2F_W(s)]E_{X^*}[x^*] \\
&= E_X[X]+\theta(1-2F_W(s))[E_{X^*}[x^*]-E_X[X]] \qquad (2.36)
\end{aligned}$$

where $E_{X^*}(x^*)=\int_0^\infty(1-F_X(x))^2dx$.

Furthermore, simplifying Eq.(2.35) means we should be able to convert the integrals into algebraic equations. We will need to make a Laplace transformation.

Definition 2.3.1 (Laplace Transformation)

Let the function $f(t)$ be defined on $[0,\infty)$, then its Laplace transform $L\{f\}$ is another function $\tilde{f}(s)$, which is defined as:

$$\tilde{f}(s)=L\{f\}:=\int_0^\infty e^{-st}f(t)dt=\lim_{A\to\infty}\int_0^A e^{-st}f(t)dt.$$

Example 2.3.1

Let $f(t)=e^{at}$ for some $a\in\mathbb{R}$. Using the integration,

$$L\{e^{at}\}=\int_0^\infty e^{-st}e^{at}dt=\int_0^\infty e^{-t(a+s)}dt=-\frac{1}{s-a}e^{-t(a+s)}|_0^\infty=\frac{1}{s-a}$$

where $s>a$.

From the example (2.3.1), the Laplace transformation of $f_W(w)$ may be expressed as $\tilde{f}_W(w) = L\{f_W\} = \frac{\beta}{(\beta+s)}$.

Substituting $\tilde{f}_W(w)$, we rewrite Eq.(2.35) as follows:

$$\begin{aligned}
\mu_Z(t) &= \int_0^t \frac{\beta}{(\beta+s)} h(s;\beta+\delta) E_X[X] ds \\
&+ \theta \left[E_{X^*}[x^*] - E_X[X] \right] \int_0^t \frac{2\beta}{2\beta+\delta} h(s;2\beta+\delta) ds \\
&- \theta \left[E_{X^*}[x^*] - E_X[X] \right] \int_0^t \frac{\beta}{\beta+\delta} h(s;\beta+\delta) ds \\
&+ \int_0^t \frac{\beta}{\beta+\delta} h(s;\beta+\delta) \mu_Z(t-s) ds
\end{aligned} \tag{2.37}$$

where the p.d.f of W is $h(t;\beta) = f_W(t) = \beta e^{-\beta t}$, that is an exponential distribution with mean $\frac{1}{\beta}$ and the Laplace transformation of W is $\tilde{h}(t;\beta) = E(e^{-tW}) = \frac{\beta}{\beta+t}$. Also, we used $\delta > \beta$. Taking another Laplace transformation but this time on both of sides of Eq.(2.37) and rearranging the terms is equivalent to:

$$\tilde{\mu}_Z(r) = \frac{\beta E_X[X]}{r(\delta+r)} + \theta \frac{\beta E_{X^*}[x^*]}{r(2\beta+\delta+r)} \tag{2.38}$$

Definition 2.3.2 (Invert Laplace Transform)
If $F(t)$ has the Laplace transform $f(s)$, that is $L\{F(t)\} = f(s)$, then the inverse Laplace transform is defined by $L^{-1}\{f(s)\} = F(t)$ and is unique apart from null functions.

Theorem 10 (Inverse Laplace Transform is Linear)

$$L^{-1}\{af_1(s) + bf_2(s)\} = aL^{-1}\{f_1(s)\} + bL^{-1}\{f_2(s)\}.$$

Proof. For suitable well behaved functions $F_1(t)$ and $F_2(t)$:

$$L\{aF_1(s) + bF_2(s)\} = aL\{F_1(s)\} + bL\{F_2(s)\} = af_1(s) + bf_2(s)$$

Taking the inverse Laplace transform gives:

$$aF_1(s) + bF_2(s) = L^{-1}\{af_1(s) + bf_2(s)\}$$

which is the same as

$$aL^{-1}\{f_1(s)\}+bL^{-1}\{f_2(s)\}=L^{-1}\{af_1(s)+bf_2(s)\}$$

$\square$

The partial fraction of the component $\frac{1}{r(\delta+r)}$ in the first term of Eq.(2.38) is given by:

$$\frac{1}{r(\delta+r)}=\frac{1}{\delta r}-\frac{1}{\delta(\delta+r)}$$

Similar from example (2.3.1) and applying the above partial fractions, we obtain:

$$\begin{aligned}
L^{-1}\{\frac{1}{r-0}\} &= e^0=1\\
L^{-1}\{\frac{1}{\delta+r}\} &= e^{\delta t}\\
L^{-1}\{\frac{1}{r(\delta+r)}\} &= \frac{1}{\delta}(1-e^{\delta t})\\
L^{-1}\{\frac{1}{r(2\beta+\delta+r)}\} &= \frac{1}{(2\beta+\delta)}(1-e^{2\beta+\delta t})
\end{aligned}$$

Hence the required first moment becomes:

$$\mu_Z(t)=\beta E_X[X]\frac{(1-e^{\delta t})}{\delta}+\theta\beta\left(E_{X^*}[x^*]-E_X[X]\right)\frac{1-e^{2\beta+\delta t}}{(2\beta+\delta)} \tag{2.39}$$

Second Moment

Our next goal is to find a formula for the second moment of $Z(t)$, $\mu_Z^{(2)}(t)$ supposing $E[X^i]<\infty$, for $i=1,2$.

$$
\begin{aligned}
\mu_Z^{(2)}(t) &= E_Z[(Z(t))^2] \\
&= E_W[E_Z[(e^{-\delta s}X_1 + e^{-\delta s}Z(t-s))^2]|W_1] \\
&= E_W[e^{-2\delta s}E_Z[X_1^2|W_1 = s] + 2E_W[(e^{-\delta s}X_1 \cdot e^{-\delta s}E_Z[Z(t-s)])|W_1] \\
&+ E_W[e^{-2\delta s}E_Z[Z(t-s)]|W_1 = s] \\
&= \int_0^t e^{-2\delta s}E[X^2|W = s]f_W(s)ds + 2\int_0^t e^{-2\delta s}E[X|W = s]\mu_Z(t-s)f_W(s)ds \\
&+ \int_0^t e^{-2\delta s}\mu_Z^{(2)}(t-s)f_W(s)ds \qquad (2.40)
\end{aligned}
$$

Similarly, we have:

$$
\begin{aligned}
E[X^2|W = s] &= \int_0^\infty x^2 f_{X|W=s}(x)dx \\
&= \int_0^\infty x^2 \frac{f_{X,W}(x,w)}{f_W(w)}dx \\
&= \int_0^\infty x^2 c_\theta^{FGM}(F_X(x), F_W(s))f_X(x)dx \\
&= E_X[X^2] + \theta\int_0^\infty x^2(2-2F_X(x))(1-2F_W(s))f_X(x)dx \\
&- \theta\int_0^\infty x^2(1-2F_W(s))f_X(x)dx \\
&= E_X[X]^2\Big(1-\theta(1-2F_W(s))\Big) \\
&+ \theta\Big(1-2F_W(s)\Big)\int_0^\infty 2x[1-F_X(x)]^2dx \\
&= E_X[X^2] + \theta(1-2F_W(s))\Big(E_{X^*}[x^{*2}] - E_X[X^2]\Big) \qquad (2.41)
\end{aligned}
$$

substituting into Eq.(2.40), we have:

$$\begin{aligned}
\mu_Z^{(2)}(t) &= \int_0^t e^{-2\delta s} E_X[X^2] f_W(s) ds \\
&+ \theta [E_{X^*}[x^{*2}] - E_X[X^2]] \int_0^t e^{-2\delta s}(1 - 2F_W(s)) f_W(w) ds \\
&+ 2\int_0^t e^{-2\delta s} E_X[X] \mu_Z(t-s) ds \\
&+ 2\theta [E_{X^*}[x^*] - E_X[X]] \int_0^t e^{-2\delta s}(1 - 2F_W(s)) f_W(w) \mu_Z(t-s) ds \\
&+ \int_0^t e^{-2\delta s} \mu_Z^{(2)}(t-s) ds \qquad (2.42)
\end{aligned}$$

Similar to the Laplace transformation steps in the first moment:

$$\begin{aligned}
\mu_Z^{(2)}(t) &= \frac{\beta}{(\beta+2\delta)} h(s;\beta+2\delta) E_X[X^2] ds \\
&+ \theta E_{X^*}[x^{*2}] \int_0^t \left(\frac{2\beta}{(2\beta+2\delta)} h(s;2\beta+2\delta) - \frac{\beta}{(\beta+2\delta)} h(s;\beta+2\delta)\right) ds \\
&+ 2\int_0^t \frac{\beta}{(\beta+2\delta)} h(s;\beta+2\delta) E_X[X] \mu_Z(t-s) ds \\
&+ 2\theta([E_{X^*}[x^*] - E_X[X]]) \int_0^t \left(\frac{2\beta}{(2\beta+2\delta)} h(s;2\beta+2\delta)\right. \\
&- \left.\frac{\beta}{(\beta+2\delta)} h(s;\beta+2\delta)\right) \mu_Z(t-s) ds \\
&+ \int_0^t \frac{\beta}{(\beta+2\delta)} h(s;\beta+2\delta) \mu_Z^{(2)}(t-s) ds \qquad (2.43)
\end{aligned}$$

Theorem 11 (Laplace Transform of a Combination of Terms)

Consider the function f defined for all non-negative real numbers:

$$\tilde{f}(r) = \frac{1}{r(\alpha_1+r)(\alpha_2+r)\ldots(\alpha_n+r)}$$

An equivalent combination of partial fractions to f is given by:

$$\tilde{f}(r) = \gamma_0 \frac{1}{r} + \gamma_1 \frac{1}{\alpha_1+r} + \gamma_2 \frac{1}{\alpha_2+r} + \ldots + \gamma_n \frac{1}{\alpha_n+r},$$

where $\gamma_0 = \frac{1}{\alpha_1\alpha_2...\alpha_n}$ *and,* $i = 1,...,n$

$$\gamma_i = -\frac{1}{\alpha_i}\prod_{j=1;j\neq i}^{n}\frac{1}{\alpha_j - \alpha_i}.$$

Also, since the inverse Laplace transform of $\frac{1}{\alpha_i + r}$ *is* $e^{-\alpha_2 t}$, *the inverse* $\tilde{f}$ *is given by:*

$$f(t) = \gamma_0 + \gamma_1 e^{-\alpha_1 t} + \gamma_2 e^{-\alpha_2 t} + ... + \gamma_n e^{-\alpha_n t}.$$

See proof of the above theorem in (Baeumer, 2003).

Using theorem(11), take a Laplace transform on both sides:

$$\begin{aligned}
\tilde{\mu}_Z^{(2)}(r) &= \frac{1}{1-\frac{\beta}{\beta+2\delta}\tilde{h}(r;2\beta+2\delta)}\left(\frac{\tilde{h}(r;2\beta+2\delta)}{r}\frac{\beta}{\beta+2\delta}E_X[X^2]\right.\\
&+ \theta([E_{X^*}[x^{*2}]-E_X[X^2]])(\frac{2\beta}{2\beta+2\delta})(\frac{2\beta}{2\beta+2\delta}\frac{\tilde{h}(r;2\beta+2\delta)}{r}\\
&- \frac{\beta}{2\beta+2\delta}\frac{\tilde{h}(r;2\beta+2\delta)}{r})\\
&+ 2E_X[X]\frac{\beta}{2\beta+2\delta}\tilde{h}(r;2\beta+2\delta)\tilde{\mu}_Z(r)\\
&+ 2\theta[E_{X^*}[x^*]-E_X[X]])(\frac{2\beta}{2\beta+2\delta}\tilde{h}(r;2\beta+2\delta)\\
&- \left.\frac{\beta}{2\beta+2\delta}\tilde{h}(r;2\beta+2\delta))\tilde{\mu}_Z(r)]\right)
\end{aligned} \tag{2.44}$$

This reduces to:

$$
\begin{aligned}
\tilde{\mu}_Z^{(2)}(r) &= \frac{\beta E_X[X^2]}{r(2\delta+r)} + \theta \frac{\beta([E_{X^*}[x^{*2}] - E_X[X^2])}{r(2\beta+2\delta+r)} + 2\frac{\beta E_X[X]}{2\delta+r}\tilde{\mu}_Z(r) \\
&+ 2\theta \frac{\beta[E_{X^*}[x^*] - E_X[X]]}{2\beta+2\delta+r}\tilde{\mu}_Z(r) \\
&= \frac{\beta E_X[X^2]}{r(2\delta+r)} + \theta \frac{\beta([E_{X^*}[x^{*2}] - E_X[X^2])}{r(2\beta+2\delta+r)} \\
&+ 2\frac{\beta E_X[X]}{2\delta+r}\left(\frac{\beta E_X[X]}{2\delta+r} + \theta \frac{\beta[E_{X^*}[x^*] - E_X[X]}{r(2\beta+2\delta+r)}\right) \\
&+ 2\theta \frac{\beta[E_{X^*}[x^*] - E_X[X]]}{2\beta+2\delta+r}\left(\frac{\beta E_X[X]}{2\delta+r} + \theta \frac{\beta[E_{X^*}[x^*] - E_X[X]]}{r(2\beta+2\delta+r)}\right) \\
&= \frac{\beta E_X[X^2]}{r(2\delta+r)} + \theta \frac{\beta([E_{X^*}[x^{*2}] - E_X[X^2])}{r(2\beta+2\delta+r)} + 2\frac{\beta^2(E_X[X])^2}{r(\delta+r)(2\delta+r)} \\
&+ 2\theta \frac{\beta^2 E_X[X](E_{X^*}[x^*] - E_X[X])}{r(2\beta+\delta+r)(2\delta+r)} \\
&+ 2\theta \frac{\beta^2 E_X[X](E_{X^*}[x^*] - E_X[X])}{r(2\beta+2\delta+r)(\delta+r)} \\
&+ 2\theta^2 \frac{\beta^2(E_{X^*}[x^*] - E_X[X])^2}{r(2\beta+\delta+r)(2\beta+2\delta+r)}
\end{aligned} \tag{2.45}
$$

Finally, after inverting the Laplace transformation, we have:

$$
\begin{aligned}
\mu_Z^{(2)}(t) &= \beta E_X[X^2](\frac{1-e^{-2\delta t}}{2\delta}) + \theta\beta([E_{X^*}[x^{*2}] - E_X[X^2])(\frac{1-e^{-(2\beta+2\delta)t}}{2\beta+2\delta}) \\
&+ 2\beta^2 E_X[X^2])(\frac{1}{2\delta^2} - \frac{e^{-\delta t}}{\delta^2} + \frac{e^{-2\delta t}}{2\delta^2}) \\
&+ 2\theta\beta^2 E_X[X][E_{X^*}[x^*] - E_X[X]](\frac{1}{2\delta(2\beta+\delta)} - \frac{e^{-(2\beta+\delta)t}}{(2\beta+\delta)(-2\beta+\delta)} \\
&+ \frac{e^{-2\delta t}}{2\delta(-2\beta+\delta)}) \\
&+ 2\theta\beta^2 E_X[X][E_{X^*}[x^*] - E_X[X]](\frac{1}{\delta(2\beta+\delta)} - \frac{e^{-\delta t}}{\delta(2\beta+\delta)} \\
&+ \frac{e^{-(2\beta+2\delta)t}}{(2\beta+\delta)(2\beta+2\delta)(2\beta+\delta)}) \\
&+ 2\theta^2\beta^2(E_X[X][E_{X^*}[x^*] - E_X[X]])^2(\frac{1}{(2\beta+2\delta)(2\beta+\delta)} \\
&- \frac{e^{-(2\beta+\delta)t}}{\delta(2\beta+\delta)} + \frac{e^{-(2\beta+2\delta)t}}{\delta(2\beta+2\delta)})
\end{aligned} \tag{2.46}
$$

See Barges et al. (2011) for how the generalized version, mth moment was derived.

2.3.3 Summary and Conclusion

In modeling for dependence between risks in the area of insurance, dependence measuring between inter-arrival claim time (which is a component of a counting process) and discounted cost of claims with the F-G-M copula adds to literature in the area of this thesis. Basic theories in the chapter 1 such as the Sklar's theorem and copula densities were useful in this work. The obvious method of parameter estimation will be the moment matching technique.

In section 4 of their paper, Barges et al. (2011) presented some applications of these moments in two major areas of actuarial studies. First in premium calculation and secondly in evaluation of Value at Risk (VaR) by considering the method of moment matching to approximate the distribution of $Z(t)$. For instance it was seen that, when the dependencies is negative (positive), if on average, for a fixed period of time, the time elapsed between each claim decreases (increases), then the size of the claim amount increases (decreases)(Barges et al., 2011) .

However, the concepts of moments and distribution of aggregate claims under actuarial theory for loss reserving can go as far as determining solvency capital requirements and so there is the need for more literature to come out with more and better forms of copulas that may be used when measuring for dependence between risks.

2.4 Modelling Dependence in Insurance Claims Processes with Lévy Copulas

2.4.1 Introduction

An event may give rise to claims of different types in a non-life insurance company. Examples of such events are: a) work-related accident resulting in claims for medical costs and allowance costs, b) Natural peril causing losses in motor and home classes of business. Any form of dependence in the costs has a potential implications for pricing, reserving, solvency, and capital allocation of an insurance company. Each of the claim type creates a stochastic processes. A natural choice in modeling is the compound Poisson process for each of the claim type.

In such situations, the dependence between multiple compound Poisson processes and how to model it becomes of interest. One may want to model the dependence in frequency and that of the dependence in severity. Embrechts et al.(2002), Denuit et al. (2005) and McNeil et al. (2005) remarked that, the reason for such an interest is that, dependence in claims processes can have an impact on both frequency (claim counts) and severity (claim amounts) and this has direct implications on pricing, reserving and capital allocation of an insurance company.

In their paper, Avanzi et al. (2011) first explored the ability of Lévy copulas to allow for a great range of dependence structures in multivariate compound Poisson process modelling and extended relating concepts to model dependence in frequency and severity. The goal of this paper is to use Lévy copulas to describe the dependence between a group of Lévy processes. We present a cross-sectional review to this paper by considering other relating papers such as: Esmaeili et al. (2010) and Velsen (2012) in the rest of this section.

2.4.2 Existing Approaches in Modeling Dependence in Multiple Compound Poisson Processes

Approach 1: The Common Shock Based Approach

This approach is intuitively represented in a common shock representation, for instance see Lindskog and McNeil (2003);Yuen and Wang (2002). In this approach we consider claims occuring at the same time in two or more different classes according to an identical arrival process and or apply a distributional copula to claim sizes occuring simultaneously. The benefits of using this approach are: a) We are able to account for a detailed separate dependence in frequency and dependence in severity, b) Also when we consider over alternative time horizons (consistently), we are able to account for dependence.

Approach 2: The Distributional Copula Approach for Aggregate Claims over a given Time Horizon

In this approach we focus on classes of business rather than the common shock events. We apply a distributional copula to the aggregate claim size of each class at a chosen time horizon to build a multivariate distribution of aggregate claim amounts. Another way to do this is to apply a distributional copula to the aggregate number of claims over a given time horizon, for instance see McNeil et al.,(2015); Bargès et al.,(2009); Bäuerle and Grübel,(2005); Genest and Nešlehová, (2007). Among many other benefits, Avanzi et al.,(2011) added this approach possesses the benefit of relative parsimony in model specification.

Limitations to the two approaches:

1. The common shock based approach is parameter intensive as the number of dimensions increases and this is due to its flexibility,

2. Inferring copula for a different time horizon for the distributional copula approach is generally not possible. This approach also requires sufficient data for the aggregate claim amounts in each class of business for the given time horizon so there is the possibility of inefficient use of data in cases where individual accident information is known.

Approach 3: The Lévy Copulas Approach

This third approach combines the benefits of the two approaches mentioned above. Lévy copulas approach similar to the original copula modeling of dependence except that this modeling approach uses the concept of tail integral instead of cumulative distribution functions. This approach has a superior advantage over the others. For instance, it is time consistent, there is a coherent modeling of dependence in frequency separate to that of the dependence in frequency, and it makes full use of the data. The goal here is to build a multivariate function that joins the marginal tail integrals (i.e. the expected number of losses over a given threshold) of the compound Poisson process for each class of business into a multivariate tail integral which completely specifies the desired multivariate (dependent) compound Poisson process model. The subsequent sections are the theoretical basics needed in this approach.

2.4.3 Dependence Modeling of Multiple Compound Poisson Processes with Lévy Copulas

Consider the following properties for a special stochastic random variable $(X_t)_{t\geq 0}$:

1. $X_0 = 0$ a.s,
2. $X_t - X_s \sim X_{t-s} - X_0 \; \forall s \leq t$, stationary increments,
3. $X_t - X_s \perp\!\!\!\perp \sigma(X_r, r \leq s) \; \forall s \leq t$, independent increments,
4. $\lim_{t \to 0} \mathbb{P}(|X_t - X_0| > \varepsilon) = 0 \; \forall \varepsilon > 0$, continuity in probability.

where '$\sim$' stands for 'same distribution' and '$\perp\!\!\!\perp$' for stochastic independence.

Definition 2.4.1 (Lévy Process)

A Lévy process $X = (X_t)_{t\geq 0}$ is a stochastic process $X_t : \Omega \to \mathbb{R}^d$ satisfying (1-4).

This is to say that X starts at zero, has stationary and independent increments and is continuous in probability.

Definition 2.4.2 (Compound Poisson Process with jump distribution μ and intensity λ)

Let $N = (N_t)_{t\geq 0}$ be a Poisson process with intensity λ and replace the jumps of size 1 by independent iid jumps of random height $H_1, H_2, \ldots$ with values in $\mathbb{R}^d$ and $H_k \sim$ a jump distribution μ. This is a compound Poisson process:

$$C_t = \sum_{k=1}^{N_t} H_k,$$

where $H_k \sim \mu$ are iid and independent of $N = (N_t)_{t\geq 0}$.

Remark 8

Compound Poisson processes are Lévy processes. See proof in Schilling (2014).

Definition 2.4.3 (Lévy Copula - Positive Lévy copula, Tankov, 2003)

For Lévy process with positive jumps, a "positive Lévy copula" is defined to be a function $C : [0,\infty]^d \to [0,\infty]$ which satisfies the following:

1. *$C(u_1, \ldots, u_d)$ is increasing in each component,*
2. *$C(u_1, \ldots, u_d) = 0$ if $u_i = 0$ for any $i = 1, \ldots, d$.*
3. *Evaluating C at ∞ for all components except for the ith component which is evaluated at u produces margins $C_i, i = 1, \ldots, d$, which satisfy $C_i(u) = u$ for all u in $[0,\infty]$.*
4. *For all $(a_1, .., a_d)$, $(b_1, .., b_d) \in [0,\infty]^{d]}$ and with $a_i \leq b_i$, $\sum_{i_1}^{2} \cdots \sum_{i_d=1}^{2} (-1)^{i_1+\cdots+i_d} C(u_1, \ldots, u_d) \leq 0$ where $u_{j1} = a_j$ and $u_{j2} = b_j$ for all $j = 1, \ldots, d$.*

2.4.4 Lévy Copulas and Compound Poisson Process

In this sub section we will consider modeling dependence with Lévy Copulas as mentioned earlier. These modeling steps may be ordered chronologically with points P1-P4.
P1: Let us consider the bivariate compound Poisson process $S = \{S_1, S_2\}$ as our multivariate compound Poisson process. Decompose it into unique and common jumps, so that:

$$\begin{cases} S_1(t) = S_1^{\perp}(t) + S^{\|}(t) \\ \\ S_2(t) = S_2^{\perp}(t) + S_2^{\|}(t) \end{cases} \tag{2.47}$$

This decomposition can be further rewritten as:

$$\begin{aligned} \begin{pmatrix} S_1(t) \\ S_2(t) \end{pmatrix} &= \begin{pmatrix} \sum_{i=1}^{N_1(t)} X_i \\ \sum_{j=1}^{N_2(t)} Y_j \end{pmatrix} \\ &= \begin{pmatrix} S_1^{\perp}(t) + S_1^{\|}(t) \\ S_2^{\perp}(t) + S_2^{\|}(t) \end{pmatrix} \\ &= \begin{pmatrix} \sum_{i=1}^{N_1^{\perp}(t)} X_i^{\perp} + \sum_{i=1}^{N_1^{\|}(t)} X_i^{\|} \\ \sum_{j=1}^{N_2^{\perp}(t)} Y_j^{\perp} + \sum_{j=1}^{N_2^{\|}(t)} Y_j^{\|} \end{pmatrix} \end{aligned} \tag{2.48}$$

where the superscript $\perp$ and superscript $\|$ refers to unique and common respectively and $S_1^{\perp}(t)$ and $S_2^{\perp}(t)$ are independent compound Poisson processes whilst $S_1^{\|}(t)$ and $S_2^{\|}(t)$ are dependent compound Poisson processes whose jumps occur at the same time, for instance see Lindskog and McNeil (2003), Esmaeili et al. (2010) . The three processes $S_1^{\perp}(t)$, $S_2^{\perp}(t)$ and $(S^{\|}(t), S_2^{\|}(t))$ are compound Poisson processes and independent, Esmaeili et al. (2010). However, we have a prior idea that, the joint survival function of the jumps (yet to be explained in a subsequent lemma) may be dependent. For the purpose of clarification and consistency in notation, we will be considering Y is X_2 and X is X_1.

Remark 9

In general, the jump size distributions of $S_1^{\perp}(t)$ and $S_1^{\|}(t)$ are not identical but may be a mixture of jump size distributions (see, for example, Mikosch (2009), Proposition 3.3.4).

P2: Moving further, we introduce the concept of tail integral of Lévy process. This measures its expected number of jumps (above a certain threshold) per unit of time. We can relate this to our compound poisson process by:

$$U_i(x) = \begin{cases} \lambda_i \bar{F}_i(x), & x \in (0,\infty) \\ \\ \infty, & x = 0 \end{cases} \tag{2.49}$$

where $\bar{F}_i(x)$ is the survival function for the jump size $S_i(t)$.
Similarly, we may want to represent the tail integral of a bivariate compound Poisson process $\{S_1, S_2\}$ by:

$$U(X_1,X_2) = \begin{cases} \lambda_i^{\|} \bar{F}(x_1,x_2), & (x_1,x_2) \in (0,\infty)^2 \\ U_1(x_1) & x_1 \in (0,\infty), x_2 = 0 \\ U_2(x_2) & x_1 = 0, x_2 \in (0,\infty) \\ \infty, & (x_1,x_2) = (0,0) \end{cases} \tag{2.50}$$

P3: Furthermore, we introduce our Lévy copula by employing Theorem (1). The main purpose of the copula we are introducing is to couple the marginal tail integrals to the joint tail integral. Let us first consider the definition below:

Theorem 12 (Sklars Theorem for Lévy Copulas)
Let $U = U(\cdot,...,\cdot)$ *be a d-dimensional tail integral with* $U_1(\cdot),...,U_d(\cdot)$ *as marginal tail integrals. then there exist a Lévy copula* $\bar{C}$ *such that:*

$$U(x_1,...,x_2) = \bar{C}(U_1(x_1),...,U_d(x_d)). \tag{2.51}$$

If $U_1(\cdot),...,U_d(\cdot)$ *are continuous on* $[0,\infty]$*, then this Lévy copula is unique.*

The bivariate case is shown below:

$$\bar{C}(U_1(x_1),U_2(x_2)) = U(X_1,X_2) \tag{2.52}$$

P4: In this additional step, Böcker and Klüppelberg (2008) contributed that, since the decomposition S_1 and S_2 into unique and common components stems directly from the Lévy copula (as in **P3**), we need to have the Lévy copula readily available and also specify the Poisson parameters and the jump size distributions as well. Below is a lemma to show how the Lévy copula affects both dependence in frequency and dependence in serverity in a bivariate compound Poisson process :

Lemma 2.4.1. Arrival rate for the common jumps $S_i^{\|}, i = 1,2$ is:

$$\lambda^{\|} = \bar{C}(\lambda_1,\lambda_2), \tag{2.53}$$

The joint survival function for the common jump sizes is:

$$\bar{F}^{\|}(x_1,x_2) = \frac{1}{\lambda^{\|}}\bar{C}(\lambda_1\bar{F}_1(x_1),\lambda_2\bar{F}_2(x_2)) \tag{2.54}$$

The marginal survival function for a common jump size are:

$$\begin{cases} \bar{F}_1^{\|}(x) = \frac{1}{\lambda^{\|}}\bar{C}(\lambda_1\bar{F}_1(x),\lambda_2), \ \ \textit{and} \\ \\ \bar{F}_2^{\|}(x) = \frac{1}{\lambda^{\|}}\bar{C}(\lambda_1,\lambda_2\bar{F}_2(x)) \end{cases} \tag{2.55}$$

Each of the unique jumps in $(S_i^{\perp}(t), i = 1,2)$ arrives at rates:

$$\lambda_i^{\perp} = \lambda_i - \lambda^{\parallel}, i = 1,2 \tag{2.56}$$

The marginal survival function for a unique jump size is:

$$\bar{F}_i^{\perp}(x) = \frac{1}{\lambda_i^{\perp}}(\lambda_i \bar{F}_i(x) - \lambda^{\parallel}\bar{F}_i^{\parallel}(x)), i = 1,2 \tag{2.57}$$

2.4.5 How Lévy Copulas are Built

We will present two methods for building Lévy Copulas in this section.

Method 1: Multivariate Lévy process using Sklar's Theorem

Consider a d-mensional spectrally positive Lévy process with continuous marginal tail integrals. A positive Lévy copula $\bar{C}$ can be constructed as:

$$\bar{C}(u_1, ..., u_d) = U(U_1^{-1}(u_1), ..., U_d^{-1}(u_d)) \tag{2.58}$$

Method 2: Constructing Archimedean families of Copulas

For a function $\phi : [0,\infty] \to [0,\infty]$ with $\phi(0) = \infty$ and $\phi(\infty) = 0$ and a defined inverse $\phi^{-1}(\cdot)$, $\bar{C}(u_1, ..., u_d) = \phi^{-1}(\phi(u_1) + ... + \phi(u_d))$, where the inverse must satisfy: $(-1)^k(\phi^{-1})^k(z) > 0, forz > 0, k = 1, ..., d$ and $(\phi^{-1})^{(k)}$ denotes the k-th derivative of the inverse of $\phi(\cdot)$ with respect to z.

Remark 10

When constructing Archimedean distributional copula, special care is needed in defining the inverse of the generator. The case of Lévy copulas is easier. Archimedean generators of Lévy copulas have a domain of $[0,\infty]$ *and a range of* $[0,\infty]$, *so there is no need for a "pseudo-inverse" (Nelsen, 1999).*

2.4.6 Comparing Dependence Structures among two (2) Lévy Copulas (LC1-2) under the current context

This section presents the facts that Lévy copula allow for a wider range of dependence structures. We will begin by first referring the reader to some examples of Lévy copulas mentioned by Kettler (2006). Also, proof to the Theorem 6.1 in Kallsen and Tankov (2006) provides justification to why most of the examples are Lévy copulas. It is useful to consider the remarks below:

Remark 11 (Lévy Copulas Density)

The bivariate Lévy copulas density may be expressed as:

$$c(u_1,u_2) = \frac{\partial^2}{\partial u_1 \partial u_2} C(u_1,u_2), \tag{2.59}$$

where $u_i = U_i(x_i), \;\; i = 1,2.$

Remark 12 (Expected Number of Common Jumps per unit Time)

In each unit time, the number of common jumps on average is expressed as:

$$\lambda^{\|} = \int_0^{\lambda_2} \int_0^{\lambda_1} c(u_1,u_2) du_1 du_2. \tag{2.60}$$

LC1: Pure Common Shock Lévy Copulas

Assuming independence between common jump sizes, we model dependence in common jump frequencies by first considering the definition below:

Definition 2.4.4

The pure common shock Lévy copulas is given by

$$\begin{aligned} \bar{C}_\delta(u_1,u_2) &= \delta(u_1 \wedge \lambda_1)(u_2 \wedge \lambda_2) \\ &+ [u_1 - \delta_2 \lambda_2 (u_1 \wedge \lambda_1)] I_{u_2=\infty} \\ &+ [u_2 - \delta_1 \lambda_1 (u_2 \wedge \lambda_2)] I_{u_1=\infty} \end{aligned} \tag{2.61}$$

for $0 \leq \delta \leq min(\frac{1}{\lambda_1}, \frac{1}{\lambda_2})$ *and* λ_1, λ_2 *are the Poisson parameters of the bivariate Compound Poisson process, also* δ *is a parameter which will determine the intensity of the common jumps. For independence* $\bar{C}(u_1, u_2) = u_1 u_2$ *we will have:*

$$\begin{aligned} \lambda^{\parallel} &= \bar{C}_\delta(\lambda_1, \lambda_2), \\ &= \delta \lambda_1 \lambda_2. \end{aligned} \tag{2.62}$$

From the above, the dependence structure for Pure Common Shock Lévy Copulas is expressed as:

1. Dependence in common jump frequencies: $\lambda^{\parallel} = \delta \lambda_1 \lambda_2$,

2. Independence in common jump severities. In other words, unique and common jump sizes are all independent and identically distributed within the different processes.

Lemma 2.4.2. The pure common shock Lévy copula Eq.(2.4.4) satisfies the necessary conditions of a positive Lévy copula.

Proof. 1. $\bar{C}_\delta(0, u_2) = \bar{C}_\delta(u_1, 0) = 0$ is satisfied

2. $\bar{C}_\delta(\infty, u_2) = u_2$ and $C_\delta(u_1, \infty) = u_1$ is also satisfied,

3. For all $(a_1, a_2), (b_1, b_2) \in [0, \infty)^2$, and with $a_1 \leq b_1$ and $a_2 \leq b_2$,

$$\begin{aligned} &\bar{C}_\delta(b_1, b_2) - \bar{C}_\delta(a_1, b_2) - \bar{C}_\delta(b_1, a_2) + \bar{C}_\delta(a_1, a_2) \\ &= \delta[(b_2 \wedge \lambda_2) - (a_2 \wedge \lambda_2)][(b_1 \wedge \lambda_1) - (a_1 \wedge \lambda_1)] \geq 0 \end{aligned} \tag{2.63}$$

4. For the case $b_1 = \infty$, $b_2 \in [0, \infty)$ and $(a_1, a_2) \in [0, \infty)^2$,

$$\begin{aligned} &\bar{C}_\delta(b_1, b_2) - \bar{C}_\delta(a_1, b_2) - \bar{C}_\delta(b_1, a_2) + \bar{C}_\delta(a_1, a_2) \\ &= b_2 - a_2 + \delta(a_1 \wedge \lambda_1)[(a_2 \wedge \lambda_2) - (b_2 \wedge \lambda_2] \geq 0 \end{aligned} \tag{2.64}$$

since $\delta(a_1 \wedge \lambda_1) \leq 1$ due to the restriction on δ. All other cases are proven in similar way. □

See proof in section 3.3 of Velsen (2012)

Lemma 2.4.3. A bivariate compound Poisson process with dependence specified by the pure common shock Lévy copula given by definition (2.4.4) with non-zero δ has independent and identically distributed common and independent jump sizes within one process, and independent common jump sizes in both processes (See proof in Avanzi et al. (2011)).

LC2: Clayton Shock Lévy Copulas

In this section, we first recall that a distributional survival copula $\bar{C}$ of $\bar{F}^{\|}(x_1,x_2)$ satisfies:

$$\bar{F}^{\|}(x_1,x_2) = \bar{C}(\bar{F}_1(x_1), \lambda_2\bar{F}_2(x_2)) \tag{2.65}$$

where $\bar{F}_1(x_1) = \bar{F}^{\|}(x_1,0)$ and $\bar{F}_2(x_2) = \bar{F}^{\|}(0,x_2)$.

With reference to a Lévy Clayton Copula (Tankov, 2003) and as already specified in section 2.1.3 definition (2.8) and by substitution in relation to lemma (2.4.1), we have:

$$\bar{F}_1^{\|}(x_1) = \Big(\frac{(\lambda_1\bar{F}_1(x_1))^{-\delta} + \lambda_2^{-\delta}}{\lambda_1^{-\delta} + \lambda_2^{-\delta}}\Big)^{-1/\delta} \equiv \mu_1 \tag{2.66}$$

and

$$\bar{F}_2^{\|}(x_2) = \Big(\frac{(\lambda_2\bar{F}_2(x_2))^{-\delta} + \lambda_1^{-\delta}}{\lambda_1^{-\delta} + \lambda_2^{-\delta}}\Big)^{-1/\delta} \equiv \mu_2 \tag{2.67}$$

and also in the lemma (2.4.1), we have:

$$\bar{C}(\mu_1,\mu_2) = (\mu_1^{-\delta} + \mu_2^{-\delta} - 1)^{-1/\delta} \tag{2.68}$$

which is the Clayton copula. The distributional copula C of $F^{\|}$ takes the form

$$C(\mu_1,\mu_2) = ((1-\mu_1)^{-\delta} + (1-\mu_2)^{-\delta} - 1) + \mu_1 + \mu_2 - 1 \tag{2.69}$$

which collapses to uv for $\delta \downarrow 0$ and to min (u,v) for $\delta \to \infty$. The frequency $\lambda^{\|}$ implied by the Clayton Lévy copula takes the form

$$\lambda^{\|} = (\lambda_1^{-\delta} + \lambda_2^{-\delta})^{-1/\delta} \tag{2.70}$$

which collapses to zero for $\delta \downarrow 0$ and to min (λ_1, λ_1) for $\delta \to \infty$. In summary, for $\delta \downarrow 0$ the Clayton Lévy copula implies $\lambda^{\|} = 0$ and independent components of $S^{\|}$.

In a summary, the dependence structure for Clayton Shock Lévy Copulas is expressed as:

1. Dependence in common jump frequencies: $\lambda^{\|} = (\lambda_1^{-\delta} + \lambda_2^{-\delta})^{\frac{1}{\delta}}$,
2. Dependence in common jump severities by the copula parameter δ.

2.4.7 Parameter Estimation

As mentioned in the paper by Avanzi et al. (2011), fitting a bivariate compound Poisson process means that the Poisson parameters, marginal jump size distribution parameters and Lévy copula parameters are to be estimated simultaneously. Observing a compound Poisson process continuously over a time period is equivalent to observing all jump times and jump sizes in this time interval, for instance see Esmaeili et al. (2010). So for us to begin, we consider a sample of jumps of a positive bivariate compound Poisson process observed up to time $t = T$.

Let n be the total number of claims (jumps) occuring in a time interval of length T. The number of jumps in each class is n_1 and n_2. The number of claims common to both classes is $n^{\|}$ and the number of claims unique to each class is expressed as $n_1^{\perp}$ and $n_2^{\perp}$ respectively. The jump sizes in the first and second components are denoted by $x_1^{\perp}, \ldots, x_{n_1}^{\perp}$ and $y_1^{\perp}, \ldots, y_{n_1}^{\perp}$ respectively, while the sizes of the observed common jumps in both components are denoted by $(x_1^{\|}, y_1^{\|}), \ldots, (x_{n^{\|}}^{\|}, y_{n^{\|}}^{\|})$.

One cannot know in advance which of the jumps resulted from a common shock event so we break the total time interval into M intervals of equal width (an example is month-by-month, or quarterly). This breaking down of total interval must make it easier for one to assume that jumps from two different interval did not occurred as a result of the same common shock event. Below are some of the approaches proposed for the estimation of parameters.

1. By Maximizing the Full Likelihood Based on all Possible Combinations of Number of Jumps within each Interval,
2. By Maximizing all the likelihoods within each interval Based on number of jumps and expected jump Sizes,
3. By Maximizing all the likelihoods within each interval based on number of jumps and maximum jump sizes.

In the first approach, within each given interval (say a month), there can be n_1 jumps in the first process and n_2 in the second process.
This means that $0 \leq n^{\|} \leq min(n_1, n_2)$ and so given a certain number of common jumps ($n^{\|}$), there are:

$$\frac{(max(n_1,n_2))!}{(max(n_1,n_2)-n^{\|})!}\frac{(min(n_1,n_2))!}{(min(n_1,n_2)-n^{\|})!}$$

possibilities of distributing the common jumps over the observed jump sizes as shown in Velsen (2012). Unfortunately this method is not computationally feasible. The second approach has a simultaneous maximization for all parameters and this would be extremely complex when each interval is modelled with different type of Lévy copula. Getting a closed form expressions for the convolutions involved is intensive and may be computationally expensive also. We will be going by the third approach (which can be carried in two different sub-versions).

Moving on to the next step, the natural question that comes up next is what likelihood function must be used. In our first version to third approach, we recall that in Eq.(2.48), the three processes $S_1^{\perp}(t)$, $S_2^{\perp}(t)$ and $(S^{\parallel}(t), S_2^{\parallel}(t))$ are compound Poisson processes and independent, see Esmaeili et al. (2010), where they derived (see Theorem 4.1) and represented the log-likelihood function for the bivariate compound Poisson process by:

$$\begin{aligned} l(\delta,\lambda_1,\lambda_2,\theta_1,\theta_2) &= n_1 ln\lambda_1 - \lambda_1 T + \sum_{i=1}^{n_1} lnf_1(x_i;\theta_1) \\ &+ n_1\lambda_2 - \lambda_2 T + \sum_{i=1}^{n_2} lnf_2(x_i;\theta_2) \\ &+ \sum_{i=1}^{n_1^{\perp}} ln\left(1 - \frac{\partial \bar{C}_\delta(u_1,\lambda_2)}{\partial u_1}\bigg|_{u_1=\lambda_1\bar{F}_1(x_i^{\perp};\theta_1)}\right) \\ &+ \sum_{i=1}^{n_2^{\perp}} ln\left(1 - \frac{\partial \bar{C}_\delta(\lambda_1,u_2)}{\partial u_2}\bigg|_{u_2=\lambda_2\bar{F}_2(y_i^{\perp};\theta_2)}\right) \\ &+ \bar{C}_\delta(\lambda_1,\lambda_2)T \\ &+ \sum_{i=1}^{n^{\parallel}} ln\frac{\partial^2 \bar{C}_\delta(u_1,u_2)}{\partial u_1 \partial u_2}\Bigg)_{u_1=\lambda_1\bar{F}_1(x_i^{\parallel};\theta_1),u_2=\lambda_2\bar{F}_2(y_i^{\parallel};\theta_2)}, \end{aligned} \tag{2.71}$$

and assuming the existence of $\frac{\partial}{\partial u_1 \partial u_2}\bar{C}(u_1,u_2)$ for all $(u_1,u_2) \in (0,\lambda_1) \times (0,\lambda_2)$.

Sections 4.1 - 4.3 of Velsen (2012), presents the third approach in a second version (which is proven that the likelihood converges to the likelihood provided in Eq.(2.71)) by deriving the forms of some discrete models (see the likelihood function in Equation (66) of Velsen (2012)). The likelihood mentioned in the third approach depends on the the following distributions:

$$F_1^{\parallel}(x,y) = \frac{\lambda_1 f_1(x)}{\lambda^{\parallel}}\left[\left(\frac{-\partial\bar{C}(u,\lambda_2\bar{F}_2(y))}{\partial u}\right)\right] + \left(\frac{\partial\bar{C}(u,\lambda_2)}{\partial u}\right)\bigg|_{u=\lambda_1\bar{F}_1(x)}\Bigg] \tag{2.72}$$

Similarly, $F_2^{\parallel}(x,y)$ takes the form of:

$$F_2^{\parallel}(x,y) = \frac{\lambda_2 f_2(y)}{\lambda^{\parallel}}\left[\left(\frac{-\partial\bar{C}(\lambda_1\bar{F}_1(x),v)}{\partial v}\right)\right] + \left(\frac{\partial\bar{C}(\lambda_1,v)}{\partial v}\right)\bigg|_{v=\lambda_2\bar{F}_2(y)}\Bigg] \tag{2.73}$$

And the density $f^{\|}$ is given by:

$$f^{\|}(x,y) = \frac{\partial^2 F_1(x,y)}{\partial x \partial y} = \frac{\lambda_1 \lambda_2 f_1(x) f_2(y)}{\lambda^{\|}} \frac{\partial^2 \bar{C}(u,v)}{\partial u \partial v}\Big|_{u=\lambda_1 \bar{F}_1(x), v=\lambda_2 \bar{F}_2(y)} \tag{2.74}$$

2.4.8 Maximizing the Likelihood Function by the Method of Inference Functions for Margins (IFM)

The parameters of the process $S = \{S_1, S_2\}$, can be estimated by maximizing the likelihood function L with respect to all its entries simultaneously. Joe and Xu, (1996) suggested the use of the IFM approach since maximizing the full likelihood can become numerically intensive for large datasets.

The IFM method to maximize the likelihood function is comprised of two steps:

1. Maximize the likelihood function (that has no dependence structure):

$$H_{j,s_j} = (\lambda_j)^{N_j(T)} e^{\lambda_j T} \prod_{l=1}^{N_j(T)} f_j((s_j)_l) \tag{2.75}$$

   where s_j is the vector of jumps of S_j within $[0,T]$ to estimate the parameters ($\lambda_1, \lambda_2, \theta_1$ and θ_2) of S_j with $j = 1, 2$,

2. Keeping the estimates ($\tilde{\lambda}_1, \tilde{\lambda}_2, \tilde{\theta}_1$ and $\tilde{\theta}_2$) constant, maximize the full likelihood function (with the dependence structure), $l(\delta, \tilde{\lambda}_1, \tilde{\lambda}_2, \tilde{\theta}_1)$, to estimate the copula parameter δ.

2.4.9 Algorithm for Implementation

In this section, our objective is to sample many times from $S = \{S_1, S_2\}$ on a period $[0, 1]$ and estimate its parameters. Assuming that the marginal jump size distribution function F_j with $j = 1, 2$ is given by $F_j = 1 - e^{-\theta_j x}$.

The following steps summarize the implemented pseudo-algorithm (van Velsen, 2012)

:

- Step 1: Draw $N_1^{\perp}$ and $N_2^{\perp}$ from a Poisson distribution with frequency $\lambda_1^{\perp}$ and $\lambda_2^{\perp}$ respectively,

- Step 2: Draw $N^{\parallel}$ from a Poisson distribution with frequency $\lambda^{\parallel}$,

- Step 3: Draw $N_1^{\perp}$ times from a uniform $[0,1]$ distribution. The resulting draws are the jump times of $S_1^{\perp}$. The $N_2^{\perp}$ jump times of $S_2^{\perp}$ are determined similarly,

- Step 4: Draw $N_1^{\perp}$ times from a uniform $[0,1]$ distribution and apply the inverse of $F_1^{\perp}$ to each draw. The resulting numbers are the jump sizes of $S_1^{\perp}$. The jump sizes of $S_2^{\perp}$ are determined similary,

- Step 5: Draw $N^{\parallel}$ times from a uniform $[0,1]$ distribution and apply the inverse of the marginal distribution defined as $F^{\parallel}(x) := F_1^{\parallel}(x) = F^{\parallel}(x,\infty)$ to each draw. The resulting x_i with $i = 1,...,N^{\parallel}$ are the jump sizes of $S_1^{\parallel}$,

- Step 6: Draw $N^{\parallel}$ times from a uniform $[0,1]$ distribution. The resulting draws are denoted by u_i with $i = 1,...,N^{\parallel}$. Apply the inverse of the distribution function to $H_{x_i}(y) = \frac{F_1^{\parallel}(x,y)}{f_1^{\parallel}(x)}$ for all $i = 1,...,N^{\parallel}$ and where $f_1^{\parallel}(x) = \frac{dF_1^{\parallel}}{dx}$. The resulting numbers y_i with with $i = 1,...,N^{\parallel}$ are the jump sizes of $S_2^{\parallel}$. In details, $H_{x_i}(y)$ can be obtained by:

$$\begin{aligned}
\mathbb{P}(\Delta S_2^{\parallel} \le y | \Delta S_1^{\parallel} = x) &= \lim_{\Delta x \downarrow 0} \mathbb{P}(\Delta S_2^{\parallel} \le y | x \le \Delta S_1^{\parallel} \le x + \Delta x) \\
&= \lim_{\Delta x \downarrow 0} \frac{\mathbb{P}(x \le \Delta S_1^{\parallel} \le x + \Delta x, \Delta S_2^{\parallel} \le y)}{\mathbb{P}(x \le \Delta S_1^{\parallel} \le x + \Delta x)} \\
&= \lim_{\Delta x \downarrow 0} \frac{F^{\parallel}(x+\Delta x, y) - F^{\parallel}(x,y)}{F_1^{\parallel}(x+\Delta x) - F_1^{\parallel}(x)} \\
&= \frac{F_1^{\parallel}(x,y)}{f_1^{\parallel}(x)} \\
&= H_x(y). \qquad (2.76)
\end{aligned}$$

Remark 13

Esmaeili et al. (2010) originally proposed the above algorithm but in the step where

$N_1^{\perp}$ and $N_2^{\perp}$ are drawn, they sampled N_1, N_2 and $N^{\parallel}$ and calculated $N_1^{\perp}$ and $N_2^{\perp}$ as $N_1^{\perp} = N_1 - N^{\parallel}$ and $N_2^{\perp} = N_2 - N^{\parallel}$. This sometimes ends up with $N_1^{\perp} < 0$ and, or $N_2^{\perp} < 0$.

Before we estimate the parameters, consider breaking down the observation period of [0,1] into M intervals of equal length and with reference to Eq.(51-2) of Velsen (2012), determine Z_{i1} and Z_{i2} for $i = 1, ..., M$. This results in an $M \times 2$ matrix z of maximum jump sizes. For all $i = 1, ..., M$. Find the samples $\tilde{N_{i1}}$ and $\tilde{N_{i2}}$ to get an $M \times 2$ matrix $\tilde{n}$ for the number of jumps.

Finally, by the IFM approach, find the vectors s_1 and s_2 of all the jump sizes of S_1 and S_2 on $[0,1]$ respectively. Repeat the entire algorithm many times and based on each $z, \tilde{n}, s_1$ and s_2, the parameters of $S = \{S_1, S_2\}$ on a period $[0,1]$ are estimated.

2.4.10 Summary and Discussion

The concepts, Sklar's theorem, copula density and survival copula from chapter 1 together with the switch from distributional function to tail integral function were a core part of this work. This paper showed us that, the dependence structure can be separated into two parts (dependence among severities and dependence among frequency of claims). Adding to the estimation methods we have reviewed from previous sections, the IFM method also provides an aid when the maximizing of a full likelihood becomes extremely difficult to achieve.

The algorithm explained above were implement by all the three (3) authors under different applications but interrelated circumstances.

2.5 Multivariate Counting Processes: Copula and Beyond

2.5.1 Introduction

In the previous section (2.4), we mentioned that in modeling insurance claim processes, one may consider more than one type of claim and may want to find out how dependencies between the claim arrival times for different claim types can be modeled. Examples are that an event may give rise to two or more claims or a claim having a connection with many other claims in different spatial locations. Claim processes are examples of stochastic processes and the above situations may lead us to analysis of multivariate stochastic processes. We modeled dependence using the Lévy copula.

Certain instances in modeling multivariate counting process calls for a researcher or a practitioner to go beyond the usual techniques of dependence modeling under copula theory. One of such instance is when we intend to model dependence in multivariate counting processes across time and across components simultaneously. This puts a limitation to the use of the modeling techniques we have reviewed in the previous section up until now.

The rest of this section is a simple overview of the original work done by Bäuerle and Grübel (2005), as they discussed a new technique for the construction of dependence models of a multivariate counting processes with Poisson marginals. Our goal will be to analyze a multivariate counting processes $X = (X_1, ..., X_d)$ with the property that, the one-dimensional marginal processes X_i are Poisson processes with constant rates λ_i respectively, $i = 1, ..., d$, where X_i counts the claims of type i. See definition(1.7.2).

2.5.2 Model Specification

Variables Definition

Definition 2.5.1 (Counting Processes - Under Current Context)
A counting process X_i as a family of non-negative, integer-valued random variables indexed by subsets A of the real line $\mathbb{R}$, where $X_i(A)$ denotes the number of events of type i with 'time stamp' in A. If observations start at time 0, then it is customary to work with $t \mapsto X_i([0,t])$ as the random measure $A \mapsto X_i(A)$ can then conveniently be described by its distribution function.

Definition 2.5.2 (Borel Set)
The Borel or topological sigma-algebra (or σ-algebra) $\mathscr{B}(S,\tau)$ of a topological space (S,τ) is the σ-algebra generated by τ. The elements of $\mathscr{B}(S,\tau)$ are called the Borel (measurable) sets of (S,τ).

Remark 14
Stationary processes is a stochastic process whose unconditional joint probability distribution does not change when shifted in time. We can express time shift stationarity as a requirement that the distribution of the random vector $(X_{1(A_1+t)},...,X_{D(A_d+t)}$ does not depend on $t \in \mathbb{R}$. Here $A_1,...,A_d$ are Borel subsets of the real line and we have written A_{i+t} for the shifted set $\{x+t : x \in A_i\}$.

Marginal Distributions

As stated earlier in chapter 1, in modeling multivariate counting process, we first have to study the probabilistic behaviour of each of one of the component of $\mathbf{Y}$ and second, is to investigate the relationship between them.

Let us begin by considering the following variable definitions:

1. $\xi = (\xi_1, \xi_2, ..., \xi_d)$: a d-dimensional random vector,

2. $F_1,...,F_d$: a one-dimensional distribution functions associated with the components $\xi_1,...,\xi_d$ of the random vector.

Modeling Dependence

From theorem (1), the joint distribution function F of a d-dimensional random vector $\xi = (\xi_1, \xi_2, ..., \xi_d)$ can be written in the form:

$$F(X_1,...,X_d) = C(F_1(X_1),...,F_d(X_d)).$$

Now, we intend to model dependence in multivariate counting processes $X = (X_1,...,X_d)$ across time and across components simultaneously.

Suppose now that we start observations at time 0 so that our multivariate counting process can be indexed by $\mathbb{R}+$. Then for each $t > 0$, the transformation to uniform marginals can be applied to the individual random vector $(X_1(t),...,X_d(t))$. This would result in a family $(C_t)_{t\geq 0}$ of copulas.

In comparison to the above static situation this has two disadvantages. First, a whole family of copulas is inconvenient and will not come in handy. Secondly, (as this was explained in Remark (d) at the end of Section 5 of the original paper) , this family would not satisfy our demands, whereas in the random vector case the distribution is completely specified by the copula and the marginal distributions, the family $(C_t \geq 0)$ and the distribution of the component processes $X_i, 1 \leq i \leq d$, together do not determine the finite-dimensional distributions of the multivariate counting process X.

Two approaches to overcome the above difficulties are as follows:

1. Lindskog and McNeil (2003) suggested that, X can be assumed to be a multivariate Lévy process (jumps) which implies that the component processes can be regarded as thinnings of a basic (univariate) Poisson process. Thus dependence

of the marginal processes can be modeled only through the synchronicity of the jump,

2. Pfeifer and Nešlehová (2004), used static d-dimensional copula to construct $X((0,T])$ for fixed and finite time interval $T > 0$. Model the dependence of the component processes by basing entirely on the dependence modeling of the total number of claims in the finite period of interest.

Now let us consider these two (2) approaches in detail.

Approach 1: Consider the following variable definitions:

1. Let $\mathbb{D} := \{1, ..., d\}$ and for all $D \subset \mathbb{D}$,

$$e(D) = \begin{pmatrix} e_1(D) \\ \cdots \\ e_d(D) \end{pmatrix}$$

where

$$e_i(D) = \begin{cases} 1, & if \ i \ \in D \\ \\ 0, & otherwise \end{cases}$$

2. Jumps $X(t) - X(t^-)$ are of the form $e(D)$ and that a jump $e(D)$ at time t means that the component processes X_i with $i \in D$ (and only these) have a jump of height 1 at time t.

Since X is assumed to be a multivariate Lévy process (jumps) in this approach-1, its representation can be written as:

$$X_i(t) = \sum_{i \in D} N_d(t) \ \ i = 1, ..., d, \tag{2.77}$$

where N_D are independent Poisson processes with rate λ_D, D as above. The rates λ_i of the marginal processes are related to the rates λ_D by $\lambda_i = \sum_{i \in D} \lambda_D \; i = 1, ..., d$. A similar result has been used in Lindskog and McNeil (2003) .

Remark 15

Lévy type counting processes have a very specific dependence structure: In the random measure notation, $X_i(A_i)$ and $X_j(A_j)$ will always be independent if $A_i \cap A_j = \emptyset$ because of the independence of the increments of X. Hence, for such processes, dependence of the marginal processes is only possible via the synchronicity of the jumps.

The condition that the marginal processes be constant rate Poisson processes implies that the paths of X are of pure jump type. If we assume in addition to the above that X is a Lévy process, then the Markov property and the homogeneity in time imply that the copula family described by its 'infinitesimal generator', the Lévy copula, which together with the other characteristics of a Lévy process is enough to specify the distribution of the whole multivariate process, Cont and Tankov (2004).

The processes N_D in turn can be obtained from univariate Poisson process N with rate $\lambda = \sum_{D \subset \mathbb{D}} \lambda_D, i = 1, ..., d$, by independent marking of the points of N with probability $p_D = \lambda_D / \lambda$ and then collecting the marked points into N_D.

Approach 2: In this second approach, we begin by making two necessary considerations:

1. Instead of an infinite time space $\mathbb{R}+$, we consider a finite time interval $[0, T]$ and condition on the final random vector $X(T) = (X_1(T), ..., X_d(T))$ of the processes.
2. Instead of the whole family of copulas $(C_t)_{t \geq 0}$, a static d-dimensional copula can be used together with the condition that the components $X_i(T)$ have a Poisson distribution with parameter $\lambda_i T, i = 1, ..., d$, to construct the law of the vector $X(T)$.

To obtain $(X(t))_{0\leq t\leq T}$ from the final random vector $X(T)$ one can use the familiar fact that conditionally on their total number in $[0,T]$ being equal to n, the n points of a Poisson process with constant rate are independent and uniformly disributed on $[0,T]$. Hence, in the d-dimentional case, we obtain $(X(t))_{0\leq t\leq T}$ from X(T) via

$$X_i(t) = \sum_{j=1}^{X_i(T)} 1_{[0,t]}(\xi_{ij}), \quad 1 \leq i \leq d, \quad 0 \leq t \leq T \tag{2.78}$$

where $\xi_{ij}, 1 \leq i \leq d, 1 \leq j \leq X_i(T)$, are independent and uniformly distributed on $[0,T]$. Here 1_A denote the indicator function associated with the set A.

Remark 16

In this approach we makes use of the fact that the superposition of independence Poisson processes is again a Poisson process.

Finally, dependence modeling of the component processes is thus based entirely on the dependence modeling of the total number of claims in the period of interest.

Limitations of the two approaches:

1. Lévy type counting processes have no dependence 'across time', and the dependence 'across components' is of a very special nature. For this class of processes dependence modeling is reduced to the choice of thinning probabilities.

2. Modeling by approach 2 is more flexible as the models can make use of the whole range of (static) copulas, however they require the choice of a compact base interval $[0,T]$. If such an interval does not suggest itself from the application of interest, then both remedies known so far, letting either T tend to ∞ or patching together several such intervals, have their disadvantages, the second getting into conflict with our assumption of time shift stationarity.

2.5.3 Dependence in Thinning and Shift (TaS) Models

Resnick (1987) and Dale and Vere-Jones (2007), extended the Lévy models by incorporating random shifts of the individual points to build a new family of models called Models with Thinning and Shifts (TaS) models. The objective for introducing this model is to overcome some of the limitations mentioned in the above approaches. This section goes a bit beyond using copulas as measure of dependence. However, only the introductory part may be seen in this thesis.

Constructing a TaS Model

To build a TaS model, we will need:

1. a background Poisson process N on $\mathbb{R}$ with constant intensity λ,
2. a thinning mechanism described by a probability distribution $(p_D)_{D \subset \mathbb{D}}$ on $\mathcal{P}(\mathbb{D})$, the power set (set of all subsets) of $\mathbb{D}$, as we mentioned in approach 1,
3. and a sequence $(Y_l)_{l \in \mathbb{Z}}$ (also called shifts) of independent d-dimensional random vectors, all with distribution Q.

The example below (original version may be seen in Bäuerle and Grübel (2005)) introduces us to the the background process:

Example 2.5.1

Suppose the points of a process represents time points of natural catastrophes like earthquakes, floods, hurricanes etc. An immediate damage to houses and cars which produces claims in the non-life insurance branch is then mostly likely followed by claims in health insurance due to epidemics or general bad health conditions after the catastrophe. Thus X_1 and X_2 would count claims in non-life and health insurance respectively. In the context of this work, an extension of our models, where a single event

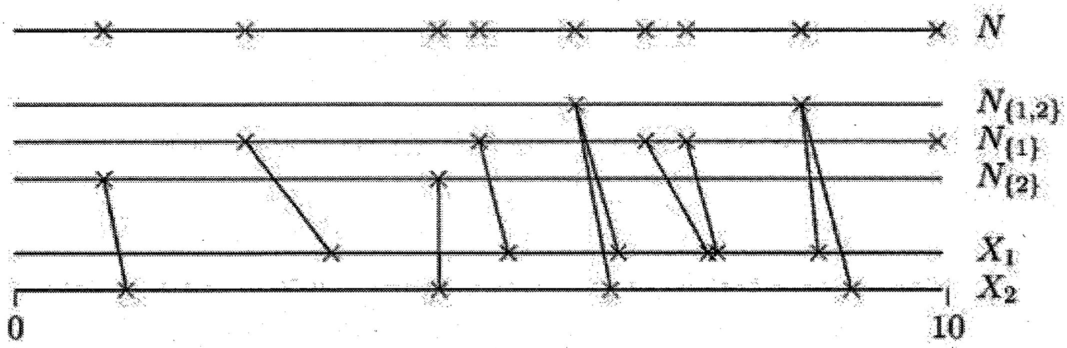


Figure 2.1 This figure is an example combining thinning and shifts

(background process) may generate more than one event in the individual component processes, may be of interest.

Next, supposing we order the event time points $(T_l)_{l\in\mathbb{Z}}$ of the background process in such a way that $-\infty < ... < T_{-2} < T_{-1} < T_0 < 0 \leq T_1 < T_2 < ... < \infty$. The points for component X_i of X are now constructed by shifting those T_l that are not deleted for this component (which happens with probability $\sum_{i\subset D} p_D$) by the amount $Y_{li}, Y_l = (Y_{l1}, ..., Y_{ld})$.

See the figure 2.1 for a better illustration. Source is Bäuerle and Grübel (2005).

Once the shifts are carried out, a Poisson process on $\mathbb{R}^d$ result whose points consist of the potential claim arrival times for each of the d (d = 2 in the above example) different claim types, with one point for each of the triggering events (the time stamps of the triggering events are lost in the process). Each point of this d-dimensional process is additionally marked by the set $D \subset \mathbb{D}$ of claim types that are going to survive the thinning step. After the second (thinning) step, we have independent Poisson processes on $\mathbb{R}^D$, one for each D with $p_D > 0$. The number $X_i(A)$ of claims of type i with time stamp in A is now obtained by taking the sum of the number of points in these individual processes that have their i^{th} coordinate in A.

It should be clear that this generalizes the Lévy model which we would obtain with Q concentrated on the zero vector. Also, this model is easy to simulate. We abbreviate the above by calling X a TaS model ('thin and shift') with parameter λ, p and Q, where λ is the intensity of the Poisson base process, $p = (p_D)_{D \subset \mathbb{D}}$ the thinning mechanism and Q the shift distribution. That these models satisfy our basic requirement of Poisson marginals is a standard fact from the general theory of point processes, see e.g. p. 138 in Resnick (1987): Shifting the points of a constant rate Poisson process (on $\mathbb{R}$) by i.i.d. amounts results in a constant rate Poisson process.

In their paper, (Bäuerle & Grübel, 2005) mentioned that a possible application of the TaS model is as follows:

Example 2.5.2 (Source: Bäuerle and Grübel (2005))
Continuing from the above example, consider Q, our shift distribution to be a product of uniform distribution (0,1) and the exponential distribution with mean 1, *and simulate* Y_1 *and* Y_2. *The table (2.1) shows a segment of a simulated path for* $d = 2$, *with* $\lambda_{\{1\}} = \lambda_{\{2\}} = \lambda_{\{1,2\}} = 1$ *as seen in Eq.(2.77). In the table, an event occurring at time 0.960 gives rise to a claim of type 2 only; the corresponding delay is 0.230, so that the claim is registered a time 1.190. The event occurring at time 6.022 is the first to trigger claims of both types.*
This example also helps to explain a technical point: Whereas, in order to emphasize the connection with the Lévy case, it suggests that first the thinning is done and then the shifts are applied, Table (2.1) lists also those shifts that because of the thinning later become irrelevant, such as the value $Y_1 = 0.316$ *in the first line. Because of our basic independence assumptions the order of the two operations irrelevant; in the proofs it will be more convenient to delete components in the second step.*

Table 2.1 Illustration from the example: Source Bäuerle and Grübel (2005)

N	Y_1	Y_2	D	X_1	X_2
0.960	0.316	0.230	{2}	-	1.190
2.481	0.916	0.206	{1}	3.397	-
4.543	0.660	0.008	{2}	-	4.551
4.987	0.308	0.463	{1}	5.295	-
6.022	0.455	0.369	{1,2}	6.477	6.391
6.773	0.663	5.642	{1}	7.436	-
7.211	0.328	0.429	{1}	7.539	-
8.440	0.181	0.533	{1,2}	8.621	8.973

Basic Structural TaS Models

Over here, our objective is to give a general description of the joint behaviour of $X_1(A_1),...,X_d(A_d)$ by expressing the joint distribution in terms of these parameters. We need these definitions:

Definition 2.5.3

For subsets $A_1,...,A_d$ of $\mathbb{R}$ and both $D,D' \subset \mathbb{D}$ with $D \subset D'$ let:

1. $M(D,D';A_1,...,A_d) := B_1 x ... x B_d, B_i := \begin{cases} A_i, for \ \ i \in D, \\ A_i^c, for \ \ i \in D' \backslash D, \\ \mathbb{R}, otherwise, \end{cases}$

2. *e=e(D)=(1,...,1): the d-dimensional vector.*

Definition 2.5.4

We define a measure $\nu(Q)$ on $\mathbb{R}^d$ by:

$$\nu(Q)(A) := \int Q(A - te)dt$$

for all d-dimensional Borel sets A; here $A - x$ denotes the set $\{a - x : a \in A\}$.

This measure plays an important role throughout the sequel. It is obvious from the result cited above on i.i.d. shiftings of constant rate Poisson processes that the marginal measures $\nu(Q)^{\pi_i}, 1 \leq i \leq d$, where π_i:$\mathbb{R}^d \to \mathbb{R}$ denotes the projection on the i^{th} coordinate, are all equal to the Lebesgue measure. The following direct computation may be instructive:

$$\begin{aligned}
\nu(Q)^{\pi_i}(B) &= \nu(Q)(\mathbb{R}^{i-1} \times B \times \mathbb{R}^{d-i-1}) \\
&= \textstyle\int_{\mathbb{R}} \int_{\mathbb{R}^d} 1_{\mathbb{R}^{i-1} \times B \times \mathbb{R}^{d-i-1}}(X - te) Q(dx) dt \\
&= \textstyle\int_{\mathbb{R}^d} \int_{\mathbb{R}} 1_{B - x_i}(t) dt Q(dx) \\
&= l(B)
\end{aligned}$$

for all one-dimensional Borel sets B.

Theorem 13 (Distributional Representation of a TaS Model)
Let X be a d-dimensional counting process of TaS type with base rate λ, thinning mechanism $p = (p_D)_{D \subset \mathbb{D}}$ and shift distribution Q. Then, for any Borel subset $A_1, ..., A_d$ of the real line, we have the following distributional representation:

$$\begin{pmatrix} X_1(A_1) \\ \ldots \\ X_d(A_d) \end{pmatrix} = \begin{pmatrix} \sum_{1 \in D} \xi(D; A_1, ..., A_d) \\ \ldots \\ \sum_{d \in D} \xi(D; A_1, ..., A_d) \end{pmatrix},$$

where the random variables $\xi(D; A_1, ..., A_d), \emptyset \neq D \subset \mathbb{D}$, are independent and Poisson distributed with

$$E\left(\xi(D; A_1, ..., A_d)\right) = \lambda \textstyle\sum_{D \subset D'} p_{D'} \nu(Q)(M(D, D'; A_1, ..., A_d)).$$

Theorem (13) has some useful qualitative and quantitative consequences, it displays, for example, the simple 'multiplicative' way that the parameters enter the dependency structure of the model. If Q is concentrated on the zero vector, in which case we have thinning only, then, for $A = A_1 \times \ldots \times A_d$,

$$\nu(Q)(A) = \int \delta_0(A - te)dt = l(\{t \in \mathbb{R} : te \in A\}) = l(A_1 \cap \ldots \cap A_d),$$

which identifies $\nu(\delta_0)$ as 'Lebesgue measure on the diagonal', i.e. the image of l under the transformation $X \in (R) \mapsto (X, \ldots,) \in \mathbb{R}^d$.

In their paper, (Bäuerle & Grübel, 2005) with extra theorems investigated into greater details, the dependence structures using TaS models. However, the scope of my work is limited to modelling dependence using copulas.

2.5.4 Conclusion

This findings in reviewing this paper also got our research in this area informed. We found that, when one intends to model dependence of a multivariate counting processes across time and across components simultaneously, further models which goes beyond what we have reviewed from sections (2.1-2.4) of this thesis. Further studies of these TaS models in other literature shows that it comes with useful dependence properties.

CHAPTER III

DATA, METHODOLOGY AND ESTIMATION

3.1 Overview

This chapter marks the beginning of the second part of the thesis. In this chapter, our main goal is to present the approaches used for arriving at answers to the research question in the second part of this thesis. The sub-sequent sections will present the data used, methods employed in this work, parameter estimation techniques and inference drawn during these procedures. We considered modeling only one aspect of dependence (which is common severities) as mentioned in LC2 of section 2.4.6 of this thesis. In other words, the chapter 3 is developed from a section under the context of the reviews made in chapter 2. The numerical part were carried using the R-software and Microsoft excel.

3.2 Data

From Embrechts et al. (1997) and McNeil (1997) we obtained the download of fire insurance claims dataset (source: http://www.ma.hw.ac.uk/ mcneil) which were collected at Copenhagen Reinsurance and comprise 2167 fire losses over the period 1980. They have been adjusted for inflation to reflect 1985 values and are expressed in millions of Danish Krone. There are five columns in this dataset namely:

1. Date: The day of claim occurence.

2. Building: The loss amount (mDKK) of the building coverage.

3. Contents: The loss amount (mDKK) of the contents coverage.

4. Profits: The loss amount (mDKK) of the profit coverage.

5. Total: The total loss amount (mDKK).

Total claim is the sum of a building loss, a loss of contents and a loss of profits. This dataset was used because of its readiness in availability and most importantly, its widely used in many literature presents some significant amount of reliability to this work.

3.3 Methodology

We begin by preparing the 2167 fire claims dataset for its usage. In an event of fire, there can be a building claim-type and or a content claim-type and in this thesis, we selected the building claims and the contents claims as our bivariate random variables. Univariate modeling and dependence modeling of these claim-type have already being well discussed under the main literatures under section (2.4) of this thesis so the choice of the univariate models and copula was an informed choice. However for the purpose of tandem in results, we still carried a univariate modeling.

In preparing the dataset, we only considered the 1502 non-zero claims in both variables (building claims and contents claims). We presented some descriptive statistics of the dataset. For the univariate modeling, two members of the transformed gamma family of distributions (exponential distribution and the Weibull distribution) were considered for each of the two variables. Our selection confirmed the choices made by Esmaeili et al. (2010) and Avanzi et al.,(2011). However, one can also infer from theorem(6) copulas are invariant under preserving the order transformations, so the Lévy Clayton copula which were used to model the log-transformed univariate claims (see (2010) and

Avanzi et al.,(2011)) was used to model the dependence between the non-zero claims of each of the building claims and that of the content claims.

Next we considered some basic estimation method like the maximum likelihood estimation to find the univariate model parameters.

After the above procedures, we proceeded to tackle three (3) different approaches in finding the Lévy Clayton copula parameter. Consider X_{1i} as the random variable for building claims and X_{2i} as the random variable for contents claims. Below are the three approaches:

Approach 1 - Pure Lévy-Clayton copula empirical fit

1. Do not assume any univariate model for both building claims and content claims,
2. Calculate the cumulative distribution value for the pairs (U_i, V_i) corresponding to the pair (X_{1i}, X_{2i}) respectively by empirical methods,
3. Using the Lévy-Clayton copula, calculate the log-likelihood value from (U_i, V_i),
4. Maximize this value to estimate the copula parameter θ (Note that this is the only parameter to estimate in this approach).

Approach 2 - Pure Lévy-Clayton copula model fit (separate modeling)

1. Assume univariate models for each random variable, building claims and content claims. Estimate the parameters of each model separately from each other using the dataset before moving to the next step,
2. Generate the cumulative distribution value for the pairs (U_i, V_i) corresponding to the pair (X_{1i}, X_{2i}) respectively by univariate modeling methods,

3. Using the Lévy-Clayton copula, calculate the log-likelihood value from (U_i, V_i),

4. Maximize this value to estimate the copula parameter θ (Note that this is the only parameter to estimate in this approach).

Approach 3 - Pure Lévy-Clayton copula model fit (all-in-one modeling)

1. Assume univariate models for each random variable, building claims and content claims. But do not estimate the univariate model parameters,

2. With the help of Eq.(1.9) and the Lévy-Clayton copula, find the joint density of the two random variables,

3. Find the full likelihood function to the above,

4. Maximize this function to estimate both the model parameters and the copula parameter θ (Note that there will be five (5) parameters to be estimated in this approach).

3.4 Inference - Estimation and Asymptotic Property

3.4.1 Maximum Likelihood Estimation

From the copula density function given by Eq.(1.8), one may want to estimate both the model parameters and the copula parameter. In our case, the five parameters to be estimated are $\theta_{building}$, $\theta_{contents}$,$\alpha_{building}$, $\alpha_{contents}$ and δ.

Let $\eta = (\theta_{building}, \theta_{contents}, \alpha_{building}, \alpha_{contents}, \delta)$. The estimation of η can be achieved by the method of maximum likelihood estimation. The log-likelihood function to the above can be written as:

$$\mathcal{L}(\eta;x) = \sum_{i=1}^{n}\sum_{k=1}^{2} log f_i(x_{ik};\alpha_k,\theta_k) + \sum_{i=1}^{n} log c(F_1(x_{1i}),F_2(x_{2i});\delta) \tag{3.1}$$

We can say that, the maximum likelihood estimator $\hat{\eta}$ of the parameter η is the solution of

$$\frac{\partial \mathcal{L}(\eta;x)}{\partial \eta} = 0. \tag{3.2}$$

let η_0 be the true value of $\hat{\eta}$. Under standard regularity conditions, consistency and asymptotic normality properties of the estimator $\hat{\eta}$ have been established; see for instance Joe (1996).

The asymptotic properties to the estimator is given by:

$$\sqrt{n}(\hat{\eta} - \eta_0) \to N(0, I^{-1}). \tag{3.3}$$

The above represent convergence in distribution, where I is the Fisher Information matrix. An alternative method of estimation is the Inference Functions for Margins (IFM) and this was mentioned in the previous chapter.

3.4.2 The Profile Likelihood

Suppose that the unkown parameter ψ, is partitioned as $\psi = (\alpha, \theta)$, where θ is the parameter of interest (eg. scale parameter from Weibull distribution) and α is the nuisance parameter (eg. shape parameter from Weibull distribution). We will need to estimate both, but our interest lies only in the parameter θ. The estimation of these parameters is done in two stages.

Suppose that $\{X_i\}$ are iid random variables, with density $f(x;\alpha,\theta)$ where our goal is to find estimates for α and θ. The log-likelihood function is given by:

$$\mathcal{L}_n(\theta, \alpha) = \sum_{i=1}^{n} log f(X_i; \alpha, \theta) \tag{3.4}$$

To estimate θ and α, one can use $(\hat{\alpha}, \hat{\theta}) = \underset{\alpha,\theta}{argmax} \mathcal{L}_n(\theta, \alpha)$. The problem associated with this is the difficulty associated with a direct maximization principles. First, we suppose that, θ is known, then we rewrite the log-likelihood function as $\mathcal{L}_n(\theta,\alpha) = \mathcal{L}_\theta(\alpha)$ (ie. θ is fixed and α varies). To estimate α we maximize $\mathcal{L}_\theta(\alpha)$ with respect to α, thus:

$$\hat{\alpha_\theta} = \underset{\alpha}{argmax} \mathcal{L}_\theta(\alpha) \tag{3.5}$$

Secondly, we do not know θ in reality, we only pretended that we know it so for each estimated θ, we require a corresponding $\hat{\alpha_\theta}$. Now since each θ comes with a new curve $\mathcal{L}_\theta(\alpha)$ over α, we can choose the θ which is the maximum over all these curves. In other words, we evaluate:

$$\hat{\theta}_n = \underset{\theta}{argmax} \mathcal{L}_\theta(\hat{\alpha}_\theta) = \underset{\theta}{\mathcal{L}_n}(\theta, \hat{\alpha}) \tag{3.6}$$

CHAPTER IV

ANALYSIS, DISCUSSION AND CONCLUSION

In this chapter, we will outline the major findings to the second part of the thesis. We first present an analysis to the results obtained from chapter 3, discuss the findings and finally conclude.

4.1 Analysis of Results

From chapter one, we understood that the copula parameter characterizes the dependence structure between two random variables. surprisingly, each of the three (3) approaches (already mentioned in chapter 3) presents different values for the Lévy-Clayton copula parameter. A major concern in a dependence study, is the ability to provide inference from a copula parameter. For instance, is there any form of independence, or what will be the strength or degree of dependence between the two variables. From the findings below, we will like to compare the different estimates of parameters we are getting within (from one iteration runs to the other) each approach and across the three (3) approaches.

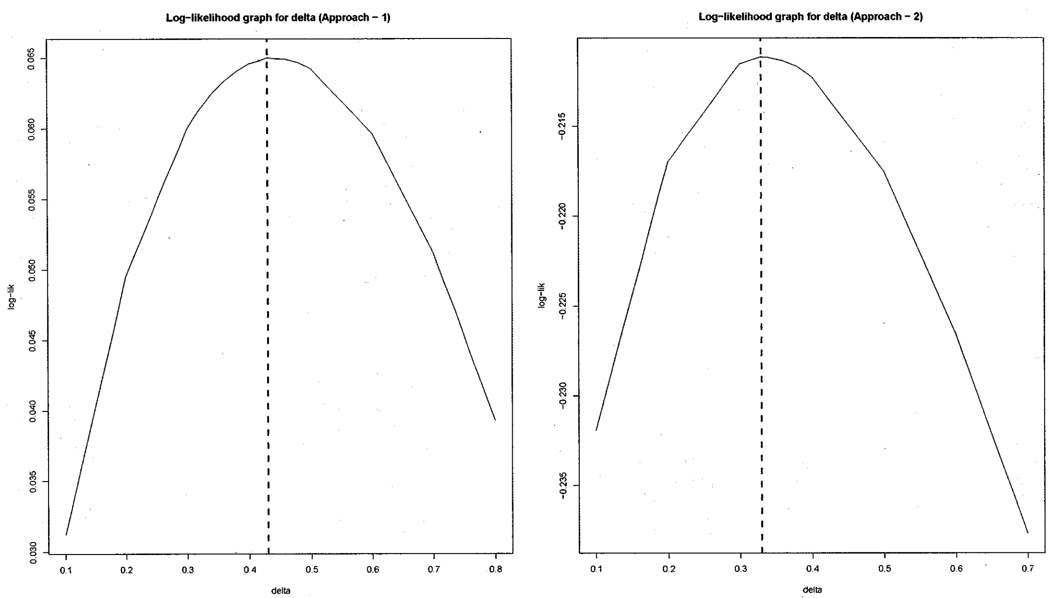


Figure 4.1 This figure shows comparison of estimated copula parameter in Approach-1(left) to that in Approach-2(right)

Table 4.1 Results from estimation: both copula and model paremters - Approach 3

	iteration runs-1	iteration runs-2	iteration runs-3	iteration runs-4	iteration runs-5
$\theta_{building}$	1.81458571	1.82502593	1.731435919	1.97127104	2.0148473405509
$\alpha_{building}$	0.95754286	0.9639855	0.965917959	1.0152914	1.021953279338
$\theta_{contents}$	1.05744286	1.08523444	1.079021551	1.07600531	1.06418441484726
$\alpha_{contents}$	0.6144857	0.62627135	0.62213932	0.64229784	0.649107150774117
δ	0.2572486	0.29452105	**0.312530019**	**0.22819507**	**0.192036752178155**
log-likelihood	-2.90013	-2.89004	**-2.8887**	**-2.87103**	**-2.871257893**

1. Comparing Run-(3) to Run-(5) with respect to their maximum log-likelihood values in the third approach, one can see that the two values are closer (ie. -2.8887, -2.87103 and -2.87125789 see table 4.1). However, the copulas parameter estimates are way too far from each other (ie. 0.312530019, 0.22819507 and 0.19203675217 respectively). Thus, very close maximum likelihood value but different parameter estimates.

   Just within one approach, approach 3, it is difficult to know which copula parameter one should use to infer the dependence structure of the random variables of interest.

2. Away from the third approach, the estimates for the copula parameter coming from the other two approaches are also different (see figure 4.1). One does not know whether starting the estimation with univariate models can produce the accurate copula parameter or not. So the question, as to which copula parameter to infer from still remains.

There is a qualitative discretion here and it seems that researcher have not address this.

4.2 Discussion

Inferring from the results and analysis presented above, as an attempt to measure dependence between variables in the area of insurance, there is a higher chance that, the applied copula methods or procedures may be theoretically appealing but, the final decisions may be subjective especially in terms of the copula parameter.

Since we were unable to provide a precise answer as to which of the copula parameters to conclude the studies with, similarly, we cannot be precise to conclude on whether anyone with large datasets should go ahead to model dependence by a parametric approach, or a non-parametric approach or strictly by data-dependent models.

My assertion on this question is that, every analyst will certainly come to this same point, but most of these analysts will loosely use the results for inference concerning the dependence structure of the random variable of interest. Also suppose one is accurately able to choose a copula from the numerous copula built, he or she will still have to provide an answer to this question. My personal view is that, there is currently no precise way to go about it.

Some researchers would also want to look into the optimization procedures at our exposure in recent times. Is the optimization tools we have at our exposure now, out of date? We understand that data are outcomes of random variables and repeats of the experiment would generate different data and hence different estimates (ie. randomness in the sampling process) induces randomness in the estimator. However in a case such as the results produced from the three (3) different approaches, the same dataset was used under varying methods but we still arrived in a situation of significant randomness in the copula parameter estimates. This may have to call for researchers to look into the optimization procedures we are exposed to once again, since optimization forms important part of dependence modelling with copulas.

4.3 Conclusion

In this thesis, we found out that copulas are extremely useful tools in modeling dependence in the area of counting processes with applications in insurance and finance. In insurance, joint behaviours associated with count of claims are of high interest to insurance companies and for that matter, the ability for researchers to come up with some techniques in such a regard is commendable.

The main contribution of this thesis is the central findings we presented during the in-depth reviews of how copulas were used in modeling for each of the counting process context discovered in the published papers of chapter two. For instance, in chapter two of this thesis, among several other random variables associated with count data in insurance, we researched on how dependence modeling is carried for bivariate risk or risk factors such current Bonus-Malus class and past count of claims, count of claims and size of claims, count of claims of two different counting processes that occurred from the same event etc. We observed how the Clayton copula, the Farlie-Gumbel-Morgenstern copula, the Gaussian copula and the Lévy-Clayton copula were used in modeling these joint behaviours. Copula parameters and model parameter estimation methods were also explored, for some of which we used the method of maximum likelihood estimation, the method of maximization by parts and also method of inference functions for margins. These methods of estimation were chosen based on the context or nature of the joint distribution functions. Computational, statistics and probabilistic techniques from the chapter one and Sklar's theorem played an important role in this work.

Despite the extreme usefulness of copulas, there is the need for improvements in the modeling techniques. For instance, the research question in the second part of the thesis remains, that is if we have a large set of dataset available and we want to estimate the copula parameters, we cannot say whether to use solely the dataset without assuming any model to arrive at a copula parameters is the best or whether to assume marginal models for data and estimate the copula parameter (separately or jointly). In our work, we selected the dependence defined in one of the reviews made in chapter two and focused on the dependence among severities of insurance for two different counting processes. Our findings were that, each approach results to significantly different copula parameter. A deeper search into this will be considered as our further studies.

Again, questions of precision in copula selection to model joint behaviours of different random variables remains unanswered. In chapter two of my thesis, the F-G-M copula was chosen because it is mathematically tractable under such a context. However no other scientific approach in selecting the appropriate copula has been agreed on till date. Similarly, areas under modeling dependence of counting process in insurance and finance such as the multivariate joint behaviour of aggregate cost of insurance claims variable remains open for research. Then again, one may require a copula, which can jointly models or unifies all other copulas mentioned in specific contexts of this thesis. This area of research also remains open.

APPENDIX A

RESULTS

Table A.1 Summary statistics of fire claim sizes in each class

Statistics	Building	Contents
Mean	1.82	1.32
Standard Deviation	4.36	4.76
Skewness	24.36	16.74
Kurtosis	755.65	385.90
Minimum	0	0
Median	1.27	0.37
Maximum	152.41	132.01

Table A.2 Results from Weibull distribution (by Profiled Likelihood Estimation)

	$\hat{\alpha}$	$\hat{\theta}$
Building Claims data	1.06473	1.9325311
Contents Claims data	0.691436	1.1211411

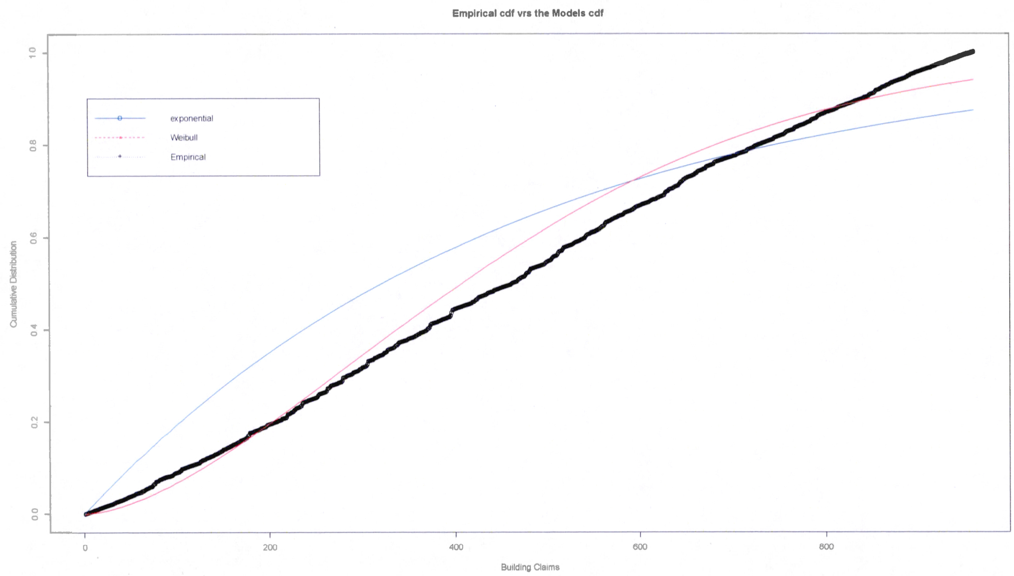


Figure A.1 This figure shows a cumulative distribution fit of Weibull model (red) and Exponential model (blue)

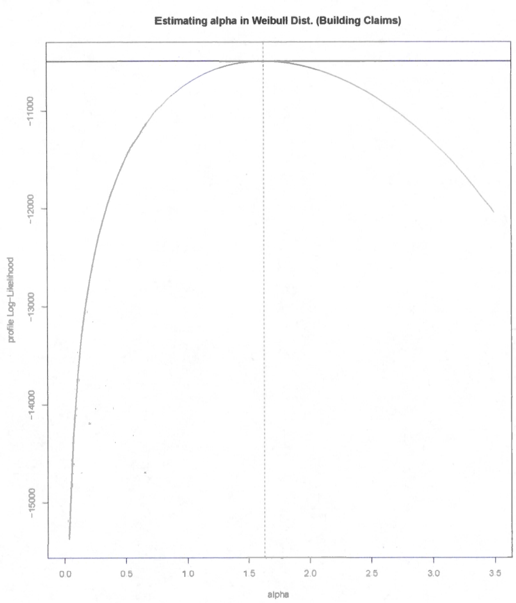


Figure A.2 This figure shows profile likelihood estimation for Weibull model parameters for Building random variable

Table A.3 Number of unique and common fire claims

		Contents		
	Item	Claim	No Claim	Total
Building	Claim	1502	488	1990
	No Claim	177	0	177
	Total	1679	488	2167

Table A.4 Results from Weibull distribution (by Mean Square Error Method)

	$\hat{\alpha}$	$\hat{\theta}$	log-likelihood Value
Building Claims data	1.875782	1.57938231	3007.402
Contents Claims data	0.99497608	0.81455864	3155.337

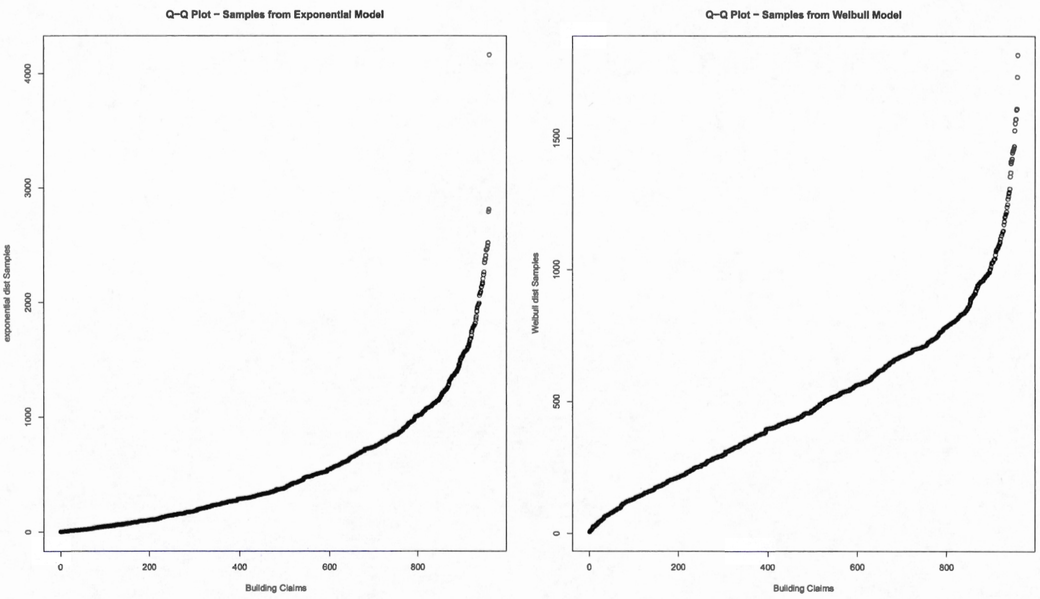


Figure A.3 This figure shows the Q-Q plots of the Exponential model and the Weibull model

Table A.5 Results from estimating the copula paremters by Approach 1

	$\hat{\delta}$	log-lik
iteration runs-1	0.1	0.0312195036805278
iteration runs-2	0.2	0.0494841338539416
iteration runs-3	0.3	0.0599466836801341
iteration runs-4	0.32	0.0613223019738393
iteration runs-5	0.34	0.0624350381143784
iteration runs-6	0.36	0.0633381647001194
iteration runs-7	0.38	0.0640175085023297
iteration runs-8	0.4	0.0645728088660071
iteration runs-9	0.42	0.0648310066346163
iteration runs-10	**0.43**	**0.0650065936923793**
iteration runs-11	0.44	0.0649530158118493
iteration runs-12	0.46	0.0649067729452187
iteration runs-13	0.48	0.0646779922063144
iteration runs-14	0.5	0.0642340856182826
iteration runs-15	0.6	0.0596522585989158
iteration runs-16	0.7	0.0511834113777575
iteration runs-17	0.8	0.0392615202052774

Table A.6 Results from estimation: copula paremters by Approach 2

	$\hat{\delta}$	log-lik
iteration runs-1	0.1	-0.231919770352844
iteration runs-2	0.2	-0.216978599766093
iteration runs-3	0.3	-0.211538373422296
iteration runs-4	0.32	-0.21127604185123
iteration runs-5	**0.33**	**-0.211150417885273**
iteration runs-6	0.34	-0.211177808877447
iteration runs-7	0.36	-0.21136167567968
iteration runs-8	0.38	-0.211665021449891
iteration runs-9	0.4	-0.212246730358873
iteration runs-10	0.5	-0.217500722607756
iteration runs-11	0.6	-0.226589587740836
iteration runs-12	0.7	-0.237670313655276

APPENDIX B

SOME FEATURES OF THE COPULAS USED IN THIS MATERIAL

B.1 Farlie-Gumbel-Morgenstern (F-G-M) Copula

B.1.1 Formula for Distribution Function

$$C(u,v) = uv[1+\alpha(1-u)(1-v], \ \ -1 \leq \alpha \leq 1. \tag{B.1}$$

B.1.2 Formula for Density Function

The density function is symmetric about the point $(\frac{1}{2}, \frac{1}{2})$. Meaning the copula density is the same at $(1-u, 1-v)$ and (u,v). So the survival copula over that point is the same as the original copula.

$$c(u,v) = 1+\alpha(1-2u)(1-2v). \tag{B.2}$$

B.1.3 Correlation Coefficient

F-G-M Copula has correlation coefficient (ρ) to be $\frac{\alpha}{3}$, and obviously ranges from $-\frac{1}{3}$ to $\frac{1}{3}$.

B.1.4 Dependence Properties

U and V are positively quadrant dependent, positively regression dependent and likelihood ratio dependent for $0 \leq \alpha \leq 1$ (Lai, 1978) and (Drouet Mari & Kotz, 2001).

B.2 Clayton Copula

B.2.1 Formula for Distribution Function

$$C(u,v) = (u^{-\delta} + v^{-\delta} - 1)^{\frac{-1}{\delta}} \quad (0,\infty) \tag{B.3}$$

B.2.2 Formula for Density Function

$$c(u,v) = \frac{\partial^2}{\partial u \partial v} = (\delta + 1)(uv)^{-(\delta+1)}(u^{-\delta} + v^{-\delta} - 1)^{\frac{-2\delta+1}{\delta}} \tag{B.4}$$

B.2.3 Kendall's Tau

$$\tau = \frac{\delta}{\delta + 2} \tag{B.5}$$

B.2.4 Low-Tail Dependence (LT)

$$LT = lim_{u \to 0+} \frac{C(u,v)}{u} = lim_{u \to 0+} \frac{(2u^{-\delta} - 1)^{\frac{-1}{\delta}}}{u} = 2^{\frac{-1}{\delta}} \tag{B.6}$$

B.2.5 Truncation-Invariance Property

This copula has a remarkable invariance under truncation (Oakes, 2005). This property makes it possible to synthesize points ina sub-region sample of a Clayton copula, with one corner at $(0,0)$.

APPENDIX C

R CODE

C.1 R Code to implement algorithm in Section 2.1

```
##############################################################
# Step 1: Generate the three indepen sets of Uniform(0,1)
##############################################################
Ti <- 8
Y_it <- runif(Ti)
V_it <- runif(Ti)
###############################################################
# Step 2: Generate the dependent Uniform Dist through
# the Clayton Copula
###############################################################
delta <- 0.5 #Low degree of dependence for now
U_it <- ( (Y_it^(-delta/(1+delta))-1)*(V_it)^(-delta) + 1)
       ^(-1/delta)
##############################################################
# Step 3: Generate the Number of Claims :
# Zero-Inflated Poisson Distribution
##############################################################
#Zero-Inflated Poisson Distribution is a Modified Poisson
```

```
# Distribution so recall theories of (a,b,1)Class
# of Discrete Dist.
h_it <- runif(Ti); x_it <- runif(Ti) ; z_it <- runif(Ti)
alpha <- 0.5 ; beta <- 0.3; d_it <- 1
lambda <- d_it * exp(x_it*alpha + beta) ;library("boot") ;
phi_it <- inv.logit(alpha*z_it) ;
k = (1 - phi_it) / (1 - exp(-lambda))
    N_it <- c(rep(NA,8))
       p <- c(rep(NA,8))
cummulat <- c(rep(NA,8))
 Counter <- c(rep(NA,8))
for(j in 1:Ti){
       p[j] <- phi_it[j]
       cummulat[j] <- phi_it[j]
       Counter[j] <- c(rep(0,8))[j]
  while(cummulat[j] <= U_it[j]){
       Counter[j] <- Counter[j] + 1
       if(Counter[j] == 1 ){
          p[j] <- k[j] * lambda[j] * exp(-lambda[j])
      }else{ p[j] <- p[j] * lambda[j] / Counter[j] }
          cummulat[j] <- cummulat[j] + p[j]
       }
N_it[j] <- Counter[j]
 }
N_it #Number of claims for insured i at time t.
#####################################################
# Step 4: Generate the Class Vector : Using (2.2)
#####################################################
```

```
  C_it_final <- c(rep(NA,8))
  C_it1 <- c(rep(NA,8))
   C_it <- 5
C_it1[1] <- C_it
C_it_final <- C_it
  counter <- 1
  while(counter <= 8){
   counter <- counter + 1
if(N_it[counter-1] == 0) {C_it1[counter] <- max(1,C_it - 1)
}else{C_it1[counter] <- min(6,C_it + 2*N_it[counter])}
C_it <- C_it1[counter]
C_it_final[counter] <- C_it1[counter]}
C_it_final ; cbind(C_it_final, c(N_it,NA))[1:8,]
samples <- cbind(C_it_final, c(N_it,NA))[1:8,]
samples
###############################################################
```

BIBLIOGRAPHY

Aas, K. (2004). Modelling the dependence structure of financial assets: A survey of four copulas.

Albrecher, H. & Boxma, O. J. (2004). A ruin model with dependence between claim sizes and claim intervals. *Insurance: Mathematics and Economics*, *35*(2), 245–254.

Avanzi, B., Cassar, L. C. & Wong, B. (2011). Modelling dependence in insurance claims processes with lévy copulas. *ASTIN Bulletin: The Journal of the IAA*, *41*(2), 575–609.

Baeumer, B. (2003). On the inversion of the convolution and laplace transform. *Transactions of the American Mathematical Society*, *355*(3), 1201–1212.

Barges, M., Cossette, H., Loisel, S. & Marceau, E. (2011). On the moments of aggregate discounted claims with dependence introduced by a fgm copula. *ASTIN Bulletin: The Journal of the IAA*, *41*(1), 215–238.

Bargès, M., Cossette, H. & Marceau, E. (2009). Tvar-based capital allocation with copulas. *Insurance: Mathematics and Economics*, *45*(3), 348–361.

Bäuerle, N. & Grübel, R. (2005). Multivariate counting processes: copulas and beyond. *ASTIN Bulletin: The Journal of the IAA*, *35*(2), 379–408.

Biard, R., Lefèvre, C., Loisel, S. & Nagaraja, H. N. (2011). Asymptotic finite-time ruin probabilities for a class of path-dependent heavy-tailed claim amounts using poisson spacings. *Applied stochastic models in business and industry*, *27*(5), 503–518.

Böcker, K. & Klüppelberg, C. (2008). Modelling and measuring multivariate operational risk with lévy copulas. *The Journal of Operational Risk*, *3*(2), 3–27.

Boudreault, M. (2003). Modeling and pricing earthquake risk. *Scor Canada Actuarial Prize*.

Bouyé, E., Durrleman, V., Nikeghbali, A., Riboulet, G. & Roncalli, T. (2000). Copulas for finance-a reading guide and some applications.

Campana, A. & Ferretti, P. (2005). On distortion risk measures for sums of discrete and identically distributed risks. *Giornale dell'Istituto Italiano degli Attuari, 68*, 89–104.

Cherubini, U., Luciano, E. & Vecchiato, W. (2004). *Copula methods in finance*. John Wiley & Sons.

Cont, R. & Tankov, P. (2004). *Financial modelling with jump processes*. Chapman and Hall.

Cossette, H., Marceau, E. & Marri, F. (2010). Analysis of ruin measures for the classical compound poisson risk model with dependence. *Scandinavian Actuarial Journal, 2010*(3), 221–245.

Czado, C., Kastenmeier, R., Brechmann, E. C. & Min, A. (2012). A mixed copula model for insurance claims and claim sizes. *Scandinavian Actuarial Journal, 2012*(4), 278–305.

Daley, D. J. & Vere-Jones, D. (2007). *An introduction to the theory of point processes: volume II: general theory and structure*. Springer Science & Business Media.

de Kort, J. (2007). *Modeling tail dependence using copulas*. Springer.

Denuit, M., Dhaene, J., Goovaerts, M. & Kaas, R. (2005). Actuarial theory for dependent risks: Measures, orders, and models. *John Wiley&Sons*.

Denuit, M., Maréchal, X., Pitrebois, S. & Walhin, J.-F. (2007). *Actuarial modelling of claim counts: Risk classification, credibility and bonus-malus systems*. John Wiley & Sons.

Drouet Mari, D. & Kotz, S. (2001). *Correlation and dependence*. World Scientific.

Durante, F. & Sempi, C. (2010). Copula theory: An introduction.

Embrechts, P., Lindskog, F. & McNeil, A. (2001). Modelling dependence with copulas. *Rapport technique, Département de mathématiques, Institut Fédéral de Technologie de Zurich, Zurich*.

Embrechts, P., McNeil, A. & Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls. *Risk management: value at risk and beyond, 176223*.

Embrechts P, Mikosch, T. & Klüppelberg, C. (1997). Modelling extremal events for insurance and finance. *Applications of mathematics—-stochastic modelling and applied probability*.

Esmaeili, H. & Klüppelberg, C. (2010). Parameter estimation of a bivariate compound poisson process. *Insurance: mathematics and economics*, *47*(2), 224–233.

Fisher, N. I. (1997). Copulas. *Encyclopedia of statistical sciences*.

Frees, E. W., Carriere, J. & Valdez, E. (1996). Annuity valuation with dependent mortality. *Journal of Risk and Insurance*, pp. 229–261.

Frees, E. W. & Valdez, E. A. (1998). Understanding relationships using copulas. *North American actuarial journal*, *2*(1), 1–25.

Frees, E. W. & Wang, P. (2005). Credibility using copulas. *North American Actuarial Journal*, *9*(2), 31–48.

Fréchet, M. (1951). Sur les tableaux de corrélation dont les marges sont donnés. *Annales de l'Universite de Lyon, Paris*, *04*, 13–84.

Genest, C. (2011). Everything you always wanted to know about copula modelling but were afraid to ask. *NIPS 2011, Sierra Nevada, Spain*, *01*, 59–78.

Genest, C. & Nešlehová, J. (2007). A primer on copulas for count data. *ASTIN Bulletin: The Journal of the IAA*, *37*(2), 475–515.

Genest, C. & Nešlehová, J. G. (2005). Modeling dependence beyond correlation. *Casualty Actuarial Society*, *01*, 59–78.

Gschlößl, S. & Czado, C. (2007). Spatial modelling of claim frequency and claim size in non-life insurance. *Scandinavian Actuarial Journal*, *2007*(3), 202–225.

Hoeffding, W. (1994). *Scale—invariant correlation theory*, In *The collected works of Wassily Hoeffding*, (pp. 57–107). Springer.

Joe, H. (2015). *Dependence Modeling with Copulas*. CRC Press.

Joe, H. & Xu, J. J. (1996). The estimation method of inference functions for margins for multivariate models.

Kallsen, J. & Tankov, P. (2006). Characterization of dependence of multidimensional lévy processes using lévy copulas. *Journal of Multivariate Analysis*, *97*(7), 1551–1572.

Kettler, P. C. (2006). Lévy-copula-driven financial processes. *Preprint series. Pure mathematics http://urn. nb. no/URN: NBN: no-8076*.

Lai, C. (1978). Morgenstern's bivariate distribution and its application to point processes. *Journal of Mathematical Analysis and Applications*, *65*(2), 247–256.

Lawless, J. F. (2014). *Statistics in Action: A Canadian Outlook*. CRC Press.

Lindskog, F. & McNeil, A. J. (2003). Common poisson shock models: applications to insurance and credit risk modelling. *ASTIN Bulletin: The Journal of the IAA*, *33*(2), 209–238.

Lundberg, F. (1903). *1. Approximerad framställning af sannolikhetsfunktionen: 2. Återförsäkring af kollektivrisker. Akademisk afhandling*. Almqvist &Wiksells.

Mai, J. & Scherer, M. (2014). *Financial engineering with copulas explained*. Springer.

McNeil, A., Frey, R. & Embrechts, P. (2005). Quantitative risk management: concepts, techniques, and tools.

McNeil, A. J. (1997). Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bulletin: The Journal of the IAA*, *27*(1), 117–137.

McNeil, A. J., Frey, R. & Embrechts, P. (2015). *Quantitative risk management: Concepts, techniques and tools*. Princeton university press.

Mikosch, T. (2009). *Non-life insurance mathematics: an introduction with the Poisson process*. Springer Science & Business Media.

Nelsen, R. B. (1998). *An Introduction to Copulas*. Springer.

Nelsen, R. B. (1999). *An Introduction to Copulas*. Springer.

Oakes, D. (1989). Bivariate survival models induced by frailties. *Journal of the American Statistical Association*, *84*(406), 487–493.

Oakes, D. (2005). On the preservation of copula structure under truncation. *Canadian Journal of Statistics*, *33*(3), 465–468.

Pfeifer, D. & Nešlehová, J. (2004). Modeling and generating dependent risk processes for irm and dfa. *ASTIN Bulletin: The Journal of the IAA*, *34*(2), 333–360.

Resnick, S. I. (1987). Extreme values, regular variation, and point processes, volume 4 of applied probability. a series of the applied probability trust.

Ross, S. M. (2014). *Introduction to probability models*. Academic press.

Samson, D. & Thomas, H. (1987). Linear models as aids in insurance decision making: the estimation of automobile insurance claims. *Journal of Business Research*, *15*(3), 247–256.

Scarsini, M. (1984). On measures of concordance. *Stochastica*, *8*(3), 201–218. Retrieved from `http://eudml.org/doc/38916`

Schepsmeier, U. & Stöber, J. (2014). Derivatives and fisher information of bivariate copulas. *Statistical Papers*, *55*(2), 525–542.

Schilling, R. L. (2014). *An Introduction to Lévy and Feller Processes*. (Master's thesis). TU Dresden, Institut fur Mathematische Stochastik, 01062 Dresden, Germany.

Song, P. X.-K., Fan, Y. & Kalbfleisch, J. D. (2005). Maximization by parts in likelihood inference. *Journal of the American Statistical Association*, *100*(472), 1145–1158.

Song, X.-K. & Song, P. X.-K. (2007). *Correlated data analysis: modeling, analytics, and applications*. Springer Science & Business Media.

Tankov, P. (2003). *Financial modelling with jump processes*, volume 2. CRC press.

van Velsen, J. (2012). Parameter estimation of a levy copula of a discretely observed bivariate compound poisson process with an application to operational risk modelling. *arXiv preprint arXiv:1212.0092*.

Wang, S. & Dhaene, J. (1998). Comonotonicity, correlation order and premium principles. *Insurance: Mathematics and Economics*, *22*(3), 235–242.

Yip, K. C. & Yau, K. K. (2005). On modeling claim frequency data in general insurance with extra zeros. *Insurance: Mathematics and Economics*, *36*(2), 153–163.

Yuen, K. C. & Wang, G. (2002). Comparing two models with dependent classes of business. In *Proceedings of the 36th Actuarial Research Conference, ARCH*, volume 134, pp. 135.

Zhao, X. & Zhou, X. (2014). Copula-based dependence between frequency and class in car insurance with excess zeros. *Operations Research Letters*, *42*(4), 273–277.