Year : 2018

Price Optimisation and Statistical Modeling of Dependent Risks

Tamraz Maissa

UNIL | Université de Lausanne

FACULTÉ DES HAUTES ÉTUDES COMMERCIALES

DÉPARTEMENT DE SCIENCES ACTUARIELLES

PRICE OPTIMISATION AND STATISTICAL MODELING OF DEPENDENT RISKS

THÈSE DE DOCTORAT

présentée à la

Faculté des Hautes Études Commerciales
de l'Université de Lausanne

pour l'obtention du grade de
Docteure ès Sciences Actuarielles

par

Maissa TAMRAZ

Directeur de thèse
Prof. Enkelejd Hashorva

Jury

Prof. Olivier Cadot, Président
Prof. François Dufresne, expert interne
Prof. Dmitry Korshunov, expert externe
Prof. Krzysztof Dębicki, expert externe
Prof. David Kalaj, expert externe

LAUSANNE
2018

UNIL | Université de Lausanne

FACULTÉ DES HAUTES ÉTUDES COMMERCIALES

DÉPARTEMENT DE SCIENCES ACTUARIELLES

PRICE OPTIMISATION AND STATISTICAL MODELING OF DEPENDENT RISKS

THÈSE DE DOCTORAT

présentée à la

Faculté des Hautes Études Commerciales
de l'Université de Lausanne

pour l'obtention du grade de
Docteure ès Sciences Actuarielles

par

Maissa TAMRAZ

Directeur de thèse
Prof. Enkelejd Hashorva

Jury

Prof. Olivier Cadot, Président
Prof. François Dufresne, expert interne
Prof. Dmitry Korshunov, expert externe
Prof. Krzysztof Dębicki, expert externe
Prof. David Kalaj, expert externe

LAUSANNE
2018

Unil
UNIL | Université de Lausanne
HEC Lausanne
Le Décanat
Bâtiment Internef
CH-1015 Lausanne

IMPRIMATUR

Sans se prononcer sur les opinions de l'autrice, la Faculté des Hautes Etudes Commerciales de l'Université de Lausanne autorise l'impression de la thèse de Madame Maissa TAMRAZ, licenciée en mathématiques de l'Université Saint-Joseph et titulaire d'un master en sciences actuarielles de l'Université de Lausanne, en vue de l'obtention du grade de docteure ès sciences actuarielles.

La thèse est intitulée :

PRICE OPTIMISATION AND STATISTICAL MODELING OF DEPENDENT RISKS

Lausanne, le 04.06.2018

Le doyen

po [signature]

Jean-Philippe Bonardi

Membres du Jury

Prof. **Olivier Cadot**
Président du jury, Unversité de Lausanne, Vice-doyen de la Faculté des Hautes Études Commerciales.

Prof. **Enkelejd Hashorva**
Directeur de thèse, Université de Lausanne, Département de Sciences Actuarielles.

Prof. **François Dufresne**
Expert interne, Université de Lausanne, Département de Sciences Actuarielles.

Prof. **Dmitry Korshunov**
Expert externe, Lancaster University, Department of Mathematics and Statistics, Lancaster, United Kingdom.

Prof. **Krzysztof Dębicki**
Expert externe, University of Wrocław, Mathematical Institute, Poland.

Prof. **David Kalaj**
Expert externe, University of Montenegro, Department of Mathematics, Podgorica, Montenegro.

University of Lausanne
Faculty of Business and Economics

Doctorate in Actuarial Science

I hereby certify that I have examined the doctoral thesis of

Maissa TAMRAZ

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature: [signature] Date: 30.05.2018

Prof. Enkelejd HASHORVA
Thesis supervisor

University of Lausanne
Faculty of Business and Economics

Doctorate in Actuarial Science

I hereby certify that I have examined the doctoral thesis of

Maissa TAMRAZ

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature: F. Dufresne Date: 31 May 2018

Prof. François DUFRESNE
Internal member of the doctoral committee

University of Lausanne
Faculty of Business and Economics

Doctorate in Actuarial Science

I hereby certify that I have examined the doctoral thesis of

Maissa TAMRAZ

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature: Dmitry Korshunov Date: 30th of May 2018

Prof. Dmitry KORSHUNOV
External member of the doctoral committee

University of Lausanne
Faculty of Business and Economics

Doctorate in Actuarial Science

I hereby certify that I have examined the doctoral thesis of

Maissa TAMRAZ

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature: Krzysztof Dębicki   Date: 31.05.2018

Prof. Krzysztof DĘBICKI
External member of the doctoral committee

University of Lausanne
Faculty of Business and Economics

Doctorate in Actuarial Science

I hereby certify that I have examined the doctoral thesis of

Maissa TAMRAZ

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature: [signature] Date: 30.05.2018

Prof. David KALAJ
External member of the doctoral committee

Acknowledgments

This is to express my sincere gratitude and appreciation towards the people that have shared this journey with me. The thesis did not write itself, and each person has brought a piece into the puzzle.

First, I would like to thank my supervisor, Prof. Enkelejd Hashorva. I had the honor to work closely with him as he has been a real mentor and yet, a vigorous challenger. His proximity and support were unparalleled, always here to bring out the best in you!
Then comes Dr. Raluca Vernic with whom I have driven several projects. Working with her was a real pleasure. I have learned the meaning of determination, but most importantly, the power that comes from positive attitude.

I profoundly acknowledge my co-authors Yaming Yang, Yizhou Bai and especially Dr. Gildas Ratovomirija for the fruitful discussions and valuable input.
I cannot forget Prof. Dufresne, Prof. Korshunov, Prof. Debicki, and Prof. Kalaj for serving as my committee members and for making this journey fruitful, smooth and enjoyable!

I cannot but express some appreciation to my colleagues and friends from the department of Actuarial Science, especially Arian, Youssouf, Heloise, Gildas, Leonard, Charbel, Jose, Anca and William for sharing long hours, stressful deadlines and unforgettable memories.

Last, but not least, I would like to thank my family, my parents Joseph and May and my siblings Joanna and Karl for their unconditional support, encouragement and love, without which I would not have come this far. Thank you for believing in me!

Lausanne, June 2018 *Maissa Tamraz*

Preface

The calculation of the economic capital falls under the Solvency II and is recently one of the main objectives of European insurers and reinsurers on a yearly basis. Typically, these regulations require them to hold a certain amount of capital to cover them from the risk of insolvency. However, the latter requires a proper understanding of the dependence structure between the insurance risks. Moreover, given the high competition in the insurance market, insurers compete intensively on price, quality of service, financial strength and other factors among every line of business. Therefore, a main objective of insurers is also to maximise their profit at the end of the calendar year and stand out from their competitors.

This thesis discusses both objectives and is divided into three parts. The first part combines Chapters 2 and 3. Both chapters explore Price Optimisation in the scope of insurance for a non-life portfolio. The considered optimisations problems are relevant for practicing actuaries and are investigated in details for the new business and the renewal business. The second part examines new dependence models derived from the distribution of the largest claims observed in two insurance portfolios using the copula approach. In Chapter 4, we investigate some distributional and extremal properties of our new model. Whereas, in Chapter 5, we introduce a new class of tractable distribution, the Discrete-Stable distribution, for the claim counting random variable and explore some new dependence properties. Finally, in the last part, we use the Sarmanov distribution to model the dependence of multivariate insurance risks on one hand and the number of claims on the other hand. In Chapter 6, we present closed-type formulas for risk aggregation and capital allocation of mixed Erlang risks joined by the Sarmanov distribution. These formulas are derived in the general framework of stop-loss reinsurance and then in the particular case with no stop-loss reinsurance. In Chapter 7, we derive closed type formulas for multivariate compound distributions with multivariate Sarmanov counting distribution. We then compare the exact values with the ones obtained from the classical recursion-discretization approach.

Contents

Chapter 1

Introduction

Nowadays, European insurers and reinsurers are governed by new regulations requiring them to hold a certain amount of capital to cover their unexpected losses. The calculation of the economic capital falls under the Solvency II framework and is based on a Value-at-Risk approach at 99.5% confidence level. Several factors are taken into account to compute the latter, however one of the main factors is associated with the modeling of the dependence structure observed between insurance risks under the enterprise risk management scheme. Measuring the dependence between insurance risks is important in practice. For instance, Nguyen and Molinari [95] show that ignoring the latter leads to an underestimation of the overall capital the insurer must hold. However, assuming complete dependence between the risks overestimates the overall risk the insurer is facing, hence their capital and as a result generates higher costs to the insurance and reinsurance companies relating to the capital they are holding. Therefore, in order to assess the level of dependence between the different insurance risks, insurers shall compute some statistical measures. The most popular measure of dependence is the linear correlation coefficient, namely *Pearson correlation coefficient*. However, relying solely on the latter is, in general, misleading especially in non-life insurance due to the asymmetry of the loss distributions for some lines of business. Therefore, to model accurately the dependence between the risks, insurers shall determine their loss distributions. In this respect, the copula approach is introduced. Copulas have a long history in the literature, see Genest and MacKay [48], Joe [66], Nelsen [94] and the references therein. Their importance lies in the fact that they capture the whole dependence structure between insurance risks. However, the main difficulty remains in the choice of the copula that will best fit the risk profile of the company, see for instance Genest and Rivest [49], Kole et al. [74] for methods related to the selection of the appropriate copula model.

Although insurers must cope with the new regulations on a yearly basis, they are also concerned in increasing their profit at the end of each calendar year. Given the high competition in the insurance market, specifically in the auto-insurance market, insurers intensively compete on several factors such as price, quality of service, financial strength and many other factors across every class of drivers. One of the main objectives they have to deal with is based on how to generate more sales, retain existing customer and increase their profit simultaneously in order to stand out from their competitors. In this respect, a new and evolving concept is introduced in the scope of insurance namely *Price Optimisation*. Price Optimisation is generally referred to as the practice of increasing or decreasing premium rates of policyholders based on non-related risk factors and some business constraints, see NAIC [92].

In the sequel, we shall explore both objectives of insurers in details. The thesis is structured as follows.

1.1 Price Optimisation

Price Optimisation is a new and evolving area in the actuarial practice for insurance companies. It has already been used for years in the retail and travel industries, see Ferreira et al. [39], Phillips [99], but has never been used in the scope of insurance until recently. Price optimisation is the process of going from traditional actuarial rates to final prices. However, this process involves a deep understanding of the market demand and customer behavior. In the literature, Price Optimisation is sometimes referred to as *Price discrimination*, see Thomas [115]. This designation is justifiable in the concept of *Inertia pricing*, i.e., the act of offering lower prices to new customers who share the same risk profile of existing ones.

In the last 12 years, many insurance companies in Europe have already implemented price optimisation techniques. So far in the literature, there is no precise mathematical description of the optimisation problems solved in such applications. Nonetheless, very recent contributions focus on the issues of price optimisation, mainly from the ethical and regulation points of view, see Marin and Bayley [84], NAIC [92], Schwartz and Harrington [108].

In this thesis, we tackle in Chapters 2 and 3 the following issue: *How can we modify the premiums for given customers within a range of reasonable prices such that the new premiums are optimal under some business constraints?*

In this regard, we consider a non-life insurance portfolio, specifically an auto-insurance portfolio. The main source of portfolio income emanates from the premium volume. In this respect, we consider in both chapters the maximisation of the expected premium volume under some business constraints. We model the demand function in both cases based on the initial price, i.e. before optimisation, the final price, i.e. after optimisation, and some other risk characteristics.

Chapter 2 is dedicated to the price optimisation of the renewal business. We define some objective and constraint functions relevant to insurers. We consider both continuous and discrete optimisation and provide some algorithmic sub-optimal solutions. Also, we explore some simulation techniques for the discrete case.

In Chapter 3, we deal with price optimisation of the new business. Typically, due to high competition in the insurance market, insurers intensively compete on several factors such as price, quality of service, etc. Therefore, we include in our model the competition in the market when setting the optimisation problems.

1.2 Dependence models

Copulas have a long history in the literature, see Genest and MacKay [48], Nelsen [94]. They are popular tools for modeling dependent insurance risks, see Denuit et al. [28], Embrechts et al. [37], Klugman and Parsa [72], Vandenberghe et al. [116] and the references therein. Typically, a copula is defined as the joint distribution function of random variables uniformly distributed on the interval $(0, 1)$. Sklar [111] shows that one can construct multivariate distributions with arbitrary marginal distributions. Indeed, if we consider a d-variate random vector $(X_1, \ldots, X_d)$ with joint distribution function F, then when the marginal distributions F_i of X_i for $i = 1, \ldots, d$ are known, we can write F with respect to the copula C of the d-variate vector as follows

$$F(x_1, \ldots, x_d) = C\left(F_1(x_1), \ldots, F_n(x_d)\right).$$

Moreover, if the marginal distributions F_i of X_i are continuous then the copula C of F is unique and can be written as

$$C(u_1, \ldots, u_d) = F\left(F_1^{-1}(u_1), \ldots, F_n^{-1}(u_d)\right).$$

Copulas share several interesting properties, mainly they are invariant through increasing transformations. The latter is relevant for instance when converting the actual losses into loss payments. Also, to assess the level of dependence between insurance risks with a copula, we can use two other dependence measures namely the Kendall's τ and Spearman's rank correlation ρ_S which are concordant/discordant measures. They are defined as follows for a given copula C

$$\tau(C) = 4\int_0^1\int_0^1 C(u,v)dC(u,v) - 1, \quad \rho_S(C) = 12\int_0^1\int_0^1 C(u,v)dudv - 3.$$

In the sequel, we shall present several applications based on the copula approach.

1.2.1 New dependence models based on mixture copulas

In insurance practice, modeling of multivariate data is crucial in several applications. We count for instance the pricing of dependent insurance risks, risk management of different portfolios, reinsurance modeling of joint risks etc. With motivation from Zhang and Lin [127], we propose, in Chapters 4 and 5, a flexible family of copulas derived from the joint distribution F of the largest claim sizes of two insurance portfolios. Their corresponding collective model over a fixed time period is given in terms of the individual claim sizes of the two portfolios $(X_i, Y_i), i \geq 1$ with distribution function G and a claim counting random variable N. By allowing N to depend on some parameter, say θ, then $F = F(\theta)$ is for various choices of N a tractable parametric family of bivariate distribution functions. If $N = 0$, then there are no claims, so the largest claims in both portfolios are equal to 0. However, when $N \geq 1$, then $(X_{N:N}, Y_{N:N})$ denotes the maximal claim amounts in both portfolios. In this respect, we define $\Lambda = N|N \geq 1$. Inspired by Joe [67][Chapter 4.2], F can be expressed with respect to the Laplace transform of the claim frequency distribution as follows

$$F(x,y) = L_\Lambda\left(-\ln G(x,y)\right), \quad \text{where} \quad \Lambda = N|N \geq 1.$$

In Chapter 4, we assume that Λ belongs to a tractable family of distributions namely Shifted Geometric, Shifted Poisson and Truncated Poisson. We present several applications of the implied parametric models to some data from the literature. Furthermore, we investigate the extremal properties of F and G by Monte Carlo simulation and show that they almost share the same dependence property.

In Chapter 5, we introduce a new class of distributions for N, namely the Discrete-Stable distribution, motivated by the Poisson distribution. By allowing N to be a compound Poisson random variable, then N is given by

$$N = \sum_{i=1}^{Y} Z_i, \quad N = 0 \text{ if } Y = 0, \tag{1.2.1}$$

where Y is a counting random variable independent of Z_i's which are discrete random variables. In this Chapter, we shall focus on a tractable choice for Z_i's, namely these are independent copies of a Sibuya random variable Z with probability generating function (pgf)

$$P_Z(z) = 1 - (1 - z)^{\alpha},$$

with $\alpha \in (0, 1]$ a fixed parameter. For such Z and Y a Poisson random variable with parameter $\lambda > 0$, then N has a discrete-stable distribution with parameters $\lambda > 0$ and $\alpha \in (0, 1]$. Furthermore, we investigate the dependence property of F with respect to both parameters α and λ and present several applications of the new model to concrete insurance data sets.

1.2.2 Dependence models based on the Sarmanov distribution

Let S denote the aggregate claims of an insurance portfolio. In the individual model, S is defined as the sum of each individual claims, say X_i for $i = 1, \dots, n$, i.e. $S = \sum_{i=1}^{n} X_i$ where $X_1, X_2, \dots, X_n$ are mutually independent random variables and n is fixed. By allowing the number of claims n to be random, say N, we define the collective model as

$$S = \sum_{i=1}^{N} X_i,$$

where the individual claims X_i are independent for different i and identically distributed and are independent of N. In the sequel, we shall explore applications of the collective model in the multivariate case with the use of Sarmanov distribution to model on one hand the dependence between the individual claim amounts and on the other hand the dependence between the number of claims in each portfolio. Sarmanov's distribution was first introduced by Sarmanov [107] in the bivariate case, then extended by Lee [77] to the multivariate case. The Sarmanov distribution caught the interest of many researchers in different fields. Vernic [117] considered capital allocation based on the TVaR rule for the Sarmanov distribution with exponential marginals; Cossette et al. [23] used the Farlie-Gumbel-Morgenstern (FGM) distribution to model the dependence between mixed Erlang distributed risks and applied it to capital allocation and risk aggregation; Hashorva and Ratovomirija [59] and Ratovomirija [100] presented aggregation and capital allocation in insurance and reinsurance for mixed Erlang distributed risks joined by the Sarmanov distribution with a specific kernel function. One key advantage of the Sarmanov distribution is its flexibility to join different types of marginals and its allowance to obtain exact results. Its applications in many insurance contexts show its flexible structure when modeling the dependence between multivariate risks given the distribution of the marginals, see the references cited above. For some kernel function ϕ, we define the Sarmanov distribution as follows

$$h(\boldsymbol{x}) = \prod_{i=1}^{n} f_i(x_i) \left(1 + \sum_{1\leq j<l\leq n} \alpha_{j,l}\phi_j(x_j)\phi_l(x_l)\right), \boldsymbol{x} \in \mathbb{R}^n,$$

where ϕ_i and $\alpha_{j,l}$ are the non-constant kernel functions and dependence parameters respectively satisfying the following conditions

$$\mathbb{E}(\phi_i(X_i)) = 0 \text{ for } i = 1, \ldots, n \quad \text{and } 1 + \sum_{1\leq j<l\leq n} \alpha_{j,l}\phi_j(x_j)\phi_l(x_l) \geq 0.$$

Following some recent works of Cossette et al. [23], Hashorva and Ratovomirija [59] and Ratovomirija [100], we present in Chapter 6 some closed-type formulas for risk aggregation and capital allocation of mixed Erlang risks joined by Sarmanov's multivariate distribution by considering a different kernel function for Sarmanov's distribution, not previously studied in this context. The risk aggregation and capital allocation formulas are derived and numerically illustrated in the general framework of stop-loss reinsurance, and then in the particular case with no stop-loss reinsurance. A discussion on the dependence structure of the considered distribution, based on Pearson's correlation coefficient, is also presented for different kernel functions and illustrated in the bivariate case.

In Chapter 7, we consider the multivariate extension of the univariate collective model. In this respect, we consider m insurance portfolios as follows

$$(S_1, ..., S_m) = \left(\sum_{l=0}^{N_1} X_{1l}, ..., \sum_{l=0}^{N_m} X_{ml}\right),$$

where N_j denotes the number of claims of type j and X_{jl} the corresponding claim sizes. We assume that the number of claims are dependent and joined by the Sarmanov distribution. In this chapter, we present closed-type formulas for multivariate compound distributions with multivariate Sarmanov counting distribution and independent Erlang distributed claim sizes. Further on, we also consider a type II Pareto dependence between the claim sizes of a certain type, see Sarabia et al. [106]. The resulting densities rely on the special hypergeometric function. Based on the Pochhammer symbol $(a)_{(n)} = a(a+1) \times ... \times (a+n-1), n \geq 1, (a)_{(0)} = 1$, the generalized hypergeometric function ${}_rF_q$ is defined by

$$ {}_rF_q\left(\{a_1,...,a_r\},\{b_1,...,b_q\};z\right)=1+\sum_{n=1}^{\infty}\frac{(a_1)_{(n)}\times...\times(a_r)_{(n)}}{(b_1)_{(n)}\times...\times(b_q)_{(n)}}\frac{z^n}{n!}. $$

The advantage of this function is that it is already implemented in the usual software. Finally, we numerically illustrate the applicability and efficiency of such formulas by evaluating a bivariate cumulative distribution function, which is also compared with the similar function obtained by the classical recursion-discretization approach.

Chapter 2

Some Mathematical Aspects of Price Optimisation

This chapter is based on E. Hashorva, G. Ratovomirija, M. Tamraz and Y. Bai: Some Mathematical Aspects of Price Optimisation, published in *Scandinavian Actuarial Journal*, 1-25, 2017.

2.1 Introduction

Commonly, insurance contracts are priced based on a tariff, here referred to as the *market tariff*. In mathematical terms such a market tariff is a function say $f : \mathbb{R}^d \to [m, M]$ where m, M are the minimal and the maximal premiums. For instance, a motor third party liability (MTPL) *market tariff* of key insurance market players in Switzerland has $d > 15$. Typically, the function f is neither linear nor a product of simple functions.

In non-life insurance, many insurance companies use different f for new business and renewal business. There are statistical and marketing reasons behind this practice. In this paper we are primarily concerned with non-life renewal business. Yet, some findings are of importance for pricing of insurance and other non-insurance products. We shall first discuss three important actuarial tasks and then present various mathematical aspects of relevance for pricing actuaries.

Practical actuarial task T1: *Given that a portfolio of N policyholders is priced under a given market tariff f, determine an optimal market tariff f^* that will be applied in the next portfolio renewal.*

Typically, actuarial textbooks are concerned with the calculation of the pure pre-

mium, which is determined by applying different statistical and actuarial methods to historical portfolio data, see e.g., Asmussen [13], Embrechts et al. [36], Rolski et al. [104], Rotar [105]. The tariff that determines the pure premium of a given insurance contract will be here referred to as the *pure risk tariff*. In mathematical terms this is a function say $g : \mathbb{R}^{d_1} \to [m_1, M_1]$ with $d_1 \geq 1$. In the actuarial practice, pure premiums are loaded, for instance for large claims, provisions, direct expenses and other costs (overheads, profit, etc.). Actuarial mathematics explains various approaches for loading insurance premiums; in practice very commonly a linear loading is applied. We shall refer to the function that is utilised for the calculation of the premium of an insurance coverage based on the costs related to that coverage as *actuarial tariff*; write $g_A : \mathbb{R}^{d_2} \to [m_2, M_2]$ for that function.
Despite the importance of task **T1**, the current actuarial literature has not dealt with its mathematical aspects. On the other side, practicing actuaries are constantly confronted with various black-box type solutions available from external services or in few cases have developed their own internal models.

Practical actuarial task T2: *Given a pure risk tariff g, construct an optimal actuarial tariff g_A that includes various premium loading.*

Since by definition there is no unique optimal *actuarial tariff*, the calculations leading to it can be performed depending on the resources of pricing and implementation team.
To this end, let us briefly mention an instance which motivates **T2**: Suppose for simplicity that the insurance portfolio in question consists of two groups of policyholders A and B with n_A and n_B policyholders, respectively. All the contracts are to be renewed, say at the next 1st of January. The pricing actuary calculates the *actuarial tariff* which shows that for group A, the yearly premium to be paid from each policyholder is 2'000 CHF and for group B, say 500 CHF. For this portfolio, overhead expenses (or expenses not directly allocated to an insurance policy) are calculated (estimated) to be X CHF for the next insurance period. The amount X can be distributed to $N = n_A + n_B$ policyholders in different ways, for instance each policyholder will have to pay $X/(n_A + n_B)$ of those expenses. Another alternative approach could be to calculate it as a fix percentage of the pure premiums. The principal challenge for pricing actuaries is that the policyholders are already in the portfolio and might be very sensitive to any change of their premiums, especially when the insurance risk does not change.
At renewal (abbreviated as $@\mathfrak{R}$ in the following) given that the insured risk does

not change, if the new offered premium is different from the current one, the policyholder can cancel the contract. Clearly, another common reason for canceling the insurance contract is also the competition in the insurance market. Consequently, the solutions of **T1-T2** need to take into account the probability of renewal of the policies at the point of renewal.

As illustrated above for **T2**, the percentage of premium increase δ_i for the i^{th} policyholder @$\mathcal{R}$ can be fixed, i.e., $\delta_i \in \Delta$ where say $\Delta = \{0\%, 5\%, 10\%\}$. Such increases are often common in practice especially if the distribution channel is primarily dominated by the tied agents. A clear advantage of such type of tariff modification is that it can be straightforwardly implemented with minimal implementation costs. Therefore, instead of **T2** a simpler task which is very often encountered in the insurance practice (but surprisingly not in actuarial literature) is as follows:

<u>Practical actuarial task T3</u>: *Modify for any $i \le N$ the premium P_i of the i^{th} policyholder @$\mathcal{R}$ by a fixed percentage, say $\delta_i \in \Delta_i$ with Δ_i a discrete set (for instance $\Delta_i = \{0\%, 5\%, 10\%\}$) so that the new set of premiums*

$$P_i^* = P_i(1+\delta_i), \quad 1 \le i \le N$$

are optimal under several business constraints. Moreover, determine the new market tariff f^* which yields P_i^*'s.

There are several difficulties related to the solutions of the actuarial tasks **T1-T3**. In practice the *market tariff* is very complex for key insurance coverage such as motor or household insurance. A typical f utilised in insurance practice is as follows (consider only two arguments for simplicity)

$$f(x,y) = \min\Big(M_0, \max(e^{ax+by}, m_0 + m_1 x + m_2 y)\Big). \tag{2.1.1}$$

The exponential term in f is very common in practice since both claim frequency and average claim sizes are modelled using generalised linear models (GLM's) with log-link function. The reason for the choice of log-link functions is the ease of IT implementation. Both min and max functions in (2.1.1) prevent the premiums from being extremes. These are often decided by empirical findings and insurance market constraints.

Even if we know the optimal P_i^*'s that solve **T3**, when the structure of f (and also of f^*) is fixed say as in (2.1.1), then the existence of an optimal f^* that gives exactly P_i^*'s is in general not guaranteed. Note that due to technical reasons, the actuaries can change the coefficients that determine f, say a, b and so on, but the structure of the tariff, i.e., the form of f in (2.1.1) is in general fixed when preparing a new renewal tariff due to implementation costs.

The main goal of this contribution is to discuss various mathematical aspects that lead to optimal solutions of both actuarial tasks **T1** and **T3**. Further we analyse eventual implementations of our optimisation problems for renewal business. Optimisation problems related to new business are more involved and will therefore be treated in a forthcoming contribution.
To this end, we note that in the last 12 years many insurance companies in Europe have already implemented price optimisation techniques. Very recent contributions focus on the issues of price optimisation, mainly from the ethical and regulation points of view, see Marin and Bayley [84], NAIC [92], Schwartz and Harrington [108]. It is important to note that optimality issues in insurance and reinsurance business, not directly related to the problems treated in this contribution, have been discussed in various contexts, see e.g., Asimit et al. [6, 7, 10], Chi and Meng [21], Golubin [52], Huang et al. [64], Zhang et al. [128] and the references therein.
Brief organisation of the rest of the paper: Section 2 describes the different optimisation settings from the insurer's point of view. In Section 3, we provide partial solutions for problem **T3**. Section 4 describes the different algorithms used to solve the optimisation problems followed by some insurance applications to the motor line of business presented in Section 5.

2.2 Objective functions and Business Constraints

2.2.1 Theoretical Settings

For simplicity, and without loss of generality, we shall assume that the renewal time is fixed for all $i = 1, \ldots, N$ policyholders already insured in the portfolio with the i^{th} policyholder paying the insurance premium P_i for the current insurance period. Each policyholder can be insured for different insurance periods. Without loss of generality, we shall suppose that @$\mathcal{R}$ each insurance contract has the option to be renewed for say one year, with a renewal premium $P_i^* := P_i(1 + \delta_i)$.
Suppose that the renewal probability for the i^{th} contract is a function of P_i and some parameters describing the risk characteristics of the policyholder. At renewal, by changing the premium, this probability will depend on the premium change δ_i, the previous premium P_i and other risks characteristics. Therefore we shall assume

that this probability is

$$\Psi_i(P_i, \delta_i), \tag{2.2.1}$$

where Ψ_i is a strictly positive function depending eventually on i (when the risk characteristics of the i^{th} policyholder are tractable). This is a common assumption in logistic regression, where Ψ_i is the inverse of the logit function (called also expit), or Ψ_i is a univariate distribution function.

In order to consider the renewal probabilities in the tariff and premium optimisation tasks, the actuary needs to know/determine $\Psi_i(P_i, \delta_i)$ for any $\delta_i \in \Delta_i$, where Δ_i is the range of possible changes of premium (commonly $0 \in \Delta_i$). Estimation of Ψ_i's is non-trivial; it can be handled for instance using logistic regression, see Section 2.4.1 below for more details.

In practice, depending on the market position and the strategy of the insurance company, different objective functions can be used for the determination of an optimal *actuarial tariff* or *market tariff*. We discuss below two important objective functions:

O1) Maximise the future expected premium volume @$\mathfrak{R}$:

In our model, the current premium volume for the portfolio in question is $V = \sum_{i=1}^{N} P_i$, whereas the premium volume in case of complete renewal is

$$V^* = \sum_{i=1}^{N} P_i^* = \sum_{i=1}^{N} P_i(1 + \delta_i).$$

Given the fact that not all policies might renew, let us denote by $N_{@\mathfrak{R}}$ the random number of policies which will be renewed. Since we can treat each contract as an independent risk, then

$$N_{@\mathfrak{R}} = \sum_{i=1}^{N} I_i,$$

with $I_1, \ldots, I_N$ independent Bernoulli random variables satisfying

$$\mathbb{P}\{I_i = 1\} = \Psi_i(P_i, \delta_i), \quad 1 \leq i \leq N.$$

Clearly, the expected percentage of the portfolio to renew is given by (set below $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_N)$)

$$\theta(\boldsymbol{\delta}) = \frac{\mathbb{E}\{N_{@\mathcal{R}}\}}{N} = \sum_{i=1}^{N} \frac{\mathbb{E}\{I_i\}}{N} = \frac{1}{N}\sum_{i=1}^{N} \Psi_i(P_i, \delta_i). \tag{2.2.2}$$

The premium volume @$\mathcal{R}$ (which is random) will be denoted by $\mathcal{V}_{@\mathcal{R}}$. It is simply given by

$$\mathcal{V}_{@\mathcal{R}} := \sum_{i=1}^{N} I_i P_i (1 + \delta_i).$$

Consequently, considering the interest in maximising the premium volume, then the objective function in this setting is given by

$$q_{vol}(\boldsymbol{\delta}) := \mathbb{E}\{\mathcal{V}_{@\mathcal{R}}\} = \sum_{i=1}^{N} P_i(1+\delta_i)\mathbb{E}\{I_i\} = \sum_{i=1}^{N} P_i(1+\delta_i)\Psi_i(P_i, \delta_i). \tag{2.2.3}$$

Note that $P_1, \ldots, P_N$ are known, therefore the optimisation will be performed with respect to δ_i's only.

O1') <u>Minimise the variance of $\mathcal{V}_{@\mathcal{R}}$</u>: If the variance of $\mathcal{V}_{@\mathcal{R}}$ is large, the whole renewal process can be ruined. Therefore along **O1** the minimisation of the variance of $\mathcal{V}_{@\mathcal{R}}$ is important. In this model we have

$$q_{var}(\boldsymbol{\delta}) := Var(\mathcal{V}_{@\mathcal{R}}) = \sum_{i=1}^{N} [P_i(1+\delta_i)]^2 \Psi_i(P_i, \delta_i)[1 - \Psi_i(P_i, \delta_i)]. \tag{2.2.4}$$

O2) <u>Maximise the expected premium difference @$\mathcal{R}$</u>: Let $\tau_i = P_i\delta_i$ be the premium difference for the i^{th} policyholder and set $\boldsymbol{\tau} := (\tau_1, \ldots, \tau_N)$. The total premium difference @$\mathcal{R}$ is $\sum_{i=1}^{N} I_i\tau_i$, with expectation

$$q_{dif}(\boldsymbol{\tau}) = \mathbb{E}\left\{\sum_{i=1}^{N} I_i\tau_i\right\} = \sum_{i=1}^{N} \tau_i \Psi_i(P_i, \delta_i). \tag{2.2.5}$$

It is not difficult to formulate other objective functions, for instance related to the classical ruin probability, Parisian ruin (see e.g., Debicki et al. [25]), or future solvency and market position of the insurance company. Moreover, the objective functions can be formulated over multiple insurance periods. Due to the nature of insurance business, there are several constraints that should be taken into account for the renewal business optimisation, see Asimit et al. [9] and the references therein. Typically, the most important business constraints relate to the strategy of the company and the concrete insurance market. We formulate few important constraints below:

C1) Expected retention level $@\mathcal{R}$ should not be less say than 70%. Although the profit and the volume of premiums at renewal are important, all insurance companies are interested in keeping most of the policyholders in their portfolios. Therefore there is commonly a lower bound imposed on the expected retention level ℓ at renewal. For instance $\ell \geq 90\%$ means that the expected percentage of customers that will not renew their contracts should not exceed 10%. In mathematical terms, this is formulated as

$$\theta_{rlevel}(\boldsymbol{\delta}) = \frac{\mathbb{E}\left\{N^*\right\}}{N} \geq \ell. \tag{2.2.6}$$

C2) A simple constraint is to require that the renewal premiums P_i^*'s are not too different from the "old" ones, i.e.,

$$\delta_i \in [a, b], \quad \tau_i := P_i \delta_i \in [A, B], \quad 1 \leq i \leq N \tag{2.2.7}$$

for instance $a = -5\%, b = 10\%$ and $A = -50, B = 300$.

Several other constraints including those related to reputational risk, decrease of provision level for tied-agents, and loss of loyal customers can be formulated similarly and will therefore not be treated in detail.

2.2.2 Practical Settings

In insurance practice the cost of optimisation itself (actuarial and other resources) needs to be also taken into account. Additionally, since the total volume of premiums at renewal is large, an optimal renewal tariff is of interest (business relevant) only if it produces a significant improvement to the current tariff. Therefore, for practical implementations, we need to redefine the objective functions. For a given positive constant c, say $c = 1'000$, we redefine (2.2.3) as

$$q_{vol}^c(\delta_1, \ldots, \delta_N) := c\left\lfloor \mathbb{E}\left\{\mathcal{V}_{@\mathcal{R}}\right\}/c\right\rfloor = c\left\lfloor \sum_{i=1}^{N} P_i(1+\delta_i)\Psi_i(P_i, \delta_i)/c\right\rfloor, \tag{2.2.8}$$

where $\lfloor x \rfloor$ denotes the largest integer smaller than x. Similarly, we redefine (2.2.4) as

$$c\Big\lfloor Var(\mathcal{V}_{@\mathfrak{R}})/c\Big\rfloor \quad = \quad c\Big\lfloor \sum_{i=1}^{N} \{P_i(1+\delta_i)\}^2 \Psi_i(P_i,\delta_i)[1-\Psi_i(P_i,\delta_i)]/c\Big\rfloor. \quad (2.2.9)$$

Finally, (2.2.5) can be written as

$$q^c_{dif}(\tau_1,\ldots,\tau_N) = c\Big\lfloor \sum_{i=1}^{N} \tau_i \Psi_i(P_i,\delta_i)/c\Big\rfloor. \qquad (2.2.10)$$

For implementation purposes and due to business constraints, τ_i's can be assumed to be certain given positive integers.
Therefore a modification of (2.2.7) can be formulated as

$$\delta_i \in [a,b] \cap (c_1^{-1}\mathbb{Z}), \quad \tau_i := P_i\delta_i \in [A,B] \cap (c_1\mathbb{Z}), \quad 1 \le i \le N, \qquad (2.2.11)$$

where $c_1 > 0$, for instance $c_1 = 100$. Such modifications of both objective functions and constraints show that for practical implementation, there is no unique optimal solution of the optimisation problem of interest.

Remarks 2.2.1. *i) If two different insurance contracts are renewed through different distributional channels, then typically different constraints are to be applied to each of those policies. Additionally, the cancellation probabilities could be different, even in the case where both policyholders have the same risk profile. Therefore, in order to allow for different distributional channels, we only need to adjust the constraints and assume an appropriate cancellation pattern.*
ii) From the practical point of view, Ψ_i's are estimated by using for instance logistic regression. At random, customers are offered $@\mathfrak{R}$ higher/lower premiums than their P_i's i.e., δ_i's are chosen randomly with respect to some prescribed distribution function. An application of the logistic regression to the data obtained (renewal/non renewal) explains the cancellation (or renewal) probability in terms of risk factors as well as other predictors (social status, etc.). In an insurance market dominated by tied-agents this approach is quite difficult to apply.
iii) Different policyholders can renew their contracts for different periods. This case is included in our assumptions above.
iv) Most tariffs utilised in practice, for instance an MTPL one, consist of hundreds of coefficients (typically more than 300). Due to a dominating product structure, modern insurance tariffs consists of many individual cells, say 200'000 in average. However, most of these tariff cells are empty and typically less than 15% of the cells

determine 80% of the total premium volume in the portfolio. For instance, it is quite rare that a Ferrari is insured for a TPL risk by a 90 years old lady, living in a very small village. With this in mind, typically, the relevant number N in practical optimisation problems does not exceed 30'000. Our algorithms and simulation methods work fairly well for such N.

2.3 Solutions for T3

The main difficulty when dealing with the actuarial task **T3** lies on the complexity of Ψ_i's since these functions are:
a) in general not known,
b) difficult to estimate if past data are partially available,
c) even when these functions are known, the constraints **C1-C2** and the objective functions **O1,O1',O2** are in general not convex. We discuss next a partial solution for **T3**.
Problem T3a: Given $P_1, \ldots, P_N$ determine $\boldsymbol{\delta}^* = (\delta_1^*, \ldots, \delta_N^*)$ such that

$$q_{vol}(\boldsymbol{\delta}^*) \text{ is maximal}, \quad q_{var}(\boldsymbol{\delta}^*) \text{ is minimal} \tag{2.3.1}$$

under the constraints

$$\theta_{rlevel}(\boldsymbol{\delta}) \geq \ell, \quad \boldsymbol{l} \leq \boldsymbol{\delta} \leq \boldsymbol{u},$$

where $\boldsymbol{l}$ and $\boldsymbol{u}$ are 2 vectors such that their components $l_i,\ u_i \in (-1,1)$ for $i \leq N$.
Problem T3b: Determine the market tariff f^* from $P_1^*, \ldots, P_N^*$.
The solution (an approximate one) of **T3b** can be easily derived. Given $P_1^*, \ldots, P_N^*$, and since the structure of the *market tariff* is known, then f^* can be determined (approximately) by running a non-linear regression analysis with response variables P_i^*'s.
Below we shall focus on task **T3a** dealing with the determination of the optimal premiums P_i^*'s at renewal. In insurance practice, the functions $\Psi_i, i \leq n$ can be assumed to be piece-wise linear and non-decreasing. This assumption is indeed reasonable, since for very small τ_i or δ_i the cancelation probability should not change. However, that assumption can be violated if for instance at renewal the competition modifies also their new business premiums. For simplicity, these cases will be excluded in our analysis, and thus we assume that the decision for accepting the renewal offer is not influenced by the competition.
We list below some tractable choices for Ψ_i's:

Ma) Suppose that for given known constants π_i, a_i, b_i

$$\Psi_i(P_i, \delta_i) = \pi_i(1 + a_i\delta_i + b_i\delta_i^2), \quad 1 \le i \le N.$$

In practice, π_i, a_i, b_i need to be estimated. Clearly, the case that b_i's are equal to 0 is quite simple and tractable.
Note in passing that a simple extension of the above model is to allow a_i and b_i to differ depending on the sign of δ_i.

Mb) One choice motivated by the logistic regression model commonly used for estimation of renewal probabilities is the expit function, i.e.,

$$\Psi_i(P_i, \delta_i) = \frac{1}{1 + c_i^{-1} e^{-T_i\delta_i}}, \quad 1 \le i \le N,$$

where c_i, T_i's are known constants (to be estimated in applications), see e.g., Guillén et al. [55] .
We note that Model **Ma)** can be seen as an approximation of Model **Mb**).

Mc) Finally, we consider the case where Ψ_i's are determined only for specific δ_i's. For instance, for the i^{th} policyholder Ψ_i depends on P_i and δ_i as follows

index	1	2	3	4	5	6	7	8	9
δ_i(in %)	-20 %	-15 %	-10%	-5%	0%	5%	10%	15%	20%
$P_i(1+\delta_i)$	80	85	90	95	100	105	110	115	120
$\Psi_i(P_i, \delta_i)$	0.999	0.995	0.990	0.975	0.950	0.925	0.900	0.875	0.825

Table 2.3.1: Renewal probabilities as a function of premiums of the i^{th} policyholder.

The Model **Ma**) is simple and tractable and can be seen as an approximation of a more complex one. Moreover, it leads to some crucial simplification of the objective functions in question.

2.4 Optimisation Algorithms

2.4.1 Maximise the expected premium volume @ℛ

In this section, we consider the objective function **O1** subject to the constraint function **C1**. Our optimisation problem can be formulated as follows

$$\begin{aligned} &\max_{\boldsymbol{\delta}} q_{vol}(\boldsymbol{\delta}), \quad \boldsymbol{\delta} := (\delta_1, \dots, \delta_N) \\ &\text{subject to } \theta_{rlevel}(\boldsymbol{\delta}) \geqslant \ell, \\ &\boldsymbol{l} \leq \boldsymbol{\delta} \leq \boldsymbol{u}, \end{aligned} \tag{2.4.1}$$

where q_{vol} and θ_{rlevel} are defined respectively in (2.2.3) and (2.2.6). Further $\boldsymbol{l} := (l_1, \dots, l_N)$, $\boldsymbol{u} := (u_1, \dots, u_N)$ are such that $l_i, u_i \in (-1, 1)$ for $i \leq N$.

Probability of renewal Ψ_i as in Ma).

We consider the case where the probability of renewal Ψ_i is

$$\Psi_i := \Psi_i(P_i, \delta_i) = \pi_i(1 + a_i\delta_i + b_i\delta_i^2).$$

- Setting $b_i = 0$, we have

$$\Psi_i := \Psi_i(P_i, \delta_i) = \pi_i(1 + a_i\delta_i). \tag{2.4.2}$$

Since $\Psi_i \in (0, 1)$ should hold for all policyholders $i \leq N$, we require that

$$a_i \in (1 - \frac{1}{\pi_i}, \quad \frac{1}{\pi_i} - 1), \quad \delta_i \in (-1, 1), \quad \pi_i > 0$$

for all $i \leq N$.
The assumption $b_i = 0$ implies that (2.4.1) is a quadratic programming (QP) problem subject to linear constraints. It has a global maximum if and only if its objective function is concave, which is the case when $a_i < 0$. Hence we shall assume that $a_i \in (1 - \frac{1}{\pi_i}, 0)$ for any $i \leq N$.
Scenario 1: We consider the optimisation problem (2.4.2) without the upper and lower bounds constraints. In view of (2.4.2), the optimisation problem (2.4.1) can be reformulated as follows

$$\begin{aligned} &\min_{\boldsymbol{\delta}} f(\boldsymbol{\delta}) = \frac{1}{2}\boldsymbol{\delta}^\top Q\boldsymbol{\delta} + \boldsymbol{c}^\top\boldsymbol{\delta}, \quad \boldsymbol{\delta} = (\delta_1, \dots, \delta_N)^\top, \\ &\text{subject to } g(\boldsymbol{\delta}) = \boldsymbol{a}^\top\boldsymbol{\delta} - b \leq 0, \end{aligned} \tag{2.4.3}$$

where

$$\boldsymbol{c} = (-\pi_1 P_1(1+a_1), \ldots, -\pi_N P_N(1+a_N))^\top$$

describes the coefficient of the linear terms of f, Q is a diagonal and positive definite matrix describing the coefficients of the quadratic terms of f determined by

$$Q = \begin{pmatrix} -2\pi_1 P_1 a_1 & 0 & 0 & \ldots & 0 \\ 0 & -2\pi_2 P_2 a_2 & 0 & \ldots & 0 \\ 0 & \ldots & -2\pi_i P_i a_i & \ldots & 0 \\ 0 & 0 & 0 & \ldots & -2\pi_N P_N a_N \end{pmatrix}.$$

Since (2.4.3) has only one constraint, $\boldsymbol{a}$ is a vector related to the linear coefficients of g and is given by

$$\boldsymbol{a} = -(\pi_1 a_1, \pi_2 a_2, \ldots, \pi_N a_N)^\top.$$

Furthermore, we have that

$$b = \sum_{i=1}^{N} \pi_i - N\ell.$$

Note in passing that the constant term of the objective function f is not accounted for in the resolution of (2.4.3).
Next, we define the Lagrangian function

$$\mathcal{L}(\boldsymbol{\delta}, \lambda) = f(\boldsymbol{\delta}) + \lambda g(\boldsymbol{\delta}),$$

where λ is the Lagrangian multiplier.
Given that Q is a positive definite matrix, the well-known Karush-Kuhn-Tucker (KKT) conditions (see for details Luenberger and Ye [81][p. 342])

$$\begin{cases} \nabla\mathcal{L}(\boldsymbol{\delta}^*, \lambda^*) & = & 0, \\ \lambda^* g(\boldsymbol{\delta}^*) & = & 0, \\ g(\boldsymbol{\delta}^*) & \leq & 0, \\ \lambda^* & \geq & 0 \end{cases} \tag{2.4.4}$$

are sufficient for a global minimum of (2.4.3) if they are satisfied for a given vector $(\boldsymbol{\delta}^*, \lambda^*)$. Thus, in the sequel, we provide an explicit solution for this type of optimisation problem. Typically, (2.4.3) can be reduced to the Markowitz mean-variance optimisation problem, see Markowitz [85], Markowitz et al. [86]. Setting $\boldsymbol{\delta_1} = \boldsymbol{\delta} + Q^{-1}\boldsymbol{c}$, then (2.4.3) can be expressed as the following standard quadratic program

$$\begin{aligned} &\min_{\boldsymbol{\delta_1}} \frac{1}{2}\boldsymbol{\delta_1}^\top Q \boldsymbol{\delta_1}, \\ &\text{subject to } \boldsymbol{a_1}^\top \boldsymbol{\delta_1} \le b_1, \end{aligned} \tag{2.4.5}$$

with $b_1 = b + \boldsymbol{a}^\top Q^{-1}\boldsymbol{c}$. It should be noted that the constant term (when replacing $\boldsymbol{\delta_1}$ by $\boldsymbol{\delta} + Q^{-1}\boldsymbol{c}$ in (2.4.5)) does not play any role in the resolution of (2.4.5).

Let $\boldsymbol{\delta_1^*}$ be the optimal solution of (2.4.5). The KKT conditions defined in (2.4.4) can be explicitly written as follows:

$$\begin{cases} Q\boldsymbol{\delta_1^*} + \lambda^* \boldsymbol{a} = \boldsymbol{0}, & (2.4.6a) \\ \lambda^*(\boldsymbol{a}^\top \boldsymbol{\delta_1^*} - b_1) = 0, & (2.4.6b) \\ \boldsymbol{a}^\top \boldsymbol{\delta_1^*} - b_1 \le 0, & (2.4.6c) \\ \lambda^* \ge 0, & (2.4.6d) \end{cases}$$

where $\boldsymbol{0} = (0, \ldots, 0)^\top \in \mathbb{R}^N$.

If $\lambda^* = 0$, then $\boldsymbol{\delta_1^*} = \boldsymbol{0}$ follows directly from (2.4.6a). Therefore,

$$\boldsymbol{\delta^*} = -Q^{-1}\boldsymbol{c}.$$

In view of (2.4.6d) the other possibility is $\lambda^* > 0$, which in view of (2.4.6b) implies $\boldsymbol{a}^\top \boldsymbol{\delta_1^*} = b_1$. Further from (2.4.6a) $\boldsymbol{\delta_1^*} = -\lambda^* Q^{-1}\boldsymbol{a}$, hence

$$\boldsymbol{\delta^*} = -Q^{-1}(\lambda^* \boldsymbol{a} + \boldsymbol{c}),$$

with $\lambda^* = -(\boldsymbol{a}^\top Q^{-1}\boldsymbol{a})^{-1} b_1$.

Scenario 2: We consider that (2.4.3) has lower and upper bounds constraints. Thus, the optimisation problem at hand can be formulated as follows

$$\begin{aligned} &\min_{\boldsymbol{\delta}} \frac{1}{2}\boldsymbol{\delta}^\top Q \boldsymbol{\delta} + \boldsymbol{c}^\top \boldsymbol{\delta}, \\ &\text{subject to } \boldsymbol{a}^\top \boldsymbol{\delta} - b \le 0, \\ &\boldsymbol{l} \le \boldsymbol{\delta} \le \boldsymbol{u}. \end{aligned} \tag{2.4.7}$$

The constraints in (2.4.7) can be grouped into one equation

$$A\boldsymbol{\delta} \geq \boldsymbol{d},$$

where A is a $(2N+1) \times N$ matrix and $\boldsymbol{d}$ a vector of dimension $2N+1$ respectively given by

$$A = \begin{pmatrix} \pi_1 a_1 & \pi_2 a_2 & \pi_3 a_3 & \dots & \pi_N a_N \\ 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \dots & \vdots \\ 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & 0 & 1 \\ -1 & 0 & 0 & \dots & 0 \\ 0 & -1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \dots & \vdots \\ 0 & 0 & \dots & -1 & 0 \\ 0 & 0 & \dots & 0 & -1 \end{pmatrix}, \quad \boldsymbol{d} = \begin{pmatrix} -b \\ l_1 \\ l_2 \\ \vdots \\ l_{N-1} \\ l_N \\ -u_1 \\ -u_2 \\ \vdots \\ -u_{N-1} \\ -u_N \end{pmatrix}.$$

Since we have a convex objective function and a convex region given by constraints, the solution $\boldsymbol{\delta}^*$ is unique and we can transform the above optimisation problem to a bound constrained one using duality. Hence, (2.4.7) can be rewritten as follows

$$\begin{aligned} &\min_{\boldsymbol{\delta}} f(\boldsymbol{\delta}) = \frac{1}{2}\boldsymbol{\delta}^\top C \boldsymbol{\delta} - \widetilde{\boldsymbol{c}}^\top \boldsymbol{\delta}, \\ &\text{subject to } g(\boldsymbol{\delta}) = \boldsymbol{\delta} \geq \mathbf{0}, \end{aligned} \tag{2.4.8}$$

with $C = AQ^{-1}A^\top$ a square matrix of dimension $(2N+1)$ and $\widetilde{\boldsymbol{c}} = AQ^{-1}\boldsymbol{c} + \boldsymbol{d}$ a vector of dimension $(2N+1)$. Since Q is an $N \times N$ positive definite matrix and A has rank N, then $C = AQ^{-1}A^\top$ is positive definite.
If $\boldsymbol{\delta}^*$ is the solution of (2.4.8), then $\boldsymbol{\delta}^* = Q^{-1}\left(A^\top \boldsymbol{\delta}^* - \boldsymbol{c}\right)$ is the solution of (2.4.7). We refer to Voglis and Lagaris [120] for the description of the algorithm and Appendix A.

- Hereafter we shall assume that $b_i \neq 0$ implying that Ψ_i is of the form

$$\Psi_i := \Psi_i(P_i, \delta_i) = \pi_i(1 + a_i\delta_i + b_i\delta_i^2). \tag{2.4.9}$$

Given $b_i \in (-1,0)$ and $\delta_i \in (-1,1)$, the condition $\Psi_i \in (0,1)$ holds if and only if

$$a_i \in \Big(\max\,(1-\frac{1}{\pi_i}, -1-b_i), \min\,(1+b_i, \frac{1}{\pi_i}-1)\Big)$$

for $i \leq N$. Clearly, under (2.4.9) we have that (2.4.1) is a non-linear optimisation problem with also non-linear constraints. The most popular method discussed in the literature for solving this type of optimisation problem is the *Sequential Quadratic Programming* (SQP) method see e.g., Bartholomew-Biggs [16], Boggs and Tolle [18], Nickel and Tolle [96]. It is an iterative method that generates a sequence of quadratic programs to be solved at each iteration. Typically, at a given iterate x_k (2.4.1) is modelled by a QP subproblem subject to linear constraints and then solution to the latter is used as a search direction to construct a new iterate x_{k+1}.

Plugging (2.4.9) in (2.4.1), the optimisation problem at hand can be reformulated as

$$\begin{aligned}
&\min_{\boldsymbol{\delta}}\ f(\boldsymbol{\delta}) = -\sum_{i=1}^{N} P_i\pi_i(1+(1+a_i)\delta_i + (a_i+b_i)\delta_i^2 + b_i\delta_i^3),\\
&\text{subject to}\\
&g(\boldsymbol{\delta}) = -\sum_{i=1}^{N}\pi_i(1+a_i\delta_i+b_i\delta_i^2) + N\ell \leq 0,\\
&h_1(\delta_i) = \delta_i - u_i \leq 0 \text{ for } i \leq N,\\
&h_2(\delta_i) = -\delta_i + l_i \leq 0 \text{ for } i \leq N,
\end{aligned} \tag{2.4.10}$$

where f, g, h_1 and h_2 are continuous and twice differentiable.

The main steps required to solve (2.4.10) are described in Appendix A.

Probability of renewal Ψ_i as in Mb).

We consider the following model for the renewal probability:

$$\Psi_i := \Psi_i(P_i,\delta_i) = \frac{1}{1+c_i^{-1}e^{-T_i\delta_i}}, \quad 1 \leq i \leq N, \tag{2.4.11}$$

where c_i is a constant that depends on the probability of renewal for $\delta_i = 0$ denoted by π_i given by

$$c_i = \frac{\pi_i}{1-\pi_i}$$

and $T_i < 0$ is a constant (to be estimated in applications) that measures the elasticity of the policyholder relative to the premium change. The greater $|T_i|$ the more elastic the policyholder is to premium change. Under (2.4.11) we have that (2.4.1) is a non-linear optimisation problem subject to non-linear constraints, which can be solved by SQP algorithm described in Appendix A.

Remarks 2.4.1. *If δ_i are close to 0, then using Taylor expansion **Mb)** can be approximated by **Ma)** as follows*

$$\Psi_i(P_i, \delta_i) \approx \frac{c_i}{1+c_i}\left(1 + \frac{c_i T_i}{1+c_i}\delta_i - \frac{T_i^2(c_i - 1)}{2(1+c_i)^2}\delta_i^2\right),$$

where

$$\pi_i = \frac{c_i}{1+c_i}, \quad a_i = \frac{c_i T_i}{1+c_i}, \quad b_i = -\frac{T_i^2(c_i-1)}{2(1+c_i)^2}. \tag{2.4.12}$$

Probability of renewal Ψ_i as in Mc).

In this model δ_i belongs to a discrete set, which we shall assume hereafter to be

$$\mathbf{D} = \{-20\%, -15\%, -10\%, -5\%,\ 0\%,\ 5\%,\ 10\%,\ 15\%,\ 20\%\}.$$

Also, the renewal probabilities Ψ_i's are fixed for each insured i based on δ_i for $i \leq N$ as defined in Table 2.3.1. In this section we deal with a *Mixed Discrete Non-Linear Programming* (MDNLP) optimisation problem. Thus, (2.4.1) can be reformulated as follows

$$\begin{aligned} &\min_{\boldsymbol{\delta}} f(\boldsymbol{\delta}) = -\sum_{i=1}^{N} P_i(1+\delta_i)\Psi_i(P_i, \delta_i), \\ &\text{subject to } g(\boldsymbol{\delta}) = -\sum_{i=1}^{N} \Psi_i(P_i, \delta_i) + N\ell \leqslant 0, \\ &\text{and } \delta_i \in \mathbf{D}, \quad 1 \leq i \leq N. \end{aligned} \tag{2.4.13}$$

In general, this type of optimisation problem is very difficult to solve due to the fact that the discrete space is non-convex. Several methods were discussed in the literature for (2.4.13), see e.g., Arora et al. [4]. The contribution Loh and Papalambros [80] proposed a new method for solving the MDNLP optimisation problem subject to non-linear constraints. It consists in approximating the original non-linear model by a sequence of mixed discrete linear problems evaluated at each point iterate $\boldsymbol{\delta_k}$.

Also, a new method for solving a MDNLP was introduced by using a penalty function, see the recent contributions Luenberger and Ye [81], Ma and Zhang [82] for more details. The algorithmic solution of (2.4.13) is described in Appendix B.

2.4.2 Maximise the retention level $@\mathcal{R}$

We consider the case where the insurer would like to keep the maximum number of policyholders in the portfolio $@\mathcal{R}$. Therefore the optimisation problem of interest consists in finding the optimal retention level $@\mathcal{R}$ whilst increasing the expected premium volume by a fixed amount say C in the portfolio. Hence, the optimisation problem can be formulated as follows

$$\begin{aligned}
&\max_{\boldsymbol{\delta}} \frac{1}{N}\sum_{i=1}^{N}\Psi_i(P_i,\delta_i),\\
&\text{subject to} \\
&\mathbb{E}(P^*) \geqslant \mathbb{E}(P) + C,\\
&\boldsymbol{l} \leq \boldsymbol{\delta} \leq \boldsymbol{u},
\end{aligned} \tag{2.4.14}$$

where $\mathbb{E}(P^*) = \sum_{i=1}^{N} P_i(1+\delta_i)\Psi_i(P_i,\delta_i)$ is the expected premium volume $@\mathcal{R}$, $\mathbb{E}(P) = \sum_{i=1}^{N} P_i\pi_i$ is the expected premium volume before premium change and C is a fixed constant which can be expressed as a percentage of the expected premium volume before premium change. We remark that C can be interpreted as a certain premium loading. Clearly, (2.4.14) is a non-linear optimisation problem, which can be solved by using the SQP algorithm already described in Appendix A).

2.5 Insurance Applications

In this section, we consider a simulated dataset that describes the production of the motor line of business of an insurance portfolio. We simulate premiums from an exponential random variable with mean $1'204$. Also, the probability of renewal before premium change, π_i for $i = 1, \ldots, N$, are known and estimated by the insurance company for each category of policyholders based on historical data. Given that the behaviour of the policyholders is unknown at the time of renewal, the probability of renewal Ψ_i, depends on π_i and δ_i for $i = 1, \ldots, N$. If δ_i is positive, then Ψ_i decreases whereas if δ_i is negative, it is more likely that the policyholder will renew the insurance policy, thus generating a greater Ψ_i. In the following paragraphs, we are going to present some results related to the optimisation problems formulated in the last section.

2.5.1 Optimisation problem Ma)

Maximise the expected premium volume @$\mathfrak{R}$.

We consider, first, the optimisation problem defined in (2.4.1). In this case, the probability of renewal Ψ_i is defined in **Ma)** and set $b_i = 0$ for $i = 1, \ldots, N$. Given that $a_i < 0$ for $i \leq N$, the probability of renewal Ψ_i increases when δ_i is negative and decreases when δ_i is positive, thus describing perfectly the behaviour of the policyholders that are subject to a decrease, respectively increase, in their premiums @$\mathfrak{R}$. The table below describes some statistics on the data for $10'000$ policyholders.

	Premium at time 0
Min	200
Q1	491
Q2	909
Q3	1'605
Max	9'061
No. Obs.	10'000
Mean	1'204
Std. Dev.	990

Table 2.5.1: Production statistics for the motor business.

We consider the constraint that the expected percentage of the policyholders to remain in portfolio @$\mathfrak{R}$ is at least 85%. By solving (2.4.1) in Matlab with the function *quadprog*, we obtain the optimal δ for each policyholder. We denote by t_0 the time before premium change and by t_1 the time after premium change. Figure 6.4.1 below is a comparative histogram describing the number of policyholders at time t_0 and at time t_1 with respect to the different premium ranges and the average optimal δ for each premium range.

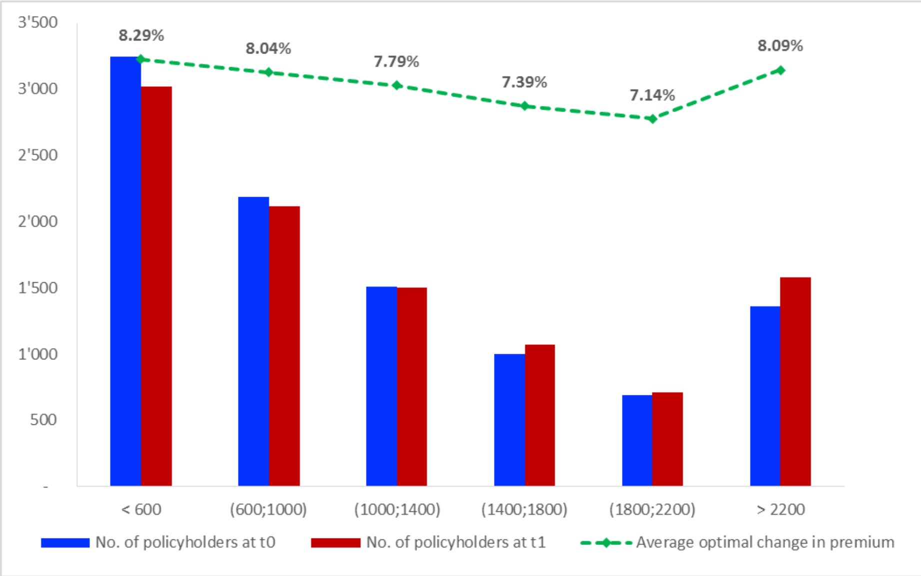


Figure 2.5.1: Number of policyholders based on premium range.

As seen in Figure 6.4.1, 32% of the policyholders have a premium below 600 CHF vs. 30% @$\mathfrak{R}$ due to an average optimal increase in premium of 8%. The average optimal δ decreases gradually for premiums between 600 CHF and 2′200 CHF. Premiums above 2′200 CHF account for only 16% of the portfolio with an average optimal δ of 8%. Typically, in practice, insurance companies are likely to increase the tariffs of policyholders with low premiums as a small increase in the price will not have a great impact on the renewal of the policy. However, for policies with large premium amount, a small increase in the price can lead to the surrender of the policy. Therefore, the results in Figure 6.4.1 are accurate from the insurance company's perspective when increasing/ decreasing the premiums paid by the policyholders. It should be noted here that we neglect the cases of bad risks and large claims. We look at a homogeneous portfolio where the occurrence of a claim is low and the claim amounts are reasonable.

Next, we consider two scenarios:
Scenario 1 The expected percentage of the policyholders (abbreviate as EPP) to remain in portfolio @$\mathfrak{R}$ is at least 75%,
Scenario 2 The EPP to remain in portfolio @$\mathfrak{R}$ is at least 85%.
Table 2.5.2 below summarises the optimal results when solving (2.4.1) and examines the effect of both scenarios on the expected premium volume and the expected number of policyholders in the portfolio @$\mathfrak{R}$.

	Scenario 1		Scenario 2	
Range of δ (%)	(-10,20)	(-20,30)	(-10,20)	(-20,30)
Growth in expected premium volume @$\mathfrak{R}$ (%)	15.78	23.03	8.70	12.96
Growth in expected number of policies @$\mathfrak{R}$ (%)	-3.52	-5.25	-0.16	-0.16
Average optimal delta (%)	19.99	29.90	7.97	11.89
Number of increases	10'000	10'000	6'196	6'528
Number of decreases	-	-	3'804	3'472

Table 2.5.2: Scenarios testing.

Scenario 1 The optimal δ for both bounds corresponds approximately to the maximum value (upper bound) of the interval. This is mainly due to the fact that EPP @$\mathfrak{R}$ to remain in portfolio is at least 75%. Therefore, the main goal is to maximise the expected premium volume at time t_1.

Scenario 2 For EPP @$\mathfrak{R}$ to remain in portfolio of at least 85%, Table 2.5.2 shows an increase in the expected premium volume which is less important than the one observed in Scenario 1. However, the expected number of policyholders in the portfolio @$\mathfrak{R}$ is higher and is approximately the same as at t_0.

Hereafter, we shall consider EPP @$\mathfrak{R}$ to remain in portfolio to be at least 85%. Commonly in practice the size of a motor insurance portfolio exceeds $10'000$ policyholders. However, solving the optimisation problems for $\boldsymbol{\delta}$ using the described algorithms when N is large requires a lot of time and heavy computation and may be costly for the insurance company. Thus, an idea to overcome this problem is to split the original portfolio into subportfolios and compute the optimal $\boldsymbol{\delta}$ for the subportfolios. One criteria that can be taken into account for the split is the amount of premium in our case. However, in practice, insurance companies have a more detailed dataset, thus more information on each policyholders, so the criterion that are of interest for the split include the age and gender of the policyholders, the car type, age and value. Table 2.5.3 and Table 2.5.4 below describe the results when splitting the original portfolio into three and four subportfolios, respectively.

		Growth in % @$\mathfrak{R}$	
Premium Range	Average optimal δ	$\mathbb{E}\{N_{@\mathfrak{R}}\}$	$\mathbb{E}\{\mathcal{V}_{@\mathfrak{R}}\}$
< 600	8.60%	-0.27%	9.17%
(600,1'200)	7.29%	-0.03%	8.25%
$> 1'200$	8.05%	-0.17%	8.99%
After the split	8.00%	-0.16%	8.84%
Before the split	7.97%	-0.16%	8.70%
Difference	-	0%	-0.13%

Table 2.5.3: Split into 3 subportfolios.

		Growth in % @$\mathfrak{R}$	
Premium Range	Average optimal δ	$\mathbb{E}\{N_{@\mathfrak{R}}\}$	$\mathbb{E}\{\mathcal{V}_{@\mathfrak{R}}\}$
< 500	8.99%	-0.34%	9.50%
(500, 800)	6.27%	0.15%	7.41%
(800, 1'400)	7.66%	-0.09%	8.49%
$> 1'400$	8.47%	-0.26%	9.31%
After the split	7.99%	-0.16%	8.95%
Before the split	7.97%	-0.16%	8.70%
Difference	-	0%	-0.23%

Table 2.5.4: Split into 4 subportfolios.

In Table 2.5.3 and 2.5.4, we consider that the insurer would like to keep 85% of the policyholders in each subportfolios, thus a total of 85% of the original portfolio. However, in practice, the constraints on the retention level @$\mathfrak{R}$ are specific to each subportfolios. In this regard, the insurance company sets the constraints on the expected number of policies for each subportfolios so that the constraint of the overall portfolio is approximately equal to 85%. The error from the split into three, respectively four subportfolios is relatively small and is of -0.13%, respectively -0.23% for the expected premium volume @$\mathfrak{R}$.

Remarks 2.5.1. *i) This application is mostly relevant when dealing with a non linear optimisation problem of a large insurance portfolio.*
ii) In the following sections, we limit the size of the insurance portfolio to 1'000 policyholders as the algorithms used thereafter to solve the optimisation problems are based on an iterative process which is computationally intensive.

Maximise the expected premium volume and minimise the variance of the premium volume

Similarly to the asset allocation optimisation problem in finance introduced by Markowitz Markowitz [87], the insurer performs a trade-off between the maximum aggregate expected premiums and the minimum variance of the total earned premiums, see e.g., Asimit et al. [11] for a different optimality criteria.
We present in Figure 2.5.2 the comparison of the optimal results computed with the function *gamultiobj* of *Matlab 2016a* for the following scenarios:

- **Scenario 1**: the expected premium volume and the variance of the premium volume are optimised simultaneously as in **Problem T3a**,
- **Scenario 2**: only the expected premium volume is maximised.

The same constraint on the retention level is used for both scenarios and $\delta \in (-30\%, 30\%)$. The histograms in Figure 2.5.2 represent the optimal variance whilst the dashed curves depict the optimal expected volume. We notice that all the optimal results are normalised with the results obtained from the assumption that the insurer will not change the premiums for next year.

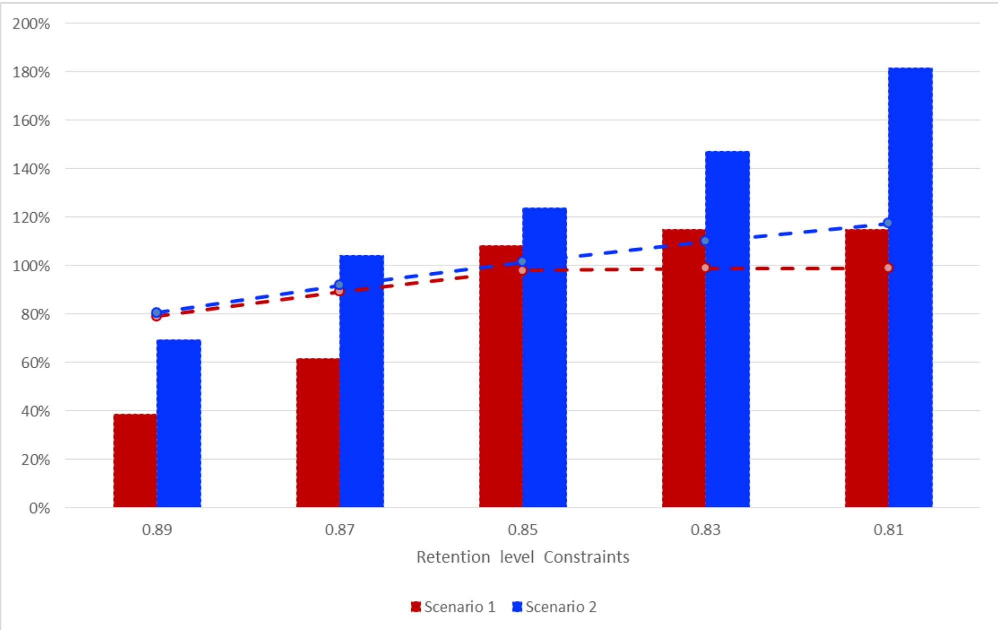


Figure 2.5.2: Optimal results with different objective function scenarios.

For both Scenarios, the maximum expected volume is associated with higher variance. In this respect the lower the retention level the higher the expected volume and the higher the variance. Furthermore, compared to Scenario 2, Scenario 1 results in smaller expected volume but yields a smaller variance. We show next in

Table 2.5.5 the optimal results for the different constraints on the retention level and the possible range of premium changes.

	75%		85%	
Bounds for δ	$(-10, 20)$	$(-20, 30)$	$(-10, 20)$	$(-20, 30)$
Aggregate expected future premiums @$\mathfrak{R}$ (%)	103.57	103.66	99.90	103.90
Variance of the aggregate future premiums @$\mathfrak{R}$ (%)	109.76	113.08	98.41	101.05
Expected number of policies @$\mathfrak{R}$ (%)	98.95	98.84	99.98	99.99
Average optimal δ (%)	6.13	6.82	1.68	1.98
Average optimal increase (%)	18.50	26.92	11.82	20.32
Average optimal decrease (%)	-8.33	-16.32	-8.92	-17.10
Number of increases	539	535	511	510
Number of decreases	461	465	489	490

Table 2.5.5: Scenario 1 optimal results based on different retention levels .

It can be seen that the optimal variance @$\mathfrak{R}$ increases with the range of the possible premium changes $\boldsymbol{\delta}$. For instance when the insurer would like to keep at least 75% of the policyholders, the variance @$\mathfrak{R}$ increases from 109.76 for $\delta \in (-10\%, 20\%)$ to 113.08 for $\delta \in (-20\%, 30\%)$, respectively. Furthermore, the increase in variance @$\mathfrak{R}$ is associated with an increase of the expected volume @$\mathfrak{R}$. This means that the riskier the portfolio the more the insurance company earns premiums.

Maximise the retention level @$\mathfrak{R}$.

We consider here that the insurer would like to maximise the EPP @$\mathfrak{R}$ to remain in portfolio whilst increasing the expected premium volume @$\mathfrak{R}$ by a certain amount C needed to cover, for instance, the operating costs and other expenses of the insurance company. Figure 2.5.3 below describes the results obtained from solving the optimisation problem (2.4.14) defined in Section 2.4.2 using the *fmincon* function in *Matlab* for $C = 95'000$ and $\delta \in (-10\%, 20\%)$.

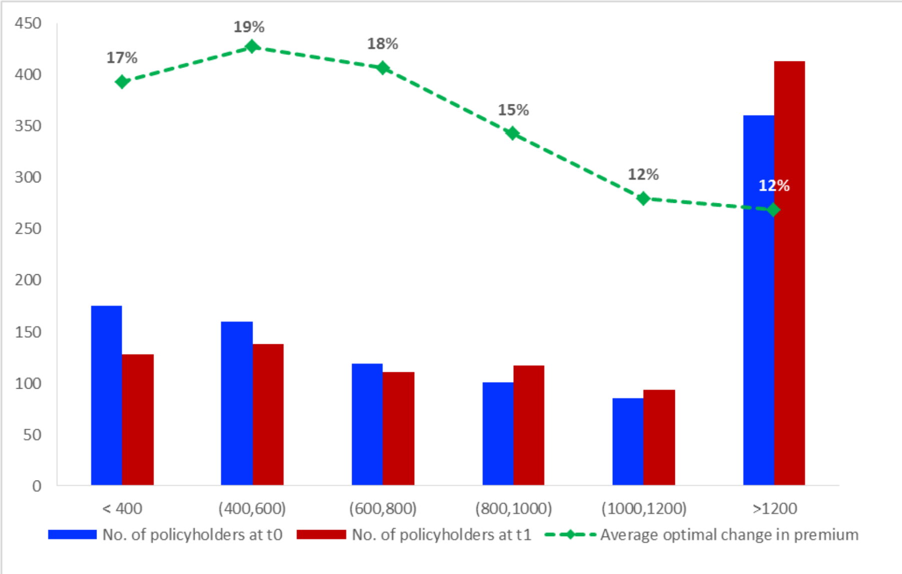


Figure 2.5.3: Number of policyholders based on premium range.

In practice, the amount C needed to cover the expenses of the company is set by the insurers. In fact, C can be expressed as a percentage of the expected premium volume at time t_0. Therefore, we consider three different loadings: 9%, 10% and 11% thus adding an amount of $85'000$, respectively $95'000$ and $105'000$ to the expected premium volume at time t_0. We consider two ranges for δ, namely $\delta \in (-10\%, -20\%)$ and $\delta \in (-20\%, -30\%)$.

	$C = 85'000$		$C = 95'000$		$C = 105'000$	
Bounds for δ	(-10,20)	(-20,30)	(-10,20)	(-20,30)	(-10,20)	(-20,30)
Growth in expected number of policies(%)	-2.19	-2.06	-2.50	-2.36	-2.82	-2.67
Growth in expected premium volume(%)	8.90	8.90	9.95	9.95	11.00	11.00
Average optimal δ (%)	13.92	15.82	15.12	17.23	16.60	18.64

Table 2.5.6: Scenario testing - Retention

Table 2.5.6 shows that when C increases, the expected number of policyholders @$\mathfrak{R}$ decreases whereas the average optimal δ increases.

2.5.2 Optimisation problem Mb)

We consider the probability of renewal Ψ_i as defined in **Mb)**. As discussed in Section 2.4.1, T_i describes the behaviour of the policyholders subject to premium change.

For instance, let us consider a policyholder whose probability of renewal without premium change π_i is 0.95. Figure 2.5.4 shows that the greater T_i the more the curve of the renewal probability goes to the right thus the less elastic the policyholder is to premium change. Conversely, as T_i decreases, the more elastic the policyholder is to premium change.

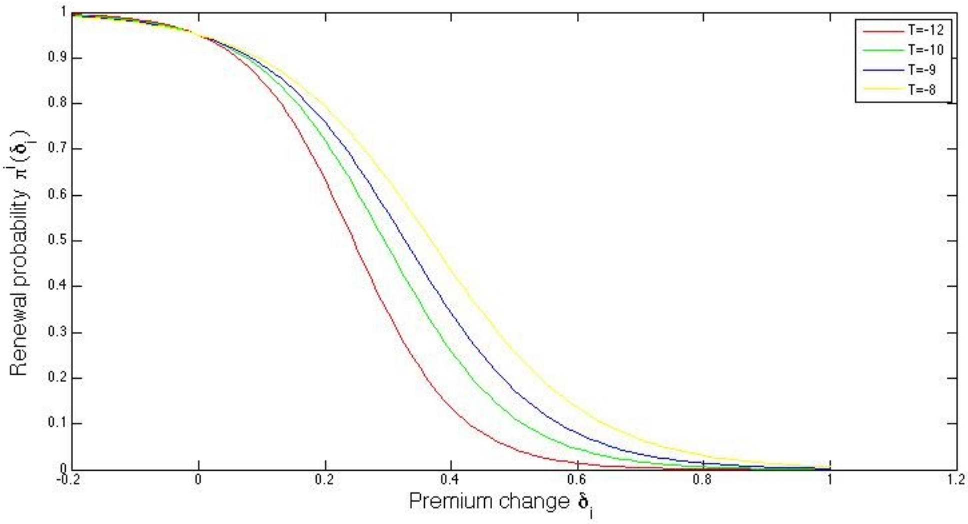


Figure 2.5.4: Renewal probability with respect to premium change for different values of T_i

In this section, we will only consider the case where the insurer would like to maximise the expected premium volume @$\mathcal{R}$. The constraint on the retention level is assumed to be of 85%.

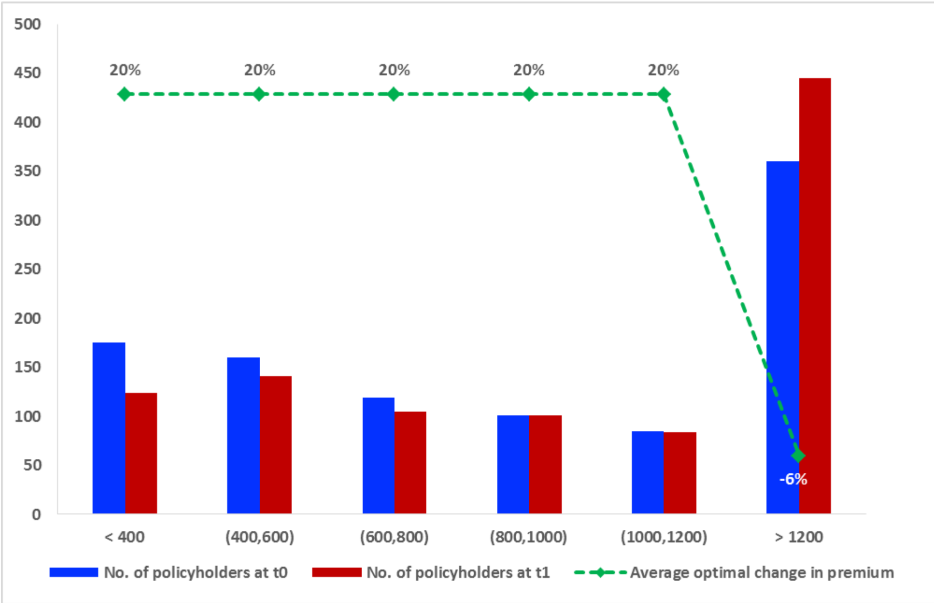


Figure 2.5.5: Number of policyholders based on premium range.

Figure 2.5.5 shows that the average optimal δ for premiums less than 1'200 CHF is constant for the different premium ranges at 20% which corresponds to the maximum value that δ can take. However, for premiums greater than 1'200 CHF, the average optimal δ decreases to -6%. As stated in Section 2.5.1, insurers are more likely to increase the premiums of policyholders with small premium amounts and decrease the premiums of policyholders with large premium amounts. Thus, the results obtained in Figure 2.5.5 are accurate as they describe the behaviour of the insurer when increasing, respectively decreasing the premiums of the policyholders.

At the time of renewal, the insurer sets the constraints on EPP to remain in portfolio. Typically, when the retention level is low, the expected premium volume @$\mathfrak{R}$ is greater compared to the case when the retention level is high. Therefore, we consider two different scenarios:

Scenario 1 The EPP @$\mathfrak{R}$ to remain in portfolio is at least 75%,

Scenario 2 The EPP @$\mathfrak{R}$ to remain in portfolio is at least 85%.

The table below summarises the optimal results when solving (2.4.1) for the different constraints.

Bounds for δ	Scenario 1 (-10,20)	Scenario 1 (-20,30)	Scenario 2 (-10,20)	Scenario 2 (-20,30)
Growth in expected premium volume @$\Re$ (%)	17.84	26.45	4.50	6.48
Growth in expected number of policies @$\Re$ (%)	-0.93	-1.41	-0.02	-0.02
Average optimal delta (%)	20.00	30.00	10.70	16.09
Number of increases	1'000	1'000	703	736
Number of decreases	-	-	297	264

Table 2.5.7: Scenarios testing.

Scenario 1 Table 2.5.7 shows that all policyholders are subject to an increase in their premiums and the average optimal δ for the whole portfolio corresponds to the maximum change in premium for both bounds of δ.
Scenario 2 As seen in Table 2.5.7, the expected number of policyholders @$\Re$ is approximately the same as the one before premium change. However, the growth in expected premium volume is lower than in Scenario 1 due to the fact that the average optimal δ for both bounds is lower.

Remarks 2.5.2. *It should be noted that the probability of renewal defined in* ***Mb)*** *can be approximated by the probability of renewal defined in* ***Ma)*** *for δ relatively small (refer to Remark 2.4.1). Therefore, consider $\delta \in (-5\%, 5\%)$ and a retention level $\ell = 85\%$ @$\Re$. The table below describes the optimal results when using the logit model* ***Mb)*** *and the polynomial model defined in* ***Ma)****.*

Model	Logit	Polynomial	Difference*
Growth in expected premium volume @$\Re$	1.53%	0.47%	1.04%
Growth in expected number of policies @$\Re$	-0.02%	-0.02%	0%
Average optimal delta	2.97%	1.30%	NR
Number of increases	796	619	NR
Number of decreases	204	381	NR

Table 2.5.8: Comparison between **Ma)** and **Mb)**.
*(*NR = Not Relevant)*

Table 2.5.8 shows that for a small range of δ, the difference between the exact results obtained from ***Mb)*** *and the approximate results obtained from* ***Ma)*** *is relatively* *small and is of around 1% for the expected premium volume @$\mathfrak{R}$ and is of 0% for the expected number of policyholders @$\mathfrak{R}$. Thus, the approximate values tend to the real ones when the range of δ tends to 0.*

2.5.3 Optimisation problem Mc) and Simulation studies

In this Section, we consider the case where the renewal probabilities Ψ_i are fixed for each insured i, as defined in Table 2.3.1. To solve the optimisation problem (2.4.13), we use the MDNLP method described in Appendix B. The table below summarises the optimal results for a portfolio of 100'000 policyholders with respect to different constraints on the retention level at renewal.

Retention level constraints (%)	85	87.5	90	92.5	95	97.5
Growth in expected premium volume @$\mathfrak{R}$ (%)	5.92	5.92	5.34	4.19	2.22	-1.24
Growth in expected number of policies @$\mathfrak{R}$ (%)	-7.89	-7.89	-5.26	-2.63	0.00	2.63
Average optimal delta (%)	15.00	15.00	10.00	4.82	-0.51	-6.37

Table 2.5.9: Scenario testing-Discrete optimisation

Table 2.5.9 shows that when the retention level increases, the expected number of policies increases whereas the expected premium volume @$\mathfrak{R}$ decreases. In fact, the average optimal $\boldsymbol{\delta}$ decreases gradually from 15% for a retention level of 85% to -6% for a retention level of 97.5%. Also, it can be seen that for a retention level of 95% the optimisation has a negligible effect on the expected number of policies and premium volume @$\mathfrak{R}$ as the average optimal $\boldsymbol{\delta}$ is approximately null. Hence, no optimisation is needed in this case.

In addition to the MDNLP approach, we have implemented a simulation technique which consists in simulating the premium change $\boldsymbol{\delta}$ for each policyholder as described in the following pseudo algorithm:

- **Step 1:** Based on a chosen prior distribution for $\boldsymbol{\delta}$, sample the premium change for each policyholder,
- **Step 2:** Repeat **Step 1** until the constraint on the retention level is satisfied,
- **Step 3:** Repeat **Step 2** m times,
- **Step 4:** Among the m simulations take the simulated $\boldsymbol{\delta}$ which gives out the maximum expected profit.

Next, we present the optimal results obtained through $1'000$ simulations for the same portfolio. We shall consider three different assumptions on the prior distribution of $\boldsymbol{\delta}$, namely:

- **Case 1: Simulation based on the Uniform distribution**
  In this simulation approach, we assume that the prior distribution of $\boldsymbol{\delta}$ is uniform. As highlighted in Table 2.9.1- 2.9.2, the parameters of the uniform distribution and the possible values of the premium change are chosen so that the constraint on the retention level is fulfilled. Actually, this choice is based on many simulations trials that we have implemented in which for a fixed range of $\boldsymbol{\delta}$, the parameter of the Uniform distribution is modified at each trial so that the retention level is reached. It should be noted that the more the elements of $\boldsymbol{\delta}$, the smaller the bounds of the Uniform distribution. We present in Table 2.5.10 the simulation results.

Retention level constraints (%)	85	87.5	90	92.5	95	97.5
Growth in expected premium volume @$\mathfrak{R}$ (%)	5.13	5.11	4.02	1.87	-0.55	-4.10
Growth in expected number of policies @$\mathfrak{R}$ (%)	-10.32	-6.64	-5.24	-2.61	0.04	2.73
Average optimal delta (%)	17.30	12.62	9.95	4.87	-0.36	-6.52

Table 2.5.10: Simulation approach based on the Uniform distribution.

- **Case 2: Simulation based on practical experience**
  Next, we assume a prior distribution for $\boldsymbol{\delta}$ which is based on the historical premium change of each policyholder. Those prior distributions are presented in Figure 2.9.1 and the results are described in Table 2.5.11 below.

Retention level constraints (%)	85	87.5	90	92.5	95	97.5
Growth in expected premium volume @$\mathfrak{R}$ (%)	5.50	5.08	4.10	1.98	-0.87	-3.75
Growth in expected number of policies @$\mathfrak{R}$ (%)	-9.19	-7.70	-5.26	-2.63	0.47	2.77
Average optimal delta (%)	16.2	13.9	10.0	4.92	-1.17	-6.25

Table 2.5.11: Simulation approach based on practical experience

- **Case 3: Simulation based on the results of the MDNLP**
  We use the empirical distribution of the optimal δ obtained from the MDNLP

algorithm as a prior distribution. The chosen distribution are shown in Figure 2.9.1 with different constraints on the retention level. Table 2.5.12 below summarises the optimal results.

Retention level constraints (%)	85	87.5	90	92.5	95	97.5
Growth in expected premium volume @$\mathfrak{R}$ (%)	5.92	5.92	3.90	1.61	-0.91	-4.05
Growth in expected number of policies @$\mathfrak{R}$ (%)	-7.89	-7.89	-5.26	-2.63	0.00	2.63
Average optimal delta (%)	15.00	15.00	10.00	4.82	-0.51	-6.35

Table 2.5.12: Simulation approach based on empirical distribution.

It can be seen that the simulation approaches yield approximately to the same results obtained from the MDNLP algorithm presented in Table 2.5.9.

2.6 Appendix A: Solution of (2.4.8)

Let $I = \{1, \ldots, N\}$. The Lagrangian function related to (2.4.8) is

$$\mathcal{L}(\boldsymbol{\delta}, \boldsymbol{\lambda}) = f(\boldsymbol{\delta}) + \boldsymbol{\lambda} g(\boldsymbol{\delta}),$$

where $\boldsymbol{\lambda} := (\lambda_1, \ldots, \lambda_N)$ are the Lagrangian multipliers. In this case, the KKT conditions below

$$\begin{cases} C\boldsymbol{\delta}^* - \widetilde{\boldsymbol{c}} + \boldsymbol{\lambda}^* & = \mathbf{0}, \\ \lambda_i^* \delta_i^* & = 0, \forall i \in I \\ \delta_i^* & \geq 0, \forall i \in I \\ \lambda_i^* & \geq 0, \forall i \in I \end{cases} \tag{2.6.1}$$

hold for $(\boldsymbol{\delta}^*, \boldsymbol{\lambda}^*)$.

Let L be a subset of I such that $\lambda_i > 0$ and $\delta_i^* = 0$ for all $i \in L$ if L is non-empty. Note that $\lambda_i = 0$ for $i \notin L$. We have that if L is empty, then by (3.6.3)

$$\boldsymbol{\delta}^* = C^{-1}\widetilde{\boldsymbol{c}}. \tag{2.6.2}$$

Next suppose that L is non-empty and set $R = I \setminus L$. If R is empty, then the solution is found to be on the boundary as above, i.e, $\delta_i^* = 0$ for all i. If R is non-empty, we have that $\lambda_i = 0$ for any $i \in R$. We need to determine $\boldsymbol{\delta}_R^*$ which is the subvector of $\boldsymbol{\delta}^*$ determined by dropping the components with indices not in R. Since C is positive-definite, then C_{RR} the submatrix of C determined by dropping the rows and columns with indices not in R is positive definite and therefore non-singular. In view of (3.6.3) we obtain the solution

$$\boldsymbol{\delta}_R^* = (C_{R,R})^{-1}(-C_{R,L}\boldsymbol{\delta}_L^* + \widetilde{\boldsymbol{c}}_R) = (C_{R,R})^{-1}\widetilde{\boldsymbol{c}}_R$$

and

$$\boldsymbol{\lambda}_L^* = \widetilde{\boldsymbol{c}}_L - C_{L,L}\boldsymbol{\delta}_L^* - C_{L,R}\boldsymbol{\delta}_R^* = \widetilde{\boldsymbol{c}}_L - C_{L,R}\left((C_{R,R})^{-1}\widetilde{\boldsymbol{c}}_R\right).$$

For practical implementation, it is necessary to determine the index set L and this can be achieved by an iterative approach, see Voglis and Lagaris [120].

2.7 Appendix B: Solution of (2.4.10)

Step 1: Let

$$\mathcal{L}(\boldsymbol{\delta}, \lambda, \boldsymbol{\mu}, \boldsymbol{\gamma}) = f(\boldsymbol{\delta}) + \lambda g(\boldsymbol{\delta}) + \sum_{i=1}^{N} \mu_i h_1(\delta_i) + \sum_{i=1}^{N} \gamma_i h_2(\delta_i)$$

be the Lagrangian function of (2.4.10) where $\lambda \in \mathbb{R}$ and $\boldsymbol{\mu}, \boldsymbol{\gamma} \in \mathbb{R}^N$ are the Lagrangian multipliers and ($\boldsymbol{\delta_0}$,λ_0,$\boldsymbol{\mu_0}$,$\boldsymbol{\gamma_0}$) an initial estimate of the solution. It should be noted that the SQP is not a feasible point method. This means that neither the initial point nor the subsequent iterate ought to satisfy the constraints of the optimisation problem.

Step 2: In order to find the next point iterate $(\boldsymbol{\delta_1}, \lambda_1, \boldsymbol{\mu_1}, \boldsymbol{\gamma_1})$, the SQP determines a step vector $\boldsymbol{s} = (\boldsymbol{s_\delta}, s_\lambda, \boldsymbol{s_\mu}, \boldsymbol{s_\gamma})$ solution of the QP subproblem evaluated at ($\boldsymbol{\delta_0}$,λ_0,$\boldsymbol{\mu_0}$,$\boldsymbol{\gamma_0}$) and defined below

$$\begin{aligned}
&\min_{\boldsymbol{s}} \frac{1}{2}\boldsymbol{s}^\top H \boldsymbol{s} + \nabla f(\boldsymbol{\delta_0})^\top \mathbf{s},\\
&\text{subject to}\\
&\nabla g(\boldsymbol{\delta_0})^\top \boldsymbol{s} + g(\boldsymbol{\delta_0}) \leq 0,\\
&\nabla h_1(\delta_{0,i})^\top \boldsymbol{s} + h_1(\delta_{0,i}) \leq 0 \text{ for } i \leq N,\\
&\nabla h_2(\delta_{0,i})^\top \boldsymbol{s} + h_2(\delta_{0,i}) \leq 0 \text{ for } i \leq N,
\end{aligned} \tag{2.7.1}$$

where H is an approximation of the Hessian matrix of $\mathcal{L}$, ∇f the gradient of the objective function and ∇g, ∇h_1 and ∇h_2 the gradient of the constraint functions. The Hessian matrix H is updated at each iteration by the BFGS quazi Newton formula. The SQP method maintains the sparsity of the approximation of the Hessian matrix and its positive definetness, a necessary condition for a unique solution.

Step 3: In order to ensure the convergence of the SQP method to a global solution, the latter uses a merit function ϕ whose reduction implies progress towards a solution. Thus, a step length, denoted by $\alpha \in (0, 1)$, is chosen in order to guarantee the reduction of ϕ after each iteration such that

$$\phi(\delta_k + \alpha s_k) \leq \phi(\delta_k),$$

with

$$\phi(x) = f(x) + r\Big(g(x) + \sum_{i=1}^{N} h_1(x_i) + \sum_{i=1}^{N} h_2(x_i)\Big) \text{ and } r > \max_{1 \leq i \leq N}(|\lambda|, |\mu_i|, |\gamma_i|).$$

Step 4: The new point iterate is given by

$$(\boldsymbol{\delta_1}, \lambda_1, \boldsymbol{\mu_1}, \boldsymbol{\gamma_1}) = (\boldsymbol{\delta_0} + \alpha \boldsymbol{s_\delta}, \lambda_0 + \alpha s_\lambda, \boldsymbol{\mu_0} + \alpha \boldsymbol{s_\mu}, \boldsymbol{\gamma_0} + \alpha \boldsymbol{s_\gamma}).$$

If the latter satisfies the KKT conditions (3.6.3), the SQP converges at that point. If not, set $k = k + 1$ and go back to **Step 2**.

Remarks 2.7.1. *It should be noted that the KKT conditions defined in (3.6.3) are known as the first order optimality conditions, see e.g., Luenberger and Ye [81]. Hence if for a given vector* $(\boldsymbol{\delta}^*, \lambda^*, \boldsymbol{\mu}^*, \boldsymbol{\gamma}^*)$*, the KKT conditions are satisfied, then* $(\boldsymbol{\delta}^*, \lambda^*, \boldsymbol{\mu}^*, \boldsymbol{\gamma}^*)$ *is a local minimum of (2.4.10).*

2.8 Appendix C: MDNLP optimisation problem (2.4.13)

Step 1: Given that Ψ_i is discrete and depends on the values of δ_i, we assume that Ψ_i can be written as a function of δ_i as follows

$$\Psi_i(\delta_i) = -0.9775\delta_i^2 - 0.4287\delta_i + 0.9534 \text{ for } \delta_i \in \mathbf{D}.$$

(2.4.13) is then treated as a continuous optimisation problem and the optimal solution is found by using one of the methods described previously. We denote by $\boldsymbol{\delta}^*$ the continuous optimal solution.

Step 2: Let $\boldsymbol{\delta_0}$ be the rounded up vector of $\boldsymbol{\delta^*}$ to the nearby discrete values of the set $\mathbf{D}$. $\boldsymbol{\delta_0}$ is considered to be the initial point iterate. If $\boldsymbol{\delta_0}$ is not a feasible point of (2.4.13), then (2.4.13) is approximated by a mixed discrete linear optimisation problem at $\boldsymbol{\delta_0}$ and is given by

$$\begin{aligned} &\min_{\boldsymbol{\delta}} \nabla f(\boldsymbol{\delta_0})^\top(\boldsymbol{\delta}-\boldsymbol{\delta_0}), \\ &\text{subject to } g(\boldsymbol{\delta_0}) + \nabla g(\boldsymbol{\delta_0})^\top(\boldsymbol{\delta}-\boldsymbol{\delta_0}) \leqslant 0, \\ &\text{ and } \boldsymbol{\delta} \in \mathbf{D}^N. \end{aligned} \tag{2.8.1}$$

Step 3: (2.8.1) is solved by using a linear programming method and the branch and bound method, see Dakin [24] for more details. We denote by $\boldsymbol{\delta_k}$ the new point iterate. If $\boldsymbol{\delta_k}$ is feasible and $||\boldsymbol{\delta_k}-\boldsymbol{\delta_{k-1}}|| < \epsilon$ with $\epsilon > 0$ small, then the iteration is stopped. Else $k = k+1$ and go back to **Step 2**.

Remarks 2.8.1. *If, for a certain point iterate $\boldsymbol{\delta}$, the constraint of (2.4.13) is satisfied and $\boldsymbol{\delta} \in \mathbf{D}^N$ then $\boldsymbol{\delta}$ is a feasible solution of the optimisation problem.*
In general, it is very hard to find the global minimum of a MDNLP optimisation problem due to the fact that there are multiple local minimums. Therefore, $\boldsymbol{\delta^*}$ is said to be a global minimum if $\boldsymbol{\delta^*}$ is feasible and $f(\boldsymbol{\delta^*}) \leq f(\boldsymbol{\delta})$ for all feasible $\boldsymbol{\delta}$.

2.9 Appendix D: Prior distribution for simulation

2.9.1 Simulation based on the Uniform distribution (simulation Case 1)

The tables below describe the range of δ with their respective distribution based on the different retention levels.

Retention level (%)	85	87.5	90
Range of $\boldsymbol{\delta}$(%)	$\{15, 20\}$	$\{10, 15\}$	$\{0, 5, 10, 15\}$
Prior distribution	$U(0.85, 0.99)$	$U(0.90, 0.99)$	$U(0.04, 0.68)$

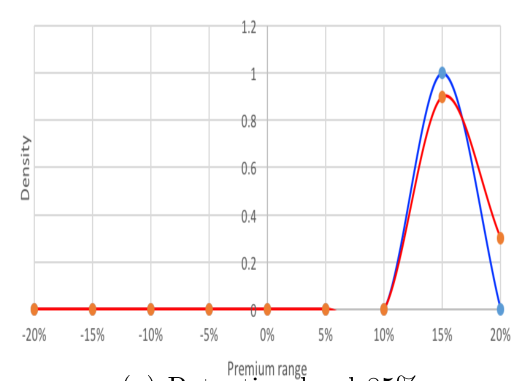
Table 2.9.1: Possible range of $\boldsymbol{\delta}$ and prior distribution uniformly distributed.

Retention level (%)	92.5	95	97.5
Range of $\boldsymbol{\delta}$(%)	$\{-5, 0, 5, 10, 15\}$	$\{-5, 0, 5, 10, 15\}$	$\{-20, -10, -5, 0, 5, 10, 15\}$
Prior distribution	$U(0.05, 0.40)$	$U(0.04, 0.21)$	$U(0.002, 0.47)$

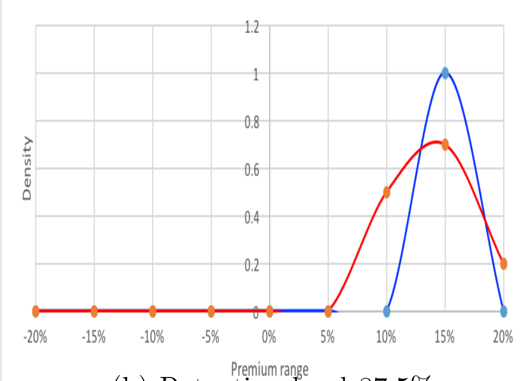
Table 2.9.2: Possible range of $\boldsymbol{\delta}$ and prior distribution uniformly distributed.

2.9.2 Simulation based on practical experience and on the optimal premium changes from the MDNLP algorithm

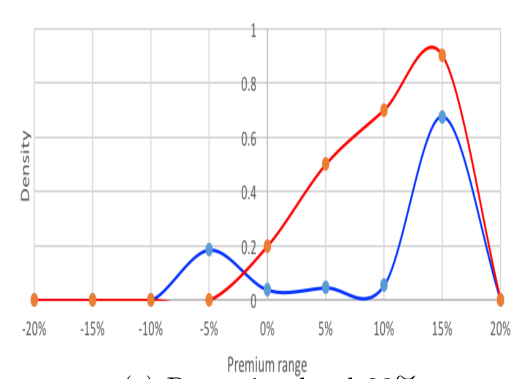
We depict in Figure 2.9.1 the prior distributions used in the simulation approach described in Section 2.5.3. The red curves represent the prior distribution from practical experience (simulation Case 2) while the blue curves are the empirical distribution of the optimal premium changes obtained with the MDNLP algorithm (simulation Case 3).



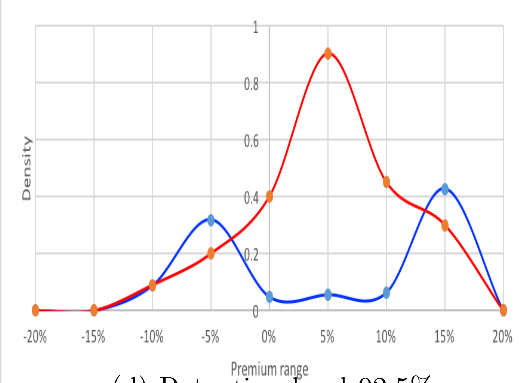
(a) Retention level 85%



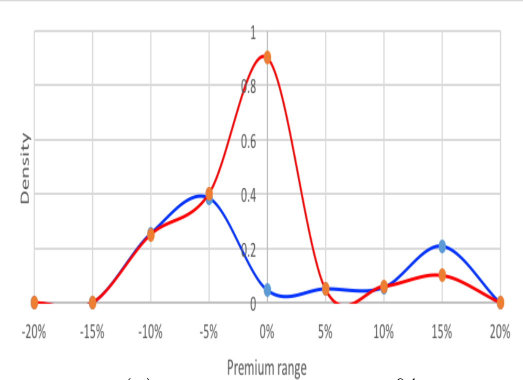
(b) Retention level 87.5%



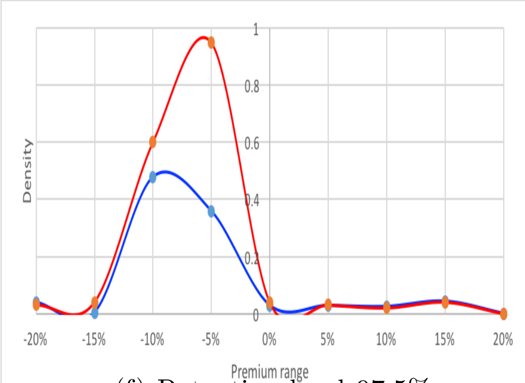
(c) Retention level 90%



(d) Retention level 92.5%



(e) Retention level 95%



(f) Retention level 97.5%

Figure 2.9.1: Prior distribution used in simulations studies: case 2 and case 3.

Chapter 3

Price Optimisation for New Business

This chapter is based on M. Tamraz and Y. Yang: Price Optimisation for New Business, *Submitted*.

3.1 Introduction

Consider that in the insurance market N customers are looking for an insurance coverage. There are $k+1$ different insurance companies that offer different premiums to each customer, say j^{th} customer receives k offers, i.e., $P_{ji}, i \leq k$ is the premium offered by the i^{th} insurance company. Of course, we shall assume that all the offers are for the same coverage. Of interest here is the possibility of a premium optimisation approach for a given company operating in the market. We write, for notation simplicity, the premiums offered by that company to the j^{th} customer as P_j instead of say P_{j1}. Let us consider a simple example. Suppose that $k = 3$, so there are altogether four companies in the market. The premiums offered by three of them are 500CHF, 520CHF, 522CHF, whereas the premium offered by the company in question that will perform an optimisation, say l, is $P_j = 519$CHF. Assume that the total profit from the contract if the premium offered is 500CHF equals 40CHF. If instead of P_j an optimal premium

$$P_j^* = P_j(1+\delta_j), \quad 1 \leq j \leq N$$

is offered, for instance $\delta_j = -0.06$ then the contract is still profitable (with approximate profit of 16 CHF) and moreover, by this offer the company is ranked first. The chances for getting this customer are therefore high. Typically, insurance companies offer also premiums that are not profitable (those risks are cross-subsidised). Therefore lowering the premium is not always the right and optimal solution. So the decision related to which δ to choose for each contract strongly depends on the strategy of the company. For the customer j let $I_j(P_j, \delta)$ denote the Bernoulli random variable which equals to 1, meaning that the customer accepts the contract for the premium $P_j(1+\delta_j)$ with acceptance probability

$$\pi_j(P_j, \delta_j) \in (0, 1].$$

Each contract offered can be seen as an independent risk. Therefore, the total number of customers that join the company as new business is given by

$$\mathcal{N}(\boldsymbol{\delta}) = \sum_{j=1}^{N} I_j(P_j, \delta_j).$$

Hereafter, in order to avoid trivialities we shall assume that $\delta_j > -1$.
Consequently, the total premium volume of the new business $\mathcal{V}(\boldsymbol{\delta})$ (which is random) is given by

$$\mathcal{V}(\boldsymbol{\delta}) = \sum_{j=1}^{N} I_j(P_j, \delta_j) P_j (1+\delta_j).$$

Of interest for the insurance company is to maximize the expected premium volume, i.e., the objective function is to maximize

$$\mathbb{E}\left\{\mathcal{V}(\boldsymbol{\delta})\right\} = \sum_{j=1}^{N} P_j(1+\delta_j)\pi(P_j, \delta_j) \tag{3.1.1}$$

under some business constraints, for instance the expected number of new customers should not be below aN, i.e.,

$$\mathbb{E}\left\{\mathcal{N}(\boldsymbol{\delta})\right\} = \sum_{j=1}^{N} \pi_j(P_j, \delta_j) \geq aN, \tag{3.1.2}$$

where $a \in (0, 1)$ is a prespecified known constant.
Price optimisation for new business has already been discussed in brief, see Marin and Bayley [84] and for renewal business see Hashorva et al. [61] for more details. However, in the literature, price optimisation is more focused on the regulations and ethical points of view, see NAIC [92], Schwartz and Harrington [108].
This paper is structured as follows. In Section 2, we define the optimisation problems from the insurer's perspective. Section 3 is dedicated to the choice of the acquisition rate π_j for each customer based on the competitors' price. Finally, Section 4 presents applications of the defined optimisation problems to a simulated data set.

3.2 Optimisation Models

Nowadays, insurers are interested in increasing their conversion rates on new business. This action leads as a result to an increase in the premium volume of the respective company. Clearly, one simple solution is to lower the premiums of all customers looking to purchase an insurance coverage in the market on one hand and increase the premiums of the existing customers at renewal on the other hand. Eventhough this method might substantially increase the conversion rate of company l and its expected premium volume, it does not represent the optimal solution as it does not differentiate between the different segments of customers in the market. Therefore, we use price optimisation in order to avoid negative profit performance and adverse selection. Moreover, given the high competition in the market, insurance companies need to constantly monitor their position to maintain their reputation. In this respect, we shall consider the competitors' premiums in the optimisation setting.

In the sequel, we assume that we have k+1 insurance companies in the market, representing k competitors for company l who will perform the optimisation. Also, we assume that we have full information about the market, so the premiums of the k competitors are known. Hereafter, we shall define two optimisation problems relevant for company l assuming that N customers are looking to purchase an insurance coverage in the market.

3.2.1 Maximise the expected premium volume

Insurers are interested in maximising their premium volume as it is one of the main source of profit of an insurance company. However, they expect to have in their portfolio a certain number of new customers and this based on the strategy of the company. Thus, the optimisation problem can be formulated as such

$$\begin{aligned} &\max_{\boldsymbol{\delta}} \sum_{j=1}^{N} P_j(1+\delta_j)\pi_j(P_j,\delta_j), \\ &\text{subject to } \frac{1}{N}\sum_{j=1}^{N} \pi_j(P_j,\delta_j) \leqslant \ell_1, \\ &\text{and } \frac{1}{N}\sum_{j=1}^{N} \pi_j(P_j,\delta_j) \geqslant \ell_2, \end{aligned} \tag{3.2.1}$$

where $\ell_1, \ell_2 < 1$ are 2 constants. For instance, ℓ_1 and ℓ_2 may denote the ratio of the expected number of customers to the total number of customers of the cheapest and the most expensive company in the market respectively. However, in practice, the total numbers of customers in a certain insurance company is not known by other companies, hence ℓ_1 and ℓ_2 are set by the insurers based on their strategy.

3.2.2 Maximise the expected number of new customers

The second objective function is concerned with the number of customers that the company is expected to get at the beginning of the period. Eventhough maximising the premium volume is important for insurers, they, nonetheless, are interested in acquiring a maximum number of new customers as they may profit the insurance company in the long run. Thus, one of the main goals of insurers is to maximise the expected number of customers that may accept the offer. Hence, the optimisation problem can be formulated as follows

$$
\begin{aligned}
&\max_{\boldsymbol{\delta}} \sum_{j=1}^{N} \pi_j(P_j, \delta_j), \\
&\text{subject to } \sum_{j=1}^{N} P_j(1+\delta_j)\pi_j(P_j, \delta_j) \leqslant C_1, \\
&\text{and } \sum_{j=1}^{N} P_j(1+\delta_j)\pi_j(P_j, \delta_j) \geqslant C_2,
\end{aligned}
\tag{3.2.2}
$$

where C_1 and C_2 are two constants. Typically, in practice, the insurer would like to maintain the reputation of the insurance company in terms of premium volume and thus stay relatively in the same position he was in the market before performing the optimisation. Therefore, C_1 and C_2 depend on the expected premium volume of the competitors. For instance, $C_1 = \sum_{j=1}^{N} P_{n,j}\pi_{n,j}(P_{n,j}, 0)$ and $C_2 = \sum_{j=1}^{N} P_{m,j}\pi_{m,j}(P_{m,j}, 0)$ where m and n denote the m^{th} and n^{th} competitors of company l in the market with $C_1 > C_2$.

Remarks 3.2.1. *(i) In the optimisation setting, we assume that the competitors do not react to the premium change of company l who performs the optimisation.*
(ii) In the insurance sector, there are multiple competitors in the market. However, we assume that customers are looking for large, nationally known insurer compared to a less expensive local known insurer. In this respect, we consider that 10 insurance companies are competing in the market.

(iii) We assume that the change in premium $\boldsymbol{\delta}$ *has an upper and lower bound and this based on the insurer strategy. For instance, the insurer doesn't want to be the cheapest in the market nor the most expensive one. This entails that, for* $j \leq N$, $\delta_j \in (m_j, M_j)$ *where* $m_j, M_j \in (-1, 1)$.

3.3 Choices for π_j

The optimisation problems are quite similar to the ones defined in Hashorva et al. [61] for the renewal business. However, the main difference lies in the choice of the probabilities π_j's for customer j. Clearly, π_j's are strongly dictated by how many companies are offering in the market, and how much is the premium difference. In the sequel, we discuss some possible tractable choices for π_j's.

3.3.1 Step function for π_j

Suppose that there are k other competitors in the market and their premiums for the j^{th} customer are known. If the current rate of company l, who performs the optimisation, is below one of the competitors' rate, then an increase in the rate level might not lead to a decrease in policies. In this respect, we model the conversion rate for customer j based on the competitors' premiums, more specifically with respect to the cheapest and highest premiums observed in the market with respective probabilities c_1 and c_2 where $c_1 > c_2$ and $c_1, c_2 \in (0, 1)$ as follows

$$\pi_j(P_j, \delta_j) = c_1 + (c_2 - c_1)\frac{P_j(1+\delta_j) - \min\limits_{1\leqslant i\leqslant k}(P_{ji})}{\max\limits_{1\leqslant i\leqslant k}(P_{ji}) - \min\limits_{1\leqslant i\leqslant k}(P_{ji})}, \quad \text{for } P_j(1+\delta_j) \in (A_j, B_j). \quad (3.3.1)$$

Clearly, π_j is a piece-wise linear, decreasing step function where the jumps are dictated by the difference in premiums between two offers. It should be noted that A_j and B_j are the jump points from one level to another. For simplicity, they are defined as the arithmetic average between two premium offers.

The corresponding shape of π_j is realistic from a practical point of view as for some premium ranges the customers' behaviour is the same relative to the different offers. The table below illustrates the latter. In this example, we consider four insurance companies and estimate the values of π_j for some premium ranges.

P_j	500	515	520	522
Range for P_j	(495,507)	(507,517)	(517,521)	(521,525)
π_j	0.75	0.50	0.40	0.30

Table 3.3.1: Values of π_j relative to customer j based on the different premium ranges.

By considering this expression for π_j, the objective functions in (3.2.1) and (3.2.2) are non linear discontinuous functions. Several methods in the literature are discussed to solve non-linear optimisation problems, see Boggs and Tolle [17], Fletcher and Powell [40], Frank and Wolfe [41]. However, these methods rely upon some assumptions on the objective function such as continuity, existence of derivatives, unimodality, etc. Therefore, in order to solve the optimisation problems at hand, we use the genetic algorithm method (abbreviated GA) described in Appendix A. GA is a widely popular method when it comes to this type of objective function. It has been explored in many areas such as optimisation, operation, engineering, evolutionary biology, machine learning, etc., see Garg [44], Mitchell [90], Reid [101], Shankar et al. [109] for more details. GA uses historical information to speculate on new search points with expected improved performance.

Discrete case for δ. Throughout the paper, we assume that δ is continuous and can take any values in the interval (m, M) where $m, M \in (0, 1)$. We shall investigate now the case where the change in premium δ can only take finite integer values from a discrete set for all customers. Let's consider the case where the competitors' premiums and company's l premium do not change. As π_j for $j \leq N$ depends constantly on the competitors' offers for the coverage in question and the position of the latter in the market, π_j varies for each customer j independently from δ_j and P_j. Thus, for illustration purposes, we compute the values of $\pi_j, j = 1, 2, 3$ for three different customers based on the different change in premium δ_i for $i = 1, \ldots, 9$.

i	1	2	3	4	5	6	7	8	9
δ_i	-20%	-15%	-10%	-5%	0%	5%	10%	15%	20%
π_1	0.750	0.750	0.750	0.656	0.536	0.416	0.300	0.300	0.300
π_2	0.750	0.750	0.723	0.613	0.504	0.394	0.300	0.300	0.300
π_3	0.750	0.728	0.618	0.508	0.398	0.300	0.300	0.300	0.300

Table 3.3.2: Values of π for 3 different customers based on δ_i for $i = 1, \ldots, 9$.

The discrete opimisation problem is solved using GA, see Appendix A for more description on the algorithm.

3.3.2 Linear function for π_j.

The simplest choice for π_j is by considering the continuous version of the step function defined above. Referring to Hashorva et al. [61], we assume that for each customer j,

$$\pi_j(P_j, \delta_j) = \alpha_j + \beta_j \delta_j,$$

where α_j and β_j are two constants to be estimated in applications. In this case, (3.2.1) is a quadratic optimisation problem subject to linear constraints, see Markowitz [88].

3.3.3 Logistic model for π_j.

The third choice is motivated by the logistic regression model where the logit function shall be used to model the conversion rate. The latter is popular in the literature for the modeling of probabilities such as the probability of renewal or lapses observed in an insurance portfolio as well as the probability of merger of non-life insurers, see Guillén et al. [55], Hashorva et al. [61], Meador et al. [89]. Its expression is given by

$$\pi_j(P_j, \delta_j) = \frac{1}{1 + c_j^{-1} e^{-T_j \delta_j}}, \tag{3.3.2}$$

where c_j and T_j are two constants to be estimated in applications. c_j includes the competition in the market and can be expressed in terms of π_j before premium change as follows

$$c_j = \frac{\pi_j(P_j,0)}{1-\pi_j(P_j,0)},$$

whereas $T_j < 0$ models the elasticity of customer j relative to the change in premium δ_j. The greater $|T_j|$, the more elastic the customer j is when purchasing an insurance policy. For this choice of π_j, the optimisation problems (3.2.1) and (3.2.2) are non-linear subject to non-linear constraints. We use the Sequential Quadratic Programming (SQP) method to solve this type of constrained optimisation problems. This method is very popular in the literature, see Boggs and Tolle [17], Nickel and Tolle [96]. It is an iterative method which solves a quadratic subproblem at each point iterate. The solution to the latter determines a step direction for the next point iterate. We refer to Appendix B for more details on the algorithm.

3.4 Insurance Applications

This section is dedicated to the application of price optimisation to insurance datasets. In this respect, we consider a simulated dataset describing the production of the motor line of business. The premiums are known and are assumed to be fair across the different segments of customers. Typically, in practice, auto-insurance markets are highly competitive. Insurers intensively compete on several factors such as price, quality of service, etc. Therefore, we consider that 10 leading insurers are competing in the market. Also, we assume that the premiums offered by the competitors are uniformly distributed around the company's l premiums who is performing the optimisation. Based on some characteristics on the insured and the type of vehicle, an offer is made by the insurance company for the coverage in question. In the sequel, we shall consider a heterogeneous portfolio consisting of $n = 1'000$ policyholders. Different premiums are offered to different segments of customers. Moreover, we note that for each customer the position of the competitors in the market change with respect to the premium charged and coverage. Therefore, we assume that for each offer the rank of the competitors and the company l in question change.
The table below presents some statistics relative to the premiums offered by company l and its competitors.

	Initial Premium	Competitors' Premiums								
	P_0	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9
Mean	1'204	1'163	1'166	1'167	1'164	1'172	1'168	1'164	1'163	1'162
Min	400	371	374	381	369	367	373	376	368	368
Q1	804	792	789	797	795	800	795	801	792	799
Q2	1'223	1'145	1'141	1'137	1'146	1'161	1'155	1'143	1'151	1'140
Q3	1'598	1'500	1'501	1'495	1'517	1'521	1'505	1'489	1'496	1'493
Max	1'999	2'248	2'190	2'233	2'252	2'233	2'235	2'176	2'217	2'217

Table 3.4.1: Premium Statistics

The above statistics rely on a simulation procedure described under the following steps:

- **Step 1:** Generate n uniform random numbers between 400 and $2'000$. These numbers account for the premiums offered by company l for a given coverage, denoted by P_0. We assume that 75% of the customers are charged a premium between 400 and $1'600$ and the rest in $(1'600, 2'000)$. This assumption is accurate in practice especially for **TPL covers and All Risks.**

- **Step 2:** Simulate n uniform random numbers u between 0.25 and 0.75. These numbers reflect the ranks of company l relative to each offer. For instance, 0 and 1 are the ranks of the cheapest and the most expensive companies competing in the market respectively whereas say 0.5 is the company ranked 5^{th} among the 10 competitors.

- **Step 3:** Based on **Step 2**, we compute the median premiums, denoted by P_{m_j}, for each offer j. We assume that the smallest and greatest premiums in the market, denoted by P_{min_j} and P_{max_j} are expressed with respect to P_{m_j} as such

  $$P_{min_j} = P_{m_j}(1+lwb) \quad \text{and} \quad P_{max_j} = P_{m_j}(1+upb) \quad \text{with} \quad lwb = -10\% \text{ and } upb = 15\%.$$

  Hence, for each offer j, P_{m_j} is computed as follows

  $$\frac{P_{m_j} - P_{0j}}{0.5 - u} = \frac{P_{max_j} - P_{min_j}}{1 - 0} \implies P_{m_j} = \frac{P_{0j}}{1 + (upb - lwb)(u - 0.5)}.$$

- **Step 4:** For $j \leq n$, we generate premiums between (P_{min_j}, P_{max_j}) with P_{min_j} and P_{max_j} as defined in **Step 3** for the remaining 7 insurance companies denoted by P_{ij} for $i = 1, \ldots, 7$.

- **Step 5:** If $P_{ij} = P_{0j}$ for $i = 1, \ldots, 7$, go back to **Step 4.**

3.4.1 π_j as defined in 3.3.1

We consider π_j for customer j to be a step function as defined in (3.3.1). Let $c_1 = 0.75$ and $c_2 = 0.3$ denote the conversion rates for the cheapest and the most expensive insurance companies offering in the market respectively. These values are accurate from the perspective of the policyholder as the latter is not only interested in paying the lowest premium offered in the market but is also interested in the reputation of the company and the quality of service. For illustration purposes, we first assume that only one customer is looking to purchase a motor insurance policy in the market. The competitors' offers are summarized below and are ranked in ascending order: 438, 457, 477, 492, 532, 596, 654, 675, 733.
Company l who will perform the optimisation offers an initial premium of $P_0 = 568$ to the corresponding customer for the coverage. The figure below shows the conversion rates based on the different premiums offered.

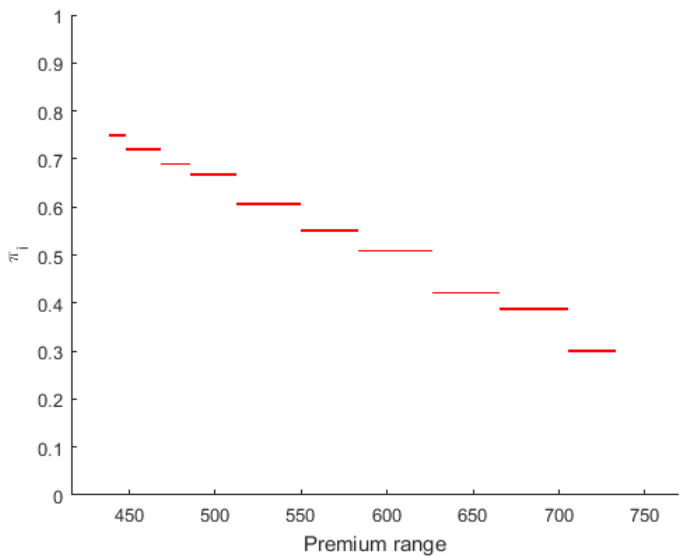


Figure 3.4.1: Values of π_j based on the premium range for customer j.

For instance, if the insurer decides to increase the premium of the current policyholder from 568 to say 680, the probability of acquiring the new business decreases from 0.55 to 0.4. Whereas, a small increase in premium, say 580, will yield the same conversion rate.

In the sequel, we assume, for simplicity, that the competitors do not react to increases/ decreases in premiums of company l. This assumption is accurate from a practical point of view as the reaction of the market to price changes is unlikely to be instantaneous due to several factors. One of the main factors is the time delay in settling claims. Indeed, the latter is important when modeling the financial state of a company.

Hereafter, we denote by t_0 the time before optimisation is performed and by t_1 the time after the optimisation. We consider the optimisation problems defined in Section 3.2.

Maximise the expected premium volume at t_1.

i) Continuous case for δ. We consider the optimisation problem defined in (3.2.1). The objective of the insurer is to maximise the premium volume of the company under some constraints on the number of customers that he expects to get at the beginning of the insurance period. In the sequel, we shall consider a conversion rate between 45% and 50%. The change in premium δ lies in $(-20\%, 20\%)$. The optimal results obtained when solving (3.2.1) using the function *ga in Matlab* are presented hereafter.

Figure 3.4.2 below highlights the positions of company l among its competitors based on the premiums offered to $n = 1'000$ customers looking to purchase an insurance coverage in the market at time t_0 and t_1.

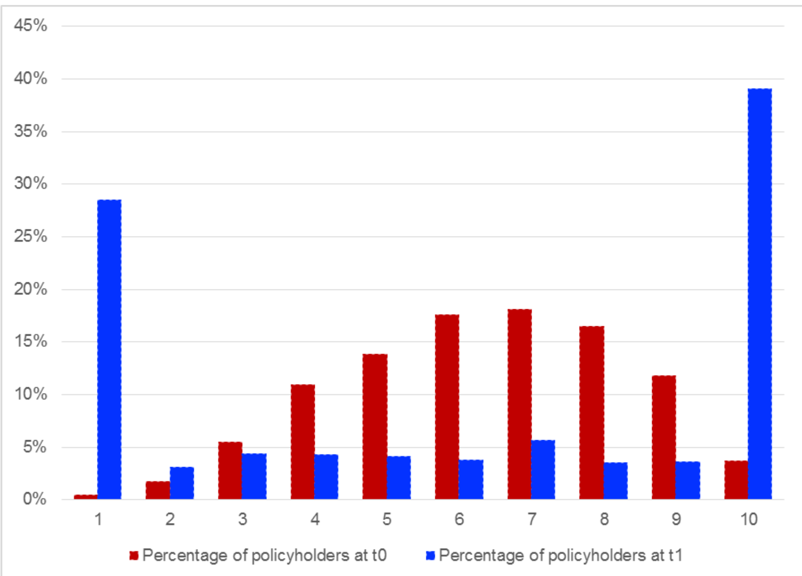


Figure 3.4.2: Premiums offered by company l compared to the competitors.

Figure 3.4.2 shows that 4% of the customers in the market are offered the highest premium by company l at t_0 compared to 39% after optimisation, at t_1. The percentage of customers that are offered the cheapest premium in the market at t_1 is of 29% compared to none at t_0. This relatively high increase will generate new sales for company l. In this particular case, the premium charged is lower than the market average premium. In practice, this decrease in premium may mainly target young new drivers. Also, if, for instance, the new customers' family, say parents or siblings, are already insured within the company, the decrease in premium is a way to enhance the loyalty of the individuals towards the company and thus increasing their future lifetime within the company. Finally, 32% are offered a premium in between at t_1 compared to 96% at t_0.

In the sequel, we shall consider two scenarios with respect to different constraints and different bounds for δ.
Scenario 1: The expected percentage of new customers (abbreviated EPN) shall be between 45% and 50%.
Scenario 2: The EPN shall be between 50% and 55%.
The table below shows the optimal results at t_1 for the two scenarios. All optimal results are normalised with the results obtained from the assumption that the insurer will not change the premiums for next year.

	Scenario 1		Scenario 2	
Bounds for δ	$(-10\%, 15\%)$	$(-20\%, 20\%)$	$(-10\%, 15\%)$	$(-20\%, 20\%)$
Aggregate expected future premium at t1 (%)	107.51	106.41	108.29	114.91
Expected number of new policies at t1 (%)	107.07	106.94	109.22	117.74
Average optimal δ (%)	1.44	0.48	0.81	-2.63
Average optimal increase (%)	13.80	13.78	13.66	13.82
Average optimal decrease (%)	-9.34	-13.37	-9.41	-13.14
Number of increases	466	510	443	390
Number of decreases	534	490	557	610

Table 3.4.2: Optimal results for Scenario 1&2 based on different bounds for δ

Table 3.4.2 shows that for both scenarios, the average optimal δ decreases with the range of possible premium changes. For instance, in Scenario 1, the average optimal δ decreases from 1.44% for $\delta \in (-10\%, 15\%)$ to 0.48% for $\delta \in (-20\%, 20\%)$. Also, the expected premium volume for Scenario 2 is greater than the one in Scenario 1 for both bounds. This is mainly due to the fact that the constraint on the expected number of new customers that will join company l is greater in Scenario

2. Typically, the probability that new customers will join the company is higher resulting in a positive effect on the expected premium volume. Also, the number of customers subject to an increase in premium is higher in Scenario 1 for both bounds.

Remarks 3.4.1. *In this application, we considered a portfolio of $n = 1'000$ customers looking to purchase an insurance policy in the market. However, in practice, n is larger. For n large, the running time may take hours (for instance, for $n = 10'000$, the running time is about four hours) whereas for $n = 1'000$ customers, using matlab, the running time is about 15 min which is reasonable. Hence, it is less time consuming if the insurance company split the offers into different categories and perform the optimisation for each of these different segments.*

ii) Discrete case for δ. In the following, we consider that the change in premium δ takes its values from a discrete set as seen in Table 3.3.2. We look at the optimisation setting (3.2.1) and set the constraints on the expected conversion rate between 45% and 50%. To solve (3.2.1), the function *ga* implemented in Matlab is used. The optimal results are summarized below.
Figure 3.4.3 highlights the distribution of the $1'000$ offers with respect to the set of discrete premium changes. For 93.5% of the portfolio, δ is between −15% and 15%. Only a small proportion of customers are offered the lowest and highest change in premium, i.e. −20% and 20%.

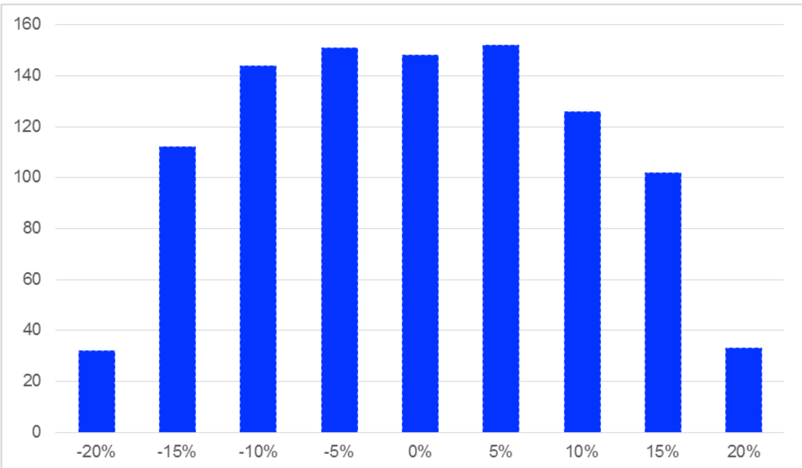


Figure 3.4.3: Optimal change in premium for the whole portfolio .

We consider two Scenarios based on the constraints of the optimisation problem.
Scenario 1: The constraints on the EPN are: $\ell_1 = 50\%$, $\ell_2 = 45\%$,
Scenario 2: The constraints on the EPN are: $\ell_1 = 55\%$, $\ell_2 = 50\%$.

	Scenario 1	Scenario 2
Aggregate expected future premium at t1 (%)	106.80	112.33
Expected number of new policies at t1 (%)	106.99	117.65
Average optimal δ (%)	-0.31	-2.85
Average optimal increase (%)	10.19	16.16
Average optimal decrease (%)	-10.28	-15.97
Number of increases	413	371
Number of decreases	587	629

Table 3.4.3: Optimal results for Scenario 1&2 based on different bounds for δ

Table 3.4.3 shows that for the same range of possible changes in premium δ, an increase in the expected conversion rate leads to a higher expected premium volume and a lower average optimal δ as seen in Scenario 1.

Maximise the expected percentage of new business at t_1

We shall now investigate the optimisation problem (3.2.2) where the insurer maximises the expected number of new customers that will join company l. Clearly, in this case, a simple approach is to offer the lowest premium in the market to attract a maximum number of new customers. However, this is not beneficial to the insurer as he would like to maintain the reputation of the insurance company in the market. In this respect, let C_1 and C_2 be two constraints relative to the expected premium volume set by the insurer for the optimisation. We assume that C_1 and C_2 depend on the expected premium volume of the competitors.
In the sequel, we shall analyse two Scenarios.
Scenario 1: We assume that the growth in the expected premium volume is between 8% and 10%, i.e, $C_1 = 595,033$ and $C_2 = 583,130$.
Scenario 2: We assume that the growth in the expected premium volume is between 10% and 16% , i.e, $C_1 = 624,410$ and $C_2 = 595,033$.

The figure below compares the number of customers observed in the different premium ranges along with the average optimal change δ in each range at time t_0 and t_1, i.e. before and after the optimisation is performed, under both Scenarios.

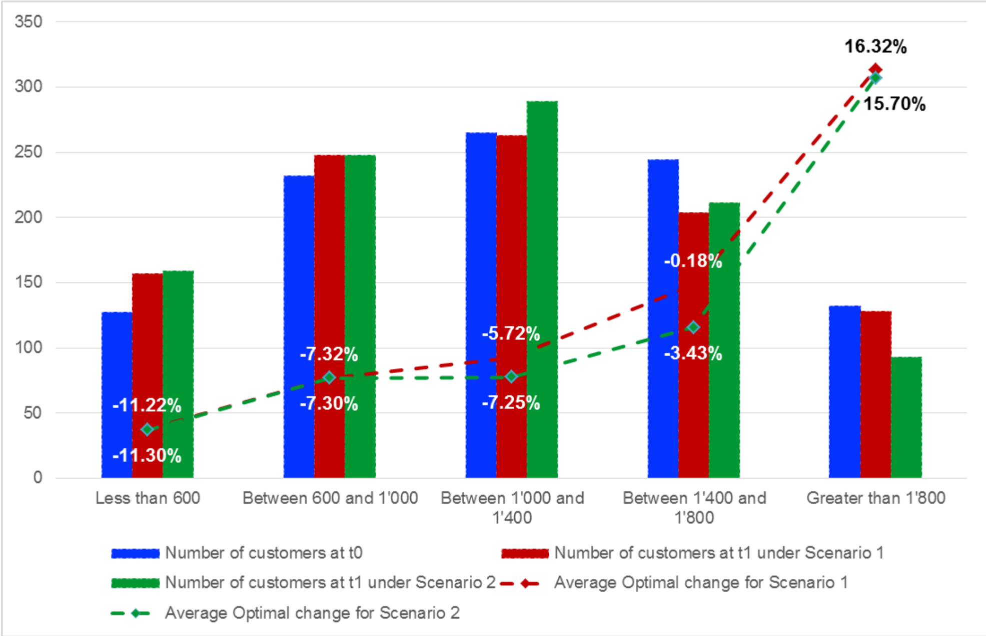


Figure 3.4.4: Number of policyholders in each premium range at t0 and at t1 for both Scenarios.

As seen in Figure 3.4.4, the number of offers with a premium less than 600 increases at t_1 whereas the number of offers with a premium above $1'800$ decreases under both Scenarios. This decrease in premium is explained by a negative average optimal change δ for almost all ranges of premiums. Also, the curve of the average optimal δ for Scenario 1 is always above the one for Scenario 2, i.e, the average optimal δ increases at a faster pace in Scenario 1 compared to Scenario 2. This is mainly explained by the constraints on the expected premium volume which is higher in Scenario 2.

	Scenario 1	Scenario 2
Bounds for δ	$(-20\%, 20\%)$	$(-20\%, 20\%)$
Aggregate expected future premium at t1 (%)	109.68	115.10
Expected number of new policies at t1 (%)	118.00	124.13
Average optimal δ (%)	-3.03	-4.97
Average optimal increase (%)	16.30	16.20
Average optimal decrease (%)	-16.46	-16.02
Number of increases	410	343
Number of decreases	590	657

Table 3.4.4: Optimal results for Scenario 1&2 based on different bounds for δ

Table 3.4.4 shows that an increase in the expected premium volume leads to an increase in the expected number of customers and a decrease in the average optimal δ. These results are accurate from the insurance company's perspective as the conversion rate increases when δ decreases leading to a higher expected premium volume.

3.4.2 π_j defined as in Section 3.3.3

In this section, we consider the logit function, as defined in (3.3.3), commonly used to model the elasticity of a customer due to price changes. In Hashorva et al. [61], the latter was used to model the probability of renewal after premium change and in Guillén et al. [55], the lapses observed in the insurance industry. We assume that the constraint on the expected percentage of customers to accept the offer of company l is of 45% and the change in premium δ lies in the interval $(-20\%, 20\%)$. Solving the optimisation problem defined in (3.2.1) yields the following optimal results. First, the figure below shows the ranks of company l among the competitors based on the premium offers to the n customers before and after performing the optimisation.

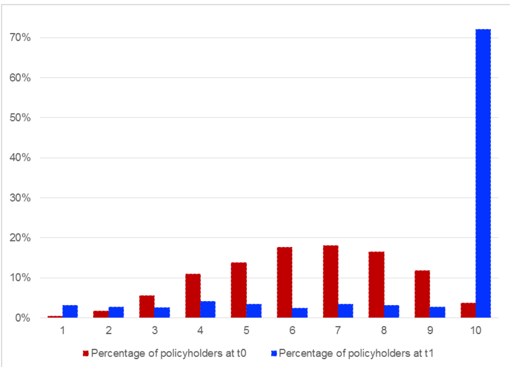


Figure 3.4.5: Acquisition rate: 45%

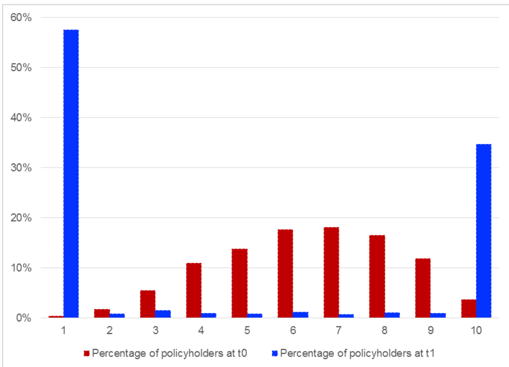


Figure 3.4.6: Acquisition rate: 50%

Figure 3.4.5 shows that a large proportion of customers are offered the highest premium in the market. This is mainly due to a relatively low acquisition rate of 45%. Clearly, this is not the case when we increase the conversion rate. As seen in Figure 3.4.6, a smaller proportion of customers are offered the highest premium in the market and a larger proportion are offered the smallest premium. This is accurate from the insurance company's perspective as the latter is more interested in acquiring new business than maximising its premium volume.

The table below summarizes the optimal results obtained at t_1 for the following two Scenarios based on different bounds for δ.

Scenario 1: The constraint on the EPN is of 45%.
Scenario 2: The constraint on the EPN is of 50%.

	Scenario 1		Scenario 2	
Bounds for δ	$(-15\%, 15\%)$	$(-20\%, 20\%)$	$(-15\%, 15\%)$	$(-20\%, 20\%)$
Aggregate expected future premium at t1 (%)	106.36	108.03	97.58	101.40
Expected number of new policies at t1 (%)	96.19	96.18	106.86	106.86
Average optimal δ (%)	10.67	12.65	-8.90	-4.82
Average optimal increase (%)	13.31	17.27	14.31	18.18
Average optimal decrease (%)	-6.15	-8.83	-14.81	-18.62
Number of increases	864	823	203	375
Number of decreases	136	177	797	625

Table 3.4.5: Optimal results for Scenario 1&2 based on different bounds for δ

As seen in Table 3.4.5, Scenario 1 yields a higher expected premium volume for both bounds compared to Scenario 2 but results in a lower acquisition rate. The average optimal change in premium is positive in Scenario 1 whereas in Scenario 2 it is negative for both bounds. This is mainly explained by the acquisition rate set in both Scenarios. Finally, the larger the range of premium change, the higher the expected premium volume and the higher the average optimal δ for both Scenarios.

3.5 Appendix A: Solution of (3.2.1) and (3.2.2) using (3.3.1).

In this section, we illustrate the GA method used to solve the optimisation problem (3.2.1) with π_j as defined in (3.3.1). The constrained optimisation (3.2.1) can be reformulated as an unconstrained one by the penalty method as follows, see Mitchell [90]

$$\min_{\boldsymbol{\delta}} f(\boldsymbol{\delta}), \text{ where } f(\boldsymbol{\delta}) = g(\boldsymbol{\delta}) + r\Big(\Phi[h_1(\boldsymbol{\delta})] + \Phi[h_2(\boldsymbol{\delta})]\Big), \tag{3.5.1}$$

where

$$\begin{cases} g(\boldsymbol{\delta}) = \sum_{j=1}^{N} P_j(1+\delta_j)\pi_j(P_j,\delta_j), \\ h_1(\boldsymbol{\delta}) = -\sum_{j=1}^{N} \pi_j(P_j,\delta_j) + \ell_1, \\ h_2(\boldsymbol{\delta}) = \sum_{j=1}^{N} \pi_j(P_j,\delta_j) - \ell_2, \end{cases}$$

and Φ and r are the penalty function and the penalty coefficient respectively.
The penalty function Φ penalize infeasible solutions (solutions that will not satisfy

the constraints of the optimisation problem) by reducing their values in the objective function, thus favoring feasible solutions in the selection process, see Yeniay [126]. In most cases, $\Phi[g(x)] = g(x)^2$ and the solution (3.5.1) tends to a feasible solution of (3.2.1) when r is large.

The algorithm is described under the following steps.
Step 1: Set a maximum number of generation n_{max}.
Step 2: Select an initial set of solution estimates chosen randomly (in contrast with the SQP method described below where one estimation point for $\boldsymbol{\delta}$ is needed). In the sequel, the set and the solution estimates are referred to as initial population and members respectively.
Step 3: The GA method relies on three operators in the following order:
1. Reproduction: This function consists in reproducing copies of the members of the initial population according to the value of their objective function. Typically, the members with the highest value of the objective function, i.e. $|g(\boldsymbol{\delta})|$, have a higher probability in contributing to the next generation. This operator can be seen as a biased roulette wheel where each member has a roulette wheel slot proportional to the value of its objective function. Therefore, to reproduce, we spin the roulette p_s times, where p_s is the size of the initial population, in order to get the new members of the next generation.
2. Crossover: It follows the reproduction one. This operator mainly produces new members of the new generation by mating and swapping each pair of strings of two members from the new generation at random with probability p_c.
3. Mutation: This operator alternates randomly the position of two or more values for the same member with probability p_m (very small in general).
Step 4: The algorithm stops when the maximum number of generation is attained. The optimal $\boldsymbol{\delta}$ is chosen based on the highest value of its objective function $|g(\boldsymbol{\delta})|$.

Remarks 3.5.1. *i) It should be noted that GA operates on a coding of the solution estimates in the form of strings often chosen to be a concatenation of binary representation.*
ii) To guarantee the success of the method, the crossover and mutations operators play an important role in finding the optimal solution as they generate new solution estimates and remove the less desired ones.
iii) The linear constraints and the upper and lower bounds are satisfied throughout the optimisation.
iv) The optimisation problem (3.2.2) is solved in a similar way.

3.6 Appendix B: Solution of (3.2.1) and (3.2.2) using (3.3.2)

(3.2.1) can be reformulated as follows:

$$\begin{aligned}
&\min_{\boldsymbol{\delta}} \ g(\boldsymbol{\delta}),\\
&\text{subject to } h_1(\boldsymbol{\delta}) \leqslant 0, \quad h_2(\boldsymbol{\delta}) \leqslant 0\\
&\text{and } f_1(\delta_j) \leqslant 0, \quad f_2(\delta_j) \leqslant 0 \quad \text{ for } j \leq N.
\end{aligned} \tag{3.6.1}$$

where g, h_1 and h_2 are defined in (3.5.1) and $f_1(\delta_j) = \delta_j - M_j$, $f_2(\delta_j) = -\delta_j + m_j$ for $j \leq N$.

Step 1: The Lagrangian function relative to (3.6.1) is given by

$$\mathcal{L} := \mathcal{L}(\boldsymbol{\delta}, \lambda, \beta, \boldsymbol{\mu}, \boldsymbol{\gamma}) = g(\boldsymbol{\delta}) + \lambda h_1(\boldsymbol{\delta}) + \beta h_2(\boldsymbol{\delta}) + \sum_{j=1}^{N} \mu_j f_1(\delta_j) + \sum_{j=1}^{N} \gamma_j f_2(\delta_j),$$

where $\lambda, \alpha \in \mathbb{R}$ and $\boldsymbol{\mu}, \boldsymbol{\gamma} \in \mathbb{R}^N$ are the Lagrangian multipliers.
Let $(\boldsymbol{\delta_0},\ \lambda_0,\ \beta_0,\ \boldsymbol{\mu_0},\ \boldsymbol{\gamma_0})$ be an estimate of the solution at $t = 0$.

Step 2: The SQP is an iterative process. Therefore, we define the next point iterate at t+1 as follows

$$\boldsymbol{\delta_{t+1}} = \boldsymbol{\delta_t} + \alpha \boldsymbol{s_t} \quad \text{ for } t \geq 0,$$

where $\alpha \in (0, 1)$ is the step length and $\boldsymbol{s_t}$ is a step vector.
$\boldsymbol{s_t} := (\boldsymbol{s_t^{\delta}}, s_t^{\lambda}, s_t^{\beta}, \boldsymbol{s_t^{\mu}}, \boldsymbol{s_t^{\gamma}})$ shall solve the following quadratic sub-problem evaluated at $(\boldsymbol{\delta_t}, \lambda_t, \beta_t, \boldsymbol{\mu_t}, \boldsymbol{\gamma_t})$ and defined as follows

$$\begin{aligned}
&\min_{\boldsymbol{s_t}} \frac{1}{2} \boldsymbol{s_t}^\top Q \boldsymbol{s_t} + \nabla g(\boldsymbol{\delta_t})^\top \mathbf{s_t},\\
&\text{subject to}\\
&\nabla h_1(\boldsymbol{\delta_t})^\top \boldsymbol{s_t} + h_1(\boldsymbol{\delta_t}) \leq 0,\\
&\nabla h_2(\boldsymbol{\delta_t})^\top \boldsymbol{s_t} + h_2(\boldsymbol{\delta_t}) \leq 0,\\
&\nabla f_1(\delta_{t,i})^\top \boldsymbol{s_t} + f_1(\delta_{t,i}) \leq 0 \text{ for } i \leq N,\\
&\nabla f_2(\delta_{t,i})^\top \boldsymbol{s_t} + f_2(\delta_{t,i}) \leq 0 \text{ for } i \leq N.
\end{aligned} \tag{3.6.2}$$

Q is an approximation of the Hessian matrix of $\mathcal{L}$ updated at each iteration by the

BFGS quazi Newton formula, ∇g the gradient of the objective function and ∇h_1, ∇h_2 , ∇f_1 and ∇f_2 the gradient of the constraint functions.
To guarantee the existence and uniqueness of a solution at each iteration, Q maintains its sparsity and positive definiteness properties.

Whereas, α is chosen as such

$$\phi(\boldsymbol{\delta_t} + \alpha \boldsymbol{s_t}) \leq \phi(\boldsymbol{\delta_t}),$$

with ϕ a merit function whose role is to ensure convergence of the SQP method to a global solution after each iteration. ϕ is given by

$$\phi(x) = g(x) + r\Big(h_1(x) + h_2(x) + \sum_{j=1}^{N} f_1(x_j) + \sum_{j=1}^{N} f_2(x_j)\Big) \text{ and } r > \max_{1\leq i\leq N}(|\lambda|, |\beta|, |\mu_i|, |\gamma_i|).$$

Step 3: If the new point iterate satisfies the KKT conditions defined in Remark 3.6.1, then it is a local minimum and the SQP converges to that point. If not, set $t = t + 1$ and go back to **Step 2**.

Remarks 3.6.1. *The success of the SQP algorithm is governed by the KKT conditions. Typically, if the point iterate doesn't satisfy these conditions, the SQP algorithm do not converge to the local/global optima. Thus, we shall define the KKT conditions as follows*

$$\begin{cases} \nabla \mathcal{L} = \mathbf{0}, \\ \lambda^* h_1(\boldsymbol{\delta^*}) = 0, \\ \beta^* h_2(\boldsymbol{\delta^*}) = 0, \\ \mu_i f_1(\delta_j^*) = 0 \textit{ for } j \leq N, \\ \gamma_j f_2(\delta_j^*) = 0 \textit{ for } j \leq N, \\ h_1(\boldsymbol{\delta^*}) \leq 0, \;\; h_2(\boldsymbol{\delta^*}) \leq 0, \\ f_1(\delta_j^*) \leq 0, \;\; f_2(\delta_j^*) \leq 0 \textit{ for } j \leq N \\ \lambda \geq 0, \quad \beta \geq 0, \quad \boldsymbol{\mu} \geq \mathbf{0}, \quad \boldsymbol{\gamma} \geq \mathbf{0}. \end{cases} \tag{3.6.3}$$

Chapter 4

On some new dependence models derived from multivariate collective models in insurance applications

This chapter is based on E. Hashorva, G. Ratovomirija and M. Tamraz: On some new dependence models derived from multivariate collective models in insurance applications, published in *Scandinavian Actuarial Journal*, 8:730-750, 2016.

4.1 Introduction

Modelling the dependence structure between insurance risks is one of the main tasks of actuaries. For instance, the determination of a risk capital in the risk management framework needed to cover unexpected losses of an insurance portfolio and the allocation of the latter to each line of business is of importance when choosing the best model of dependence for multivariate insurance risks. As discussed in Nelsen [93], copulas are a popular multivariate distribution when modelling the dependency between insurance risks as they separate the marginals from the dependence structure, see Embrechts [34], Genest et al. [46] and references therein. With motivation from Zhang and Lin [127], in this contribution we propose a flexible family of copulas derived from the joint distribution of the largest claim sizes of two insurance portfolios.

Next, in order to introduce our model, we consider the classical collective model over a fixed time period of two insurance portfolios with (X_i, Y_i) modelling the i^{th} claim size of both portfolios and N the total number of such claims. If $N = 0$, then there are no claims, so the largest claims in both portfolios are equal to 0. When $N \geq 1$, then $(X_{N:N}, Y_{N:N})$ denotes the maximal claim amounts in both portfolios. Commonly, claim sizes are assumed to be positive, however here we shall simply assume that $(X_i, Y_i), i \geq 1$ are independent with common distribution function (df) G and N is independent of everything else. Such a model is common for proportional reinsurance. In that case $Y_i = cX_i$ with c being a positive constant. Another instance is if X_i's model claim sizes and Y_i's model the expenses related to the settlement of X_i's, see Denuit et al. [27] for statistical treatments and further applications. The df of $(X_{N:N}, Y_{N:N})$ denoted by F^* is given by

$$F^*(x,y) = L_N(-\ln G(x,y)), \quad x,y \geq 0, \tag{4.1.1}$$

with L_N the Laplace transform of N. Clearly, F^* is a mixture df given by

$$F^*(x,y) = \mathbb{P}\{N=0\} + \mathbb{P}\{N \geq 1\} F(x,y), \quad x,y \geq 0,$$

where

$$F(x,y) = L_\Lambda(-\ln G(x,y)), \quad x,y \geq 0, \tag{4.1.2}$$

with $\Lambda = N|N \geq 1$ and L_Λ its Laplace transform.
Since both distributional and asymptotic properties of F^* can be easily derived from those of F, in this paper we shall focus on F assuming throughout that $\Lambda \geq 1$ is an integer-valued random variable.
When the df G is a product distribution, F above corresponds to the frailty model, see e.g., Denuit et al. [27], whereas the special case that Λ is a shifted geometric random variable is dealt with in Zhang and Lin [127]. We mention three tractable cases for Λ:
Model A: In Zhang and Lin [127], Λ is assumed to have a shifted Geometric distribution with parameter $\theta \in (0,1)$ which leads to

$$F(x,y) = \frac{\theta G(x,y)}{1-(1-\theta)G(x,y)}, \quad x,y \geq 0. \tag{4.1.3}$$

Model B: Λ has a shifted Poisson distribution with parameter $\theta > 0$, i.e., $\Lambda = 1+K$ with K being a Poisson random variable with mean $\theta > 0$, which implies

$$F(x,y) = G(x,y)e^{-\theta[1-G(x,y)]}, \quad x,y \geq 0. \tag{4.1.4}$$

Model C: Λ has a truncated Poisson distribution with $\mathbb{P}\{\Lambda = k\} = e^{-\theta}\theta^k/(k!(1 - e^{-\theta})), k \geq 1$ and thus

$$F(x, y) = \frac{e^{-\theta}}{1 - e^{-\theta}}[e^{\theta G(x,y)} - 1], \quad x, y \geq 0. \tag{4.1.5}$$

Since the distributions F and their copulas are indexed by an unknown parameter θ, the new mixture copula family has several interesting properties. In particular, it allows to model highly dependent insurance risks and therefore our model is suitable for numerous insurance applications including risk aggregation, capital allocation and reinsurance premium calculations. In this contribution we investigate first the basic distributional and extremal properties of F for general Λ. As it will be shown in Section 3, interestingly the extremal properties of F are similar to those of G. With some motivation from Zhang and Lin [127], which investigates **Model A** and its applications, in this paper, we shall discuss parameter estimation and Monte Carlo simulations for parametric families of bivariate df's induced by F. In particular, we apply our results to actuarial modelling of concrete data sets from actuarial literature. Moreover we shall consider the implications of our findings for a new real data set from a Swiss insurance company. In several cases **Model B** and **Model C** give both satisfactory fit to the data. For the case of Loss and ALAE data set we model further the stop loss and the excess of loss reinsurance premium. One of the applications of the joint distribution of the largest claims $(X_{N:N}, Y_{N:N})$ of two insurance portfolios is the analysis of the impact of their sum on the risk profile of the portfolios. Over the last decades, many contributions have been devoted on the study of the influence of the largest claims on aggregate claims, see e.g., Peng [98], Asimit and Chen [12] for an overview of existing contributions on the topic. This analysis is important when designing risk management and reinsurance strategies especially in non proportional reinsurance. Ammeter [2] is one of the first contribution which addressed the impact of the largest claim $X_{N:N}$ on the moments of the total loss of an insurance portfolio $\sum_{i=1}^{N} X_i$, see also Asimit and Chen [12] for recent results. In this paper we demonstrate by simulation the influence of the sum of the largest claims observed in two insurance portfolios $X_{N:N} + Y_{N:N}$ on the distribution of S_N. Moreover, using the covariance capital allocation principle we quantify the impact of $X_{N:N}$ and $Y_{N:N}$ on the total loss S_N. The paper is organised as follows. We discuss next some basic distributional properties of F. An investigation of the coefficient of upper tail dependence and the max-domain of attractions of F is presented in Section 3. Section 4 is dedicated to parameter estimation and Monte Carlo simulation with special focus on the cases covered by **Model A-C** above. We present three applications to concrete insurance data set in Section 5. All the proofs are relegated to Appendix.

4.2 Basic Properties of F

Let G denote the df of (X_1, Y_1) and write G_1, G_2 for its marginal df's. Suppose that G_i's are continuous and thus the copula Q of G is unique. For $\Lambda = N|N \geq 1$, we have that the marginal df's of F are

$$F_i(x) = L_\Lambda(-\ln G_i(x)), \quad i = 1, 2, x \in \mathbb{R}.$$

Hence, the generalised inverse of F_i is

$$F_i^{-1}(q) = G_i^{-1}(e^{-L_\Lambda^{-1}(q)}), \quad q \in (0, 1),$$

where G_i^{-1} is the generalised inverse of $G_i, i \leq d$. Consequently, since the continuity of G_i's implies that of F_i's, the unique copula C of F is given by

$$\begin{aligned} C(u_1, u_2) &= F(F_1^{-1}(u_1), F_2^{-1}(u_2)) \\ &= L_\Lambda\big(-\ln G(G_1^{-1}(v_1), G_2^{-1}(v_2))\big) \\ &= L_\Lambda\big(-\ln Q(v_1, v_2)\big), \qquad u_1, u_2 \in [0, 1], \end{aligned} \tag{4.2.1}$$

where we set

$$v_i = e^{-L_\Lambda^{-1}(u_i)}.$$

Remarks 4.2.1. *The df of the bivariate copula in (5.1.2) can be extended to the multivariate case. Let $X_j^{(i)}$ be the $j^t h$ claim size of the portfolio $i, i = 1, \ldots, d$ and $j = 1, \ldots, N$. Thus, the df of $(X_{N:N}^{(1)}, \ldots, X_{N:N}^{(d)})$ is given by*

$$F(z_1, \ldots, z_d) = L_\Lambda\big(-\ln G(z_1, \ldots, z_d)\big), \quad z_1, \ldots, z_d \in \mathbb{R},$$

where G is the df of $(X_1^{(1)}, \ldots, X_1^{(d)})$. Similarly to the bivariate case one may express the copula of F as follows

$$C(u_1, \ldots, u_d) = L_\Lambda\big(-\ln Q(v_1, \ldots, v_d)\big), \qquad u_1, \ldots, u_d \in [0, 1],$$

where Q is the copula of G. Without loss of generality, we present in the rest of the paper the results for the bivariate case.

Next, if G has a probability density function (pdf) g, then Q has a pdf q given by

$$q(u_1,u_2)=\frac{g(G_1^{-1}(u_1),G_2^{-1}(u_2))}{g_1(G_1^{-1}(u_1))g_2(G_2^{-1}(u_2))},\quad u_1,u_2\in[0,1],$$

with g_1,g_2 the marginal pdf's. Consequently, the pdf c of C is given by (set $t=-\ln Q(v_1,v_2)$)

$$c(u_1,u_2)=\frac{\frac{\partial v_1}{\partial u_1}\frac{\partial v_2}{\partial u_2}}{Q^2(v_1,v_2)}\left(\left(L'_\Lambda(t)+L''_\Lambda(t)\right)\frac{\partial Q(v_1,v_2)}{\partial v_1}\frac{\partial Q(v_1,v_2)}{\partial v_2}-L'_\Lambda(t)Q(v_1,v_2)q(v_1,v_2)\right), \tag{4.2.2}$$

where

$L'_\Lambda(s)=-\mathbb{E}\left\{\Lambda e^{-s\Lambda}\right\}$ and $L''_\Lambda(s)=\mathbb{E}\left\{\Lambda^2 e^{-s\Lambda}\right\}$. The explicit form of c for tractable copulas Q and Laplace transform L_Λ is useful for the pseudo-likelihood method of parameter estimation treated in Section 4. To this end, we briefly discuss the correlation order and its implication for the dependence exhibited by F. Clearly, for any x,y non-negative

$$F(x,y)\le G(x,y).$$

Consequently, in view of the correlation order, see e.g., Denuit et al. [27] we have that Kendall's tau $\tau(X_{\Lambda:\Lambda},Y_{\Lambda:\Lambda})$, Spearman's rank correlation $\rho_S(X_{\Lambda:\Lambda},Y_{\Lambda:\Lambda})$ and the correlation coefficient $\rho(X_{\Lambda:\Lambda},Y_{\Lambda:\Lambda})$ (when it is defined) are bounded by the same dependence measures calculated to (X_1,Y_1) with df G, respectively.
Moreover, if $\mathbb{E}\{\Lambda\}<\infty$, then by applying Jensen's inequality (recall $\Lambda\ge 1$ almost surely) for any x,y non-negative

$$G^a(x,y)\le G^{\mathbb{E}\{\Lambda\}}(x,y)=e^{\mathbb{E}\{\Lambda\}\ln G(x,y)}\le\mathbb{E}\left\{e^{\Lambda\ln G(x,y)}\right\}\le F(x,y), \tag{4.2.3}$$

with a the smallest integer larger than $\mathbb{E}\{\Lambda\}$. Since G^a is a df, say of (S,T), then again the correlation order implies that $\tau(X_{\Lambda:\Lambda},Y_{\Lambda:\Lambda})\ge\tau(S,T)$, and similar bounds hold for Spearman's rank correlation and the correlation coefficient. In the following we shall write also $\tau(C)$ and $\tau(Q)$ (if $a=1$) instead of $\tau(X_{\Lambda:\Lambda},Y_{\Lambda:\Lambda})$ and $\tau(S,T)$, respectively. Similarly, we denote $\rho_S(C)$ and $\rho_S(Q)$ instead of $\rho_S(X_{\Lambda:\Lambda},Y_{\Lambda:\Lambda})$ and $\rho_S(S,T)$, respectively.

4.3 Extremal Properties of F

In this section, we investigate the extremal properties of F and its copula. Assume that $\Lambda=\Lambda_n$ depends on n and write C_n instead of C. Suppose for simplicity that

$\mathbb{E}\{\Lambda_n\} = n$ and G has unit Fréchet margins. Assume additionally the following convergence in probability

$$\frac{\Lambda_n}{n} \xrightarrow{p} 1, \quad n \to \infty. \tag{4.3.1}$$

The above conditions can be easily verified in concrete examples, in particular it holds if $\Lambda_n = n$ almost surely.

In order to understand the dependence of C_n, we can calculate Kendall's tau $\tau(C_n)$ as $n \to \infty$. For instance, as shown in the simulation results in Table 5.2.1, if the copula Q of G has a coefficient of upper tail dependence $\mu_Q = 0$, then $\lim_{n\to\infty} \tau(C_n) = 0$. Note that by definition if μ_Q exists, then it is calculated by

$$\mu_Q \quad = \quad 2 - \lim_{u \downarrow 0} u^{-1}[1 - Q(1-u, 1-u)] \in [0,1]. \tag{4.3.2}$$

The following result establishes the convergence of both Kendall's tau for C_n and Spearman's rank correlation $\rho_S(C_n)$ to the corresponding measures of dependence with respect to an extreme value copula Q_A which approximates Q, i.e.,

$$\lim_{n\to\infty} \sup_{u_1,u_2\in[0,1]} \left| (Q(u_1^{1/n}, u_2^{1/n}))^n - Q_A(u_1,u_2) \right| = 0, \tag{4.3.3}$$

where

$$Q_A(u_1,u_2) = (u_1 u_2)^{A(y/(x+y))}, \quad x = \ln u_1, y = \ln u_2 \tag{4.3.4}$$

for any $(u_1,u_2) \in (0,1]^2 \setminus (1,1)$, with $A : [0,1] \to [1/2,1]$ a convex function which satisfies

$$\max(t, 1-t) \le A(t) \le 1, \quad \forall t \in [0,1]. \tag{4.3.5}$$

In the literature, see e.g. Falk et al. [38], Molchanov [91], Aulbach et al. [14, 15], Bücher and Segers [19], A is referred to as the Pickands dependence function.

Proposition 4.3.1. *If the copula Q satisfies (4.3.3) and further (4.3.1) holds, then*

$$\lim_{n\to\infty} \tau(C_n) = \tau(Q_A), \quad \lim_{n\to\infty} \rho_S(C_n) = \rho_S(Q_A). \tag{4.3.6}$$

If Q_A is different from the independence copula, and therefore $A(t) < 1$ for any

$t \in (0,1)$, then we have (see e.g., Molchanov [91])

$$\tau(Q_A) = \int_0^1 \frac{t(1-t)}{A(t)} dA'(t), \quad \rho_S(Q_A) = 12 \int_0^1 \frac{1}{(1+A(t))^2} dt - 3. \qquad (4.3.7)$$

To illustrate the results stated above, we compare by simulations the dependence properties of both C and Q. To this end, we simulate random samples from both copulas and compute the empirical dependence measures. Specifically, we generate a random sample from C in which Step 1-Step 4 in Subsection 4.4.2 are repeated $10'000$ times. Also, we simulate Λ from **Model B** and two cases of Q namely, a Gumbel copula with parameter 10 and a Clayton copula with parameter 10. Table 5.2.1 describes the simulated empirical Kendall's tau and Spearman's rho for the random samples generated from C and Q.

	Q: Gumbel copula with $\theta = 10$				Q: Clayton copula with $\theta = 10$			
$\mathbb{E}\{\Lambda\}$	$\tau(C)$	$\tau(Q)$	$\rho_S(C)$	$\rho_S(Q)$	$\tau(C)$	$\tau(Q)$	$\rho_S(C)$	$\rho_S(Q)$
10	0.9059	0.9022	0.9871	0.9862	0.3533	0.8343	0.5030	0.9588
100	0.8980	0.9002	0.9848	0.9854	0.0518	0.8348	0.0775	0.9589
1'000	0.9007	0.9004	0.9856	0.9856	0.0043	0.8334	0.0064	0.9577
10'000	0.9016	0.9018	0.9857	0.9859	0.0019	0.8324	0.0027	0.9573
100'000	0.8997	0.8996	0.9851	0.9854	-0.0104	0.8316	-0.0156	0.9569

Table 4.3.1: Empirical Kendall's Tau and Spearman's rho according to Λ.

The table above shows that for the Gumbel copula case, the level of dependence of a bivariate risk governed by C is lower or approximately equal to the one corresponding to Q when $\mathbb{E}\{\Lambda\}$ increases. For the case of Clayton copula, the bigger $\mathbb{E}\{\Lambda\}$, the weaker the dependence associated with C. In particular, for a copula Q with no upper tail dependence, Clayton copula in our example, it can be seen that when $\mathbb{E}\{\Lambda\}$ increases, C tends to the independence copula. However, when Q is an extreme value copula, Gumbel copula in our illustration, the rate of decrease in the level of dependence with respect to $\mathbb{E}\{\Lambda\}$ is small. These empirical findings are due to the correlation order demonstrated in (4.2.3). To verify the results obtained from simulations, we show that, under (4.3.7), for $\alpha = 10$, we obtain $\tau(Q_A) = 0.9$ and $\rho_S(Q_A) = 0.9855$ for the Gumbel copula which are in line with the simulation

results presented in Table 5.2.1.
It should be noted that for the Gumbel copula, the Pickands dependence function can be written as follows

$$A(t) = (t^{1/\alpha} + (1-t)^{1/\alpha})^{\alpha}, \quad t \in (0,1), \alpha \in (0,1)$$

leading to a closed form for $\tau(Q_A)$ given by

$$\tau(Q_A) = 1 - \frac{1}{\alpha}.$$

Also, it is well-known that for Clayton copula (4.3.3) holds with Q_A being the independence copula, hence for this case by (4.3.6) we have $\lim_{n\to\infty} \tau(C_n) = 0$, which confirms the findings in Table 1. This section is concerned with the extremal properties of the df F introduced in (4.1.1) in terms of G and Λ. The natural question which we want to answer here is whether the extremal properties of G and F are the same. Therefore, we shall assume that G is in the max-domain of attraction of some max-stable bivariate distribution H. Without loss of generality we shall assume that H has unit Fréchet marginal df's. Hence, our assumption is that

$$\lim_{n\to\infty} G^n(nx, ny) = H(x,y), \quad x, y \in [0, \infty). \tag{4.3.8}$$

The max-stability of H and the fact that its marginal df's are unit Fréchet imply

$$H^t(tx, ty) = H(x,y), \quad \forall x, y, t \in (0, \infty) \tag{4.3.9}$$

see e.g., Falk et al. [38]. In case Λ is a shifted geometric random variable as in **Model A**, then the above assumptions imply for any x, y non-negative (set $q := 1-\theta$)

$$\begin{aligned} n[1 - F(nx, ny)] &= n\Big[1 - \frac{\theta G(nx,ny)}{1 - qG(nx,ny)}\Big] \\ &= n\frac{1 - G(nx,ny)}{1 - qG(nx,ny)} \\ &\to -\frac{1}{\theta}\ln H(x,y), \quad n \to \infty. \end{aligned}$$

Hence

$$\lim_{n\to\infty} F^n(nx, ny) = H^{1/\theta}(x,y) \tag{4.3.10}$$

or equivalently, using (4.3.9)

$$\lim_{n\to\infty} F^n(nx/\theta, ny/\theta) = H^{1/\theta}(x/\theta, y/\theta) = H(x,y), \quad x,y \in (0,\infty) \tag{4.3.11}$$

and thus F is also in the same max-domain of attraction as G. Our result below shows that the extremal properties of G are preserved for the general case when $\mathbb{E}\{\Lambda\}$ is finite. This assumption is natural in collective models, since otherwise we cannot insure such portfolios.

Proposition 4.3.2. *If $\mathbb{E}\{\Lambda\}$ is finite then $\mu_Q = \mu_C$. Moreover, if (4.3.8) holds, then*

$$\lim_{n\to\infty} F^n(a_n x, a_n y) \quad = \quad H(x,y), \quad x,y \in (0,\infty), \tag{4.3.12}$$

where $a_n = \mathbb{E}\{\Lambda\} n$.

Remarks 4.3.3. *i) It is well-known, see e.g. Falk et al. [38] that if G is in the max-domain of attraction of H, then the coefficient of upper tail dependence μ_Q of G with copula Q exists and*

$$\mu_Q \quad = \quad 2 + \ln H(1,1) = 2 - 2A(1/2).$$

By the above proposition, F is also in the max-domain of attraction of H, and thus

$$\mu_C \quad = \quad 2 - 2A(1/2) = \mu_Q \in [0,1]. \tag{4.3.13}$$

ii) Although F and G are in the same max-domain of attraction, the above proposition shows that the normalising constant $a_n = \mathbb{E}\{\Lambda\} n$ for F is different that for G (here $a_n = n$) if $\mathbb{E}\{\Lambda\} \neq 1$.

4.4 Parameter Estimation & Monte Carlo Simulations

4.4.1 Parameter Estimation

This section focuses on the estimation of the parameters of the new copula C i.e., θ of N and α of the copula Q. Hereafter, we denote $\Theta = (\theta, \alpha)$. There are three widely used methods for the estimation of the copula parameters. The classical one is the maximum likelihood estimation (MLE). Another popular method is the inference function for margins (IFM), which is a step-wise parametric method. First, the parameters of the marginal df's are estimated and then the copula parameter Θ are obtained by maximizing the likelihood function of the copula with the marginal parameters replaced by their first-stage estimators. Typically, the success of this method depends upon finding appropriate parametric models for the marginals, see Kim et al. [70].

Finally, the pseudo-maximum likelihood (PML) method, introduced by Oakes [97] consists also of two steps. In the first step, the marginal df's are estimated non-parametrically. The copula parameters are determined in the second step by maximizing the pseudo log-likelihood function. Specifically, let $X \sim G_1$ and $Y \sim G_2$ where G_1 and G_2 are the unknown marginals df's of X and Y. For instance, if the data is not censored, a commonly used non-parametric estimator of G_1 and G_2 is their sample empirical distributions which are specified as follows

$$\widehat{G_1}(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(X_i \leq x), \quad \widehat{G_2}(y) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(Y_i \leq y). \tag{4.4.1}$$

Therefore, in order to estimate the parameter Θ, we maximize the following pseudo log-likelihood function

$$l(\Theta) = \sum_{i=1}^{n} \ln c_{\Theta}(U_{1i}, U_{2i}), \quad U_{1i} = \frac{n}{n+1}\widehat{G_1}(x_i), \quad U_{2i} = \frac{n}{n+1}\widehat{G_2}(y_i), \tag{4.4.2}$$

where c_{Θ} denotes the pdf of the copula. This rescaling is used to avoid difficulties arising from the unboundedness of the pseudo log-likelihood function in (4.4.2) as $\widehat{G_1}(x_i)$ or $\widehat{G_2}(y_i)$ tends to 1, see Genest et al. [45].

Kim et al. [70] show in a recent simulation study that the PML approach is better than the well-known IFM and MLE methods when the marginal df's are unknown, which is almost always the case in practice. Moreover, it is shown in Genest et al. [45] that the resulting estimators from the PML approach are consistent and asymptotically normally distributed.

Therefore, for our study, we shall use the PML method for the estimation of Θ which takes into account the empirical counterparts of the marginal df's to find the parameter estimators.

As described in the Introduction, we consider three types of distributions for the random variable Λ:

- **Model A:** Λ follows a shifted Geometric distribution with parameter $\theta \in (0, 1)$.

  The pdf of the Geometric copula is given by

$$c_\Theta(u_1,u_2) = W(v_1,v_2)\left(\frac{(1-(1-\theta)v_1)^2(1-(1-\theta)v_2)^2}{\theta(1-(1-\theta)Q_\alpha(v_1,v_2))^3}\right), \quad (4.4.3)$$

where $v_i = \frac{u_i}{\theta+(1-\theta)u_i}, i = 1,2$ and

$$W(v_1,v_2) = (1-(1-\theta)Q_\alpha(v_1,v_2))\left(\frac{\partial^2 Q_\alpha(v_1,v_2)}{\partial v_1 \partial v_2}\right)+2(1-\theta)\left(\frac{\partial Q_\alpha(v_1,v_2)}{\partial v_1}\frac{\partial Q_\alpha(v_1,v_2)}{\partial v_2}\right),$$

which yields the following pseudo log-likelihood function

$$\begin{aligned} l(\Theta) &= \sum_{i=1}^{n}\Big(2\ln(1-(1-\theta)v_{1i}) + 2\ln(1-(1-\theta)v_{2i}) - \ln(\theta) \\ &\quad -3\ln(1-(1-\theta)Q_\alpha(v_{1i},v_{2i})) + \ln W(v_{1i},v_{2i})\Big). \end{aligned}$$

- **Model B:** Λ follows a Shifted Poisson distribution with parameter $\theta > 0$. The pdf of the shifted Poisson copula is of the form

$$c_\Theta(u_1,u_2) = W(v_1,v_2)\left(\frac{e^{\theta(Q_\alpha(v_1,v_2)+1-v_1-v_2)}}{(1+\theta v_1)(1+\theta v_2)}\right), \quad (4.4.4)$$

where $v_j = f^{-1}(u_j)$ with $f(x) = x\exp(\theta(x-1))$ and

$$W(v_1,v_2) = (1+\theta Q_\alpha(v_1,v_2))\left(\frac{\partial^2 Q_\alpha(v_1,v_2)}{\partial v_1 \partial v_2}\right)+\theta(2+\theta Q_\alpha(v_1,v_2))\left(\frac{\partial Q_\alpha(v_1,v_2)}{\partial v_1}\frac{\partial Q_\alpha(v_1,v_2)}{\partial v_2}\right).$$

The corresponding pseudo log-likelihood of the above copula is thus given by

$$l(\Theta) = \sum_{i=1}^{n}\Big(\theta(Q_\alpha(v_{1i},v_{2i})+1-v_{1i}-v_{2i}) - \ln(1+\theta v_{1i}) - \ln(1+\theta v_{2i}) + \ln W(v_{1i},v_{2i})\Big).$$

- **Model C:** Λ follows a Truncated Poisson distribution with parameter $\theta > 0$. The joint density of the truncated Poisson copula is given by

$$c_\Theta(u_1,u_2) = \frac{1}{\theta}(1-e^{-\theta})W(v_1,v_2)e^{\theta(1-v_1-v_2+Q_\alpha(v_1,v_2))}, \quad (4.4.5)$$

where

$$W(v_1, v_2) = \theta \frac{\partial Q_\alpha(v_1, v_2)}{\partial v_1} \frac{\partial Q_\alpha(v_1, v_2)}{\partial v_2} + \frac{\partial^2 Q_\alpha(v_1, v_2)}{\partial v_1 \partial v_2},$$

$$v_j = \frac{1}{\theta} \ln\left(1 + \frac{u_j(1 - e^{-\theta})}{e^{-\theta}}\right), \quad j = 1, 2.$$

The resulting pseudo log-likelihood of the above copula can be written as follows

$$l(\Theta) = \sum_{i=1}^{n} \Big(\ln\Big(\frac{1 - e^{-\theta}}{\theta}\Big) + \theta(1 - v_{1i} - v_{2i}) + \theta Q_\alpha(v_{1i}, v_{2i}) + \ln W(v_{1i}, v_{2i}) \Big).$$

Remarks 4.4.1. *The copula C_θ of Model A and Model B include the corresponding original copula Q. In particular, if $\theta = 1$ the pdf c_θ in (4.4.3) becomes the pdf of the original copula Q, see e.g. Zhang and Lin [127], while the copula C_θ of Model B reduces to the original copula Q when $\theta = 0$.*

Next, we generate random samples from the proposed copula models C.

4.4.2 Monte Carlo Simulations

Based on the distributional properties of F derived in Section 2, we have the following pseudo-algorithm for the simulation procedure which depends on the choice of Λ and Q:

- **Step 1:** Generate a value λ from Λ.
- **Step 2:** Generate λ random samples $(U_{1,i}, U_{2,i})$, $i = 1, \ldots, \lambda$, from the original copula Q.
- **Step 3:** Calculate (M_1, M_2) as follows

$$M_j = \max_{i=1,\ldots,\lambda} U_{j,i}, \quad j = 1, 2.$$

- **Step 4:** Return (V_1, V_2), such that

$$V_j = L_\Lambda(-\ln M_j), \quad j = 1, 2.$$

Simulation results are important for exploring the dependence of F. The simulation results in the table below complete those presented already in Table 5.2.1. In this regard, we generate random samples from the Joe copula with parameter $\alpha = 10$.

	Q: Joe copula with $\alpha = 10$			
$\mathbb{E}\{\Lambda\}$	$\tau(C)$	$\tau(Q)$	$\rho_S(C)$	$\rho_S(Q)$
10	0.8982	0.8194	0.9849	0.9504
100	0.9005	0.8190	0.9857	0.9509
1'000	0.8997	0.8164	0.9855	0.9492
10'000	0.9004	0.8209	0.9857	0.9520
100'000	0.8999	0.8206	0.9852	0.9513

Table 4.4.1: Empirical Kendall's Tau and Spearman's rho according to $\mathbb{E}\{\Lambda\}$.

For the Joe copula, the Pickands dependence function can be written as follows

$$A(t) = 1 - ((\psi_1(1-t))^{-\alpha} + (\psi_2 t)^{-\alpha})^{-\frac{1}{\alpha}}$$

where $\psi_1, \psi_2 \leq 1$, $t \in (0,1)$ and $\alpha \in (0,1)$.
By using (4.3.7) and for $\alpha = 10$ and $\psi_1 = \psi_2 = 1$, we obtain $\tau(Q_A) = 0.9066$ and $\rho_S(Q_A) = 0.9874$ which are in line with the simulation results observed in Table 4.4.1 for $\tau(C)$ and $\rho_S(C)$ as $\mathbb{E}\{\Lambda\}$ increases.
Another benefit of our simulation algorithm is that we can assess the accuracy of our estimation method proposed above. Therefore, we simulate random samples of size n from the copula C with different distributions for Λ: **Model A**, **Model B** and **Model C** and two types of copula for Q: the Gumbel copula and the Joe copula. Hereof, the parameters θ of Λ and α of Q are estimated from the dataset described in Subsection 4.5.1 and are presented in Table 4.4.2 .

	Q: Joe copula		Q: Gumbel copula	
Model for Λ	θ	α	θ	α
Model A	0.3254	2.3727	0.7630	2.2758
Model B	0.9537	2.6634	0.1490	2.3276
Model C	1.8660	2.5885	0.3133	2.3240

Table 4.4.2: Parameters used for sampling from C.

n	Model A $\hat{\theta}$	Diff.	$\hat{\alpha}$	Diff.	Model B $\hat{\theta}$	Diff.	$\hat{\alpha}$	Diff.	Model C $\hat{\theta}$	Diff.	$\hat{\alpha}$	Diff.
100	0.2461	-24%	2.2255	-6%	1.0765	13%	2.6597	0%	1.7400	-7%	2.2535	-13%
1'000	0.3353	3%	2.3262	-2%	0.9906	4%	2.6999	1%	1.9238	3%	2.6491	2%
10'000	0.3304	2%	2.3260	-2%	0.9795	3%	2.6651	0%	1.8996	2%	2.5999	0%
100'000	0.3285	1%	2.3462	-1%	0.9541	0%	2.6600	0%	1.8721	0%	2.5877	0%

Table 4.4.3: Parameters used for sampling from C where Q is the Joe Copula.

n	Model A $\hat{\theta}$	Diff.	$\hat{\alpha}$	Diff.	Model B $\hat{\theta}$	Diff.	$\hat{\alpha}$	Diff.	Model C $\hat{\theta}$	Diff.	$\hat{\alpha}$	Diff.
100	0.9565	25%	2.3595	4%	0.1712	15%	2.4308	4%	0.3164	1%	2.3675	2%
1'000	0.7376	-3%	2.3076	1%	0.1563	5%	2.3458	1%	0.3084	-2%	2.3126	0%
10'000	0.7660	0%	2.3083	1%	0.1545	4%	2.3476	1%	0.3136	0%	2.3185	0%
100'000	0.7596	0%	2.2639	-1%	0.1506	1%	2.3232	0%	0.3279	5%	2.3063	-1%

Table 4.4.4: Parameters used for sampling from C where Q is the Gumbel Copula.

It can be seen from Table 4.4.3 and Table 4.4.4 that the estimated parameters from the simulated samples tend to the true value of the parameters as the sample size n increases, thus indicating the accuracy of our proposed models.

4.4.3 Influence of $X_{N:N} + Y_{N:N}$ on Total Loss

In this subsection, we focus on the distribution of the aggregate claim of two insurance portfolios by excluding the largest claim of each portfolio. Specifically, we analyse the aggregate influence of $M_N := X_{N:N} + Y_{N:N}$ on some risk measures of the total loss $S_N = \sum_{i=1}^{N}(X_i + Y_i)$. Moreover, by considering the joint distribution of $(X_{N:N}, Y_{N:N})$ we quantify the individual impact of $X_{N:N}$ and $Y_{N:N}$ on the distribution of S_N. Let S_N^* be the aggregate claim excluding the largest claims, based on some risk measure $\rho(.)$ and suppose that X_i, Y_i's have a finite second moment, the influence of the largest claims on the aggregate claim is evaluated as follows

$$I^* = \rho(S_N) - \rho(S_N^*).$$

By the covariance capital allocation principle, the contribution of $X_{N:N}$ on the change of the distribution of S_N is given by

$$I(X_{N:N}, M_N) = \frac{cov(X_{N:N}, M_N)}{var(M_N)} I^*.$$

To illustrate our results we have implemented the following simulation pseudo-algorithm:

- **Step 1:** Generate the number of claims N from Λ.
- **Step 2:** Generate N random samples $(u_{1,i}, u_{2,i}), i = 1, \dots, N$, from the original copula Q.
- **Step 3:** For each portfolio, simulate N claim sizes by using the inverse method as follows

  $$X_i^b = F_1^{-1}(u_{1,i}), \quad Y_i^b = F_2^{-1}(u_{2,i}), i = 1, \dots, N,$$

  where $F_i, i = 1, 2$, is the df of X and Y, respectively.
- **Step 4:** Evaluate the total loss with and without the largest claims, respectively

  $$S_N^b = \sum_{i=1}^{N} (X_i^b + Y_i^b), \quad S_N^{*b}.$$

To obtain the simulated distribution of S_N and S_N^* **Step 1-4** are repeated B times. The results presented in Table 4.4.5 is in millions and is obtained from the following assumptions:

- number of simulations $B = 100'000$,
- the original copula is a Gumbel copula with dependence parameter $\alpha = 2.324$,
- the number of claims follows the Shifted Poisson (**Model B**) with parameter $\theta = 1'000$,
- the claim sizes are Pareto distributed as follows

$$X_i \sim Pareto(10'000, 2.2), \quad Y_i \sim Pareto(50'000, 2.5).$$

Risk measures (ρ)	S_N	S_N^*	I^*	I^* (in %)	$I(X_{N:N}, M_N)$	$I(Y_{N:N}, M_N)$
Mean	101.77	100.21	1.57	1.54	0.38	1.19
Standard deviation	4.41	3.91	0.50	11.24	0.15	0.35
VaR (99 %)	112.75	109.65	3.10	2.74	0.90	2.20
TVaR (99 %)	117.08	111.03	6.05	5.17	1.75	4.30

Table 4.4.5: Influence of the largest claims on the total loss.

It can be seen that a significant proportion of the aggregate claims is consumed by $X_{N:N} + Y_{N:N}$. For instance, based on the standard deviation as risk measure, 11.24% of the total loss is driven by the largest claims. In this regards, $X_{N:N}$ has more important contribution to I^* than $Y_{N:N}$. This result is helpful for the insurance company when choosing the appropriate reinsurance treaty in the sense that the main source of volatility of the correlated portfolios is quantified.

4.5 Real insurance data applications

In this section, we illustrate the applications of the new copula families in the modelling of three real insurance data. Specifically, we shall consider four copula families for Q_α: Gumbel, Frank, Student and Joe and three mixture copulas in which Λ with parameter θ follows one of the three distributions: Shifted Geometric, Shifted Poisson and Truncated Poisson. The AIC criteria is used to assess the quality of each model fit relative to each of the other models.

4.5.1 Loss ALAE from accident insurance

We shall model real insurance data from a large insurance company operating in Switzerland. The dataset consists of 33'258 accident insurance losses and their corresponding allocated loss adjustment expenses (ALAE) which includes mainly the cost of medical consultancy and legal fees. The observation period encompasses the claims occurring during the accident period 1986-2014[1].
Let X_i be the i^{th} loss observed and Y_i its corresponding ALAE.
Some statistics on the data are summarised in Table 4.5.1.

[1] Data set can be downloaded here http://dx.doi.org/10.13140/RG.2.1.1830.2481

	Loss	ALAE
Min	10	1
Q1	13'637	263
Q2	32'477	563
Q3	95'880	1'509
Max	133'578'900	2'733'282
No. Obs.	33'258	33'258
Mean	292'715	5'990
Std. Dev.	2'188'622	42'186

Table 4.5.1: Statistics for Loss ALAE data from accident line.

The scatterplot of (ALAE, loss) on a log scale is depicted in Figure 4.5.1. It can be seen that large values of loss are likely to be associated with large values of ALAE. In addition, the empirical estimator of some dependence measures in Table 4.5.2 suggests a positive dependence between X_i and Y_i. For instance, the empirical estimator of the upper tail dependence of 0.6869 indicates that there is a strong dependence in the tail of the distribution of X_i and Y_i.

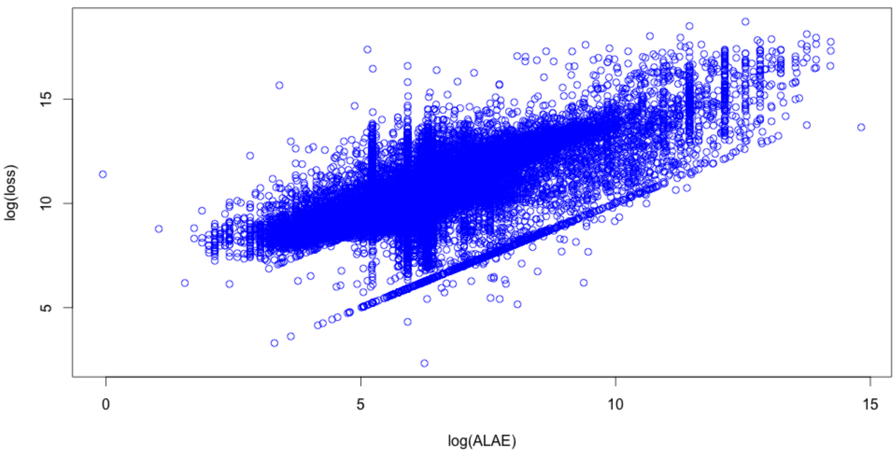


Figure 4.5.1: Scatterplot for log ALAE and log Loss: accident insurance data.

Dependence measures	Values
Pearson's Correlation	0.7460
Spearman's Rho	0.7465
Kendall's Tau	0.6012
Upper tail dependence	0.6869

Table 4.5.2: Empirical dependence measures for Loss ALAE data from accident line.

Referring to the marginals' estimator in (4.4.1), the estimation results for each copula model are found by maximizing (4.4.2) and are summarized in Table 4.5.3 below.

Model	θ	α	m	AIC
Gumbel	-	2.3876	-	-32'073
Gumbel Geometric	0.7630	2.2758	-	-32'128
Gumbel Truncated Poisson	0.3133	2.3240	-	-32'104
Gumbel Shifted Poisson	0.1490	2.3276	-	-32'059
Frank	-	8.0774	- -30'137	
Frank Geometric	0.9999	8.0772	-	-30'134
Frank Truncated Poisson	0.0001	8.0773	-	-30'135
Frank Shifted Poisson	0.0001	8.0773	-	-30'135
Student	-	0.8142	1.9805	-32'909
Student Geometric	0.1137	0.5492	1.9992	-38'088
Student Truncated Poisson	0.0001	0.7841	9.6744	-28'672
Student Shifted Poisson	0.0001	0.7885	8.7113	-29'042
Joe	-	3.0967	-	-30'655
Joe Geometric	0.3254	2.3727	-	-33'015
Joe Truncated Poisson	1.8660	2.5885	-	-32'578
Joe Shifted Poisson	0.9537	2.6634	-	-32'411

Table 4.5.3: Copula families parameters estimates.

It can be seen that the model which best fits the data is the Student Geometric copula followed by the Joe Geometric copula. We note in passing that the Student copula Q_α has an additional parameter m which is the degree of freedom.

4.5.2 Loss ALAE from general liability insurance

This data set describes the general liability claims associated with their ALAE retrieved from the Insurance Services Office available in the R package. In this respect, the sample consists of 1'466 uncensored data points and 34 censored observations. We refer to Denuit et al. [28] for more details on the description of the data. Let X_i be the i^{th} loss observed and Y_i the ALAE associated to the settlement of X_i. Each loss is associated with a maximum insured claim amount (policy limit) M. Thus, the loss variable X_i is censored when it exceeds the policy limit M. We define the censored indicator of the loss variable by

$$\delta_i = \begin{cases} 1 & \text{if} \quad X_i \leqslant M, \\ 0 & \text{if} \quad X_i > M, i = 1, \ldots, 1'500. \end{cases}$$

Next, we shall use the Kaplan-Meir estimator $\hat{G}_X$ to estimate G_1 and the empirical distribution $\hat{G}_Y$ for G_2 as in (4.4.1). In particular, the corresponding pseudo log-likelihood function is given by

$$l(\Theta) = \sum_{i=1}^{n} \Big(\delta_i \ln(c_\Theta(U_{1i}, U_{2i}) + (1 - \delta_i) \ln\Big(1 - \frac{C_\Theta(U_{1i}, U_{2i})}{\partial U_{2i}}\Big)\Big), \qquad (4.5.1)$$

where $U_{1i} = \frac{n}{n+1}\hat{G}_X(x_i)$ and $U_{2i} = \frac{n}{n+1}\hat{G}_Y(y_i)$ for $i = 1, \ldots, n$, see Denuit et al. Denuit et al. [28]. By maximizing (4.5.1), the resulting estimators of Θ for the considered copula models are presented in Table 5.4.10.

Model	θ	α	m	AIC
Gumbel	-	1.4284	-	-210.18
Gumbel Geometric	0.5425	1.3127	-	-278.23
Gumbel Truncated Poisson	0.0001	1.4422	-	-360.49
Gumbel Shifted Poisson	0.1410	1.4083	-	-361.20
Frank	-	3.0440	-	-321.44
Frank Geometric	0.7800	2.7464	-	-174.40
Frank Truncated Poisson	0.0001	3.0375	-	-306.40
Frank Shifted Poisson	0.0001	3.0375	-	-306.41
Student	-	0.4642	10.0006	-180.99
Student Geometric	0.7095	0.4252	9.1897	-228.82
Student Truncated Poisson	1	0.4094	13.9922	-271.40
Student Shifted Poisson	1	0.4016	13.9983	-295.42
Joe	-	1.6183	-	-179.00
Joe Geometric	0.4379	1.3864	-	-292.41
Joe Truncated Poisson	0.0607	1.6356	-	-331.21
Joe Shifted Poisson	0.8075	1.4629	-	-361.76

Table 4.5.4: Copula families parameters estimates.

Since the Joe Shifted Poisson copula has the the smallest AIC it represents the best model for describing the dependence in the dataset followed by the Gumbel Shifted Poisson copula.

4.5.3 Danish fire insurance data

The corresponding data set describes the Danish fire insurance claims collected from the Copenhagen Reinsurance Company for the period 1980-1990. It can be retrieved from the following website: *www.ma.hw.ac.uk/ ∼ mcneil/*. This data set has first been considered byEmbrechts et al. [35] (Example 6.2.9) and explored by Haug et al. [62]. It consists of three components: loss to buildings, loss to contents and loss to profit. However, in this case, we model the dependence between the first two components. The total number of observations is of 1’501. We only consider the observations where both components are non-null. As indicated by the empirical dependence measures in Table 4.5.5, the level of dependence between these two losses is low.

Dependence measures	Values
Pearson's Correlation	0.1413
Spearman's Rho	0.1417
Kendall's Tau	0.0856
Upper tail dependence	0.1998

Table 4.5.5: Dependence measures for the Danish fire insurance.

The estimation results for each copula is summarized in Table 5.4.8 below.

Model	θ	α	m	AIC
Gumbel	-	1.1762	-	-133.18
Gumbel Geometric	0.9999	1.1762	-	-131.17
Gumbel Truncated Poisson	0.0001	1.1762	-	-131.18
Gumbel Shifted Poisson	0.0001	1.1762	-	-131.17
Frank	-	0.8807	-	-29.12
Frank Geometric	0.9999	0.8804	-	-27.12
Frank Truncated Poisson	0.0001	0.8806	-	-27.12
Frank Shifted Poisson	0.0001	0.8805	-	-27.12
Student	-	0.1574	9.5998	-47.86
Student Geometric	0.9999	0.1576	10.0063	-45.84
Student Truncated Poisson	0.0001	0.1570	9.0048	-45.81
Student Shifted Poisson	0.0001	0.1562	8.9833	-45.42
Joe	-	1.3585	-	-204.85
Joe Geometric	0.9999	1.3585	-	-202.83
Joe Truncated Poisson	0.0001	1.3585	-	-202.84
Joe Shifted Poisson	0.0001	1.3585	-	-202.83

Table 4.5.6: Copula families parameters estimates.

It can be seen that the model that best fits the data is the Joe copula followed by the Joe Truncated Poisson copula. The Frank mixture copulas and Student mixture copulas are not a good fit for the data as their AIC is higher by far compared to the Gumbel and Joe mixture copulas families.

4.5.4 Reinsurance premiums

In this section, we examine the effects of the dependence structure on reinsurance premiums by using the proposed copula models. In practice, it is well known that insurance risks dependency has an impact on reinsurance. For instance, Dhaene and Goovaerts [31] have shown that stop loss premium is greater under the dependence assumption than under the independence case. In what follows, we consider the insurance claims data described in Subsection 4.5.1 where we denote X the loss variable, Y the associated ALAE and K the number of claims for the next accident year. In addition, two types of reinsurance treaties are analyzed namely:

- Excess-of-loss reinsurance, where the claims from Y_i 's are attributed proportionally to the insurer and the reinsurer. For a given observation (X_i, Y_i) the payment for the reinsurer is described as follows, see Cebrian et al. [20]

$$g(X_i, Y_i, r) = \begin{cases} 0 & \text{if} \quad X_i \leqslant r, \\ X_i - r + \left(\frac{X_i - r}{X_i}\right) Y_i & \text{if} \quad X_i > r \end{cases}$$

  leading to a reinsurance premium of the form

$$\kappa(r) = \mathbb{E}\left\{K\right\} \mathbb{E}\left\{g(X_i, Y_i, r)\right\}, \tag{4.5.2}$$

  where $r > 0$ is the retention level.

- Stop loss reinsurance, where the premium is given by

$$\pi(d) = \mathbb{E}\left\{\left(\sum_{i=1}^{K}(X_i + Y_i) - d\right)_+\right\} \tag{4.5.3}$$

  and d is a positive deductible.

In order to calculate the reinsurance premiums defined above, Monte Carlo simulations have been implemented. Hereof, we assume that K is Poisson distributed with a mean of 156.2, representing the expected number of claims estimated by the insurance company. Additionally, we use the empirical distributions of X_i and Y_i for the simulation of the claims amount. Regarding the dependence model, the following copulas are considered: independent copula, Joe copula, Geometric Joe Copula, Truncated Poisson Joe copula and the Shifted Poisson Joe copula where the parameters are summarized in Table 4.5.3. The following steps summarize the implemented pseudo-algorithm:

- **Step 1:** Generate the number of claims $K \sim Poisson(156.2)$.
- **Step 2:** Simulate $(u_i, v_i), i = 1, \ldots, K$ from the considered copula C.
- **Step 3:** Generate the loss and ALAE claims as follows

$$(x_i = \hat{F}^{-1}_{X,n}(u_i), y_i = \hat{F}^{-1}_{Y,n}(v_i)), i = 1, \ldots, K,$$

where $\hat{F}^{-1}_{X,n}$ and $\hat{F}^{-1}_{Y,n}$ are the inverse of the empirical df of X and Y respectively, with

$$\hat{F}_{X,n}(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(X_i \leq x), \quad \hat{F}_{Y,n}(y) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(Y_i \leq y).$$

- **Step 4:** Calculate the reinsurance premiums $\kappa^b(r)$ and $\pi^b(d)$ as in (4.5.2) and (4.5.3) respectively.
- **Step 5:** **Step 1** -**Step 4** are repeated B times and the estimators of the reinsurance premiums are given by

$$\hat{\kappa}(r) = \frac{1}{B}\sum_{b=1}^{B} \kappa^b(r), \quad \hat{\pi}(d) = \frac{1}{B}\sum_{b=1}^{B} \pi^b(d).$$

The estimation results presented in Table 4.5.7 are obtained from repeating **Step 1** -**Step 4** 100'000 times. These amounts are expressed in CHF million.

	$\hat{\kappa}(r)$			$\hat{\pi}(d)$		
Copula model	$r = 1$	$r = 5$	$r = 10$	$d = 10$	$d = 20$	$d = 30$
Independent	13.1137	6.5692	3.0971	14.7530	7.5145	3.5738
Joe	13.6950	6.7776	3.2396	15.1056	7.7691	3.8233
Joe Geometric	13.4483	6.7365	3.1619	14.8975	7.6797	3.7177
Joe Truncated Poisson	13.4038	6.7183	3.0929	14.8016	7.6698	3.6493
Joe Shifted Poisson	13.4776	6.6789	3.1081	14.9250	7.6266	3.6702

Table 4.5.7: Reinsurance premiums with respect to copula models.

Table 13 shows that the reinsurance premiums $\hat{\kappa}(r)$ and $\hat{\pi}(d)$ are lower under the independence hypothesis. Hence, the portfolio is less risky when the loss variable X_i and the ALAE variable Y_i are assumed to be independent. Furthermore, when the retention limit r increases for the excess of loss treaty, the reinsurance premiums estimates $\hat{\kappa}(r)$ under the Joe mixture copula models tend to the estimated values under the independence assumption. Conversely, for the stop loss treaty, the higher the deductible d the higher the deviation from the independence hypothesis. Furthermore, by comparing the results for each copula model, it can be seen that the Joe copula generates the highest reinsurance premiums. This result is expected given that the strongest dependence structure is obtained under the Joe copula. On the other hand, the weakest dependence model for this data is observed under the Joe truncated Poisson copula as the reisurance premiums $\hat{\kappa}(r)$ and $\hat{\pi}(d)$ are the smallest for different values of r and d.

4.6 Appendix

4.6.1 Proofs

Derivation of (4.4.4)-(4.4.5): We show first (5.1.3). The corresponding joint density c of the df C is given by

$$c(u_1,u_2)=\frac{\partial C(u_1,u_2)}{\partial u_1 \partial u_2}=\frac{\partial L_\Lambda(-\ln Q_\alpha(v_1,v_2))}{\partial u_1\partial u_2}, \tag{4.6.1}$$

where

$$C(v_1,v_2)=L_\Lambda(-\ln Q_\alpha(v_1,v_2)),\quad v_i=e^{-L_\Lambda^{-1}(u_i)},\quad i=1,2.$$

In view of (4.6.1), the partial derivative of C with respect to u_1 is

$$\frac{\partial L_\Lambda(-\ln Q_\alpha(v_1,v_2))}{\partial u_1}=\frac{1}{Q_\alpha(v_1,v_2)}L'_\Lambda(-\ln Q_\alpha(v_1,v_2))\frac{-\partial Q_\alpha(v_1,v_2)}{\partial v_1}\frac{\partial v_1}{\partial u_1}$$

leading to

$$\begin{aligned}
c(u_1,u_2) &= \frac{\partial}{\partial v_2}\left(\frac{\partial L_\Lambda(-\ln Q_\alpha(v_1,v_2))}{\partial u_1}\right)\frac{\partial v_1}{\partial u_1}\frac{\partial v_2}{\partial u_2}\\
&= \frac{\partial v_2}{\partial u_2}\left(L''_\Lambda(-\ln Q_\alpha(v_1,v_2))\frac{\frac{\partial Q_\alpha(v_1,v_2)}{\partial v_1}\frac{\partial Q_\alpha(v_1,v_2)}{\partial v_2}}{Q^2_\alpha(v_1,v_2)}\right.\\
&\quad \left.+L'_\Lambda(-\ln Q_\alpha(v_1,v_2))\frac{\frac{-\partial^2 Q_\alpha(v_1,v_2)}{\partial v_1\partial v_2}Q_\alpha(v_1,v_2)+\frac{\partial Q_\alpha(v_1,v_2)}{\partial v_1}\frac{\partial Q_\alpha(v_1,v_2)}{\partial v_2}}{Q^2_\alpha(v_1,v_2)}\right)\\
&= \frac{\frac{\partial v_1}{\partial u_1}\frac{\partial v_2}{\partial u_2}}{Q^2_\alpha(v_1,v_2)}\left(\Big(L''_\Lambda(-\ln Q_\alpha(v_1,v_2))+L'_\Lambda(-\ln Q_\alpha(v_1,v_2))\Big)\frac{\partial Q_\alpha(v_1,v_2)}{\partial v_1}\frac{\partial Q_\alpha(v_1,v_2)}{\partial v_2}\right.\\
&\quad \left.-L'_\Lambda(-\ln Q_\alpha(v_1,v_2))Q_\alpha(v_1,v_2)\frac{\partial^2 Q_\alpha(v_1,v_2)}{\partial v_1\partial v_2}\right).
\end{aligned}$$

We derive next the pdf c_Θ in (4.4.4): In this case, Λ follows a shifted Poisson distribution. In view of (5.1.3), we need to compute at first the following components:

$$\begin{aligned}
L'_\Lambda(-\ln Q_\alpha(v_1,v_2)) &= -e^{-\theta(1-Q_\alpha(v_1,v_2))}Q_\alpha(v_1,v_2)(1+\theta Q_\alpha(v_1,v_2)),\\
L''_\Lambda(-\ln Q_\alpha(v_1,v_2)) &= e^{-\theta(1-Q_\alpha(v_1,v_2))}Q_\alpha(v_1,v_2)(1+3\theta Q_\alpha(v_1,v_2)+\theta^2 Q_\alpha^2(v_1,v_2)),
\end{aligned}$$

where for $i=1,2$, $v_i = e^{-L_\Lambda^{-1}(u_i)}$ which implies $u_i = v_i e^{-\theta(1-v_i)}$ and thus $\frac{\partial v_i}{\partial u_i} = \frac{e^{-\theta(1-v_i)}}{1+\theta v_i}$. By replacing these components into (5.1.3), we have

$$\begin{aligned}
c_\Theta(u_1,u_2) &= \frac{1}{Q_\alpha(v_1,v_2)^2}\frac{e^{\theta(2-v_1-v_2)}}{(1+\theta v_1)(1+\theta v_2)}\\
&\quad\times\Bigg(\Big(e^{-\theta(1-Q_\alpha(v_1,v_2))}Q_\alpha(v_1,v_2)(1+3\theta Q_\alpha(v_1,v_2)+\theta^2Q_\alpha^2(v_1,v_2))\\
&\quad -e^{-\theta(1-Q_\alpha(v_1,v_2))}Q_\alpha(v_1,v_2)(1+\theta Q_\alpha(v_1,v_2))\Big)\frac{\partial Q_\alpha(v_1,v_2)}{\partial v_1}\frac{\partial Q_\alpha(v_1,v_2)}{\partial v_2}\\
&\quad +e^{-\theta(1-Q_\alpha(v_1,v_2))}Q_\alpha(v_1,v_2)(1+\theta Q_\alpha(v_1,v_2))Q_\alpha(v_1,v_2)\frac{\partial^2 Q_\alpha(v_1,v_2)}{\partial v_1\partial v_2}\Bigg)\\
&= \frac{1}{Q_\alpha(v_1,v_2)^2}\frac{e^{\theta(2-v_1-v_2)}}{(1+\theta v_1)(1+\theta v_2)}\Bigg(e^{-\theta(1-Q_\alpha(v_1,v_2))}Q_\alpha(v_1,v_2)\\
&\quad\times\Big(1+3\theta Q_\alpha(v_1,v_2)+\theta^2Q_\alpha^2(v_1,v_2)-1-\theta Q_\alpha(v_1,v_2)\Big)\\
&\quad\times\frac{\partial Q_\alpha(v_1,v_2)}{\partial v_1}\frac{\partial Q_\alpha(v_1,v_2)}{\partial v_2}\\
&\quad +e^{-\theta(1-Q_\alpha(v_1,v_2))}Q_\alpha^2(v_1,v_2)(1+\theta Q_\alpha(v_1,v_2))\frac{\partial^2 Q_\alpha(v_1,v_2)}{\partial v_1\partial v_2}\Bigg)\\
&= \frac{1}{Q(v_1,v_2)^2}\frac{e^{\theta(2-v_1-v_2)}}{(1+\theta v_1)(1+\theta v_2)}\\
&\quad\times\Bigg(e^{-\theta(1-Q_\alpha(v_1,v_2))}Q_\alpha(v_1,v_2)\Big(2\theta Q_\alpha(v_1,v_2)+\theta^2Q_\alpha^2(v_1,v_2)\Big)\\
&\quad\times\frac{\partial Q_\alpha(v_1,v_2)}{\partial v_1}\frac{\partial Q_\alpha(v_1,v_2)}{\partial v_2}\\
&\quad +e^{-\theta(1-Q_\alpha(v_1,v_2))}Q_\alpha^2(v_1,v_2)(1+\theta Q_\alpha(v_1,v_2))\frac{\partial^2 Q_\alpha(v_1,v_2)}{\partial v_1\partial v_2}\Bigg)\\
&= \frac{1}{Q(v_1,v_2)^2}\frac{e^{\theta(2-v_1-v_2)}}{(1+\theta v_1)(1+\theta v_2)}\\
&\quad\times\Bigg(e^{-\theta(1-Q_\alpha(v_1,v_2))}Q_\alpha^2(v_1,v_2)\Big(2\theta+\theta^2Q_\alpha(v_1,v_2)\Big)\frac{\partial Q_\alpha(v_1,v_2)}{\partial v_1}\frac{\partial Q_\alpha(v_1,v_2)}{\partial v_2}
\end{aligned}$$

$$
\begin{aligned}
&+e^{-\theta(1-Q_\alpha(v_1,v_2))}Q_\alpha^2(v_1,v_2)(1+\theta Q_\alpha(v_1,v_2))\frac{\partial^2 Q_\alpha(v_1,v_2)}{\partial v_1 \partial v_2}\Bigg) \\
= & \frac{e^{-\theta(1-Q_\alpha(v_1,v_2))}Q_\alpha^2(v_1,v_2)}{Q(v_1,v_2)^2}\frac{e^{\theta(2-v_1-v_2)}}{(1+\theta v_1)(1+\theta v_2)} \\
&\times\Bigg(\theta\frac{\partial Q_\alpha(v_1,v_2)}{\partial v_1}\frac{\partial Q_\alpha(v_1,v_2)}{\partial v_2}\Big(2+\theta Q_\alpha(v_1,v_2)\Big) \\
&+(1+\theta Q_\alpha(v_1,v_2))\frac{\partial^2 Q_\alpha(v_1,v_2)}{\partial v_1 \partial v_2}\Bigg) \\
= & \frac{e^{\theta(2-v_1-v_2)}e^{\theta(Q_\alpha(v_1,v_2)-1)}}{(1+\theta v_1)(1+\theta v_2)}\Bigg(\frac{\partial^2 Q_\alpha(v_1,v_2)}{\partial v_1 \partial v_2}(1+\theta Q_\alpha(v_1,v_2)) \\
&+\theta\frac{\partial Q_\alpha(v_1,v_2)}{\partial v_1}\frac{\partial Q_\alpha(v_1,v_2)}{\partial v_2}(2+\theta Q_\alpha(v_1,v_2))\Bigg).
\end{aligned}
$$

Next, we show (4.4.5): Since Λ follows a truncated Poisson distribution, in light of (5.1.3), the joint density c_Θ is expressed in terms of (set $\eta_\theta = e^{-\theta}/(1-e^{-\theta})$)

$$
\begin{aligned}
L'_\Lambda(-\ln Q_\alpha(v_1,v_2)) &= -\eta_\theta\theta Q_\alpha(v_1,v_2)e^{\theta Q_\alpha(v_1,v_2)}, \\
L''_\Lambda(-\ln Q_\alpha(v_1,v_2)) &= \eta_\theta\theta Q_\alpha(v_1,v_2)e^{\theta Q_\alpha(v_1,v_2)}(1+\theta Q_\alpha(v_1,v_2)),
\end{aligned}
$$

where for $i=1,2$, $v_i = e^{-L_\Lambda^{-1}(u_i)}$ and $u_i = \frac{e^{-\theta}}{1-e^{-\theta}}(v_i^\theta - 1)$ with $\frac{\partial v_i}{\partial u_i} = \frac{1-e^{-\theta}}{\theta}e^{\theta(1-v_i)}$. By substituting the above components in the joint density expressed in (5.1.3), we obtain

$$
\begin{aligned}
c_\Theta(u_1,u_2) = & \left(\frac{1-e^{-\theta}}{\theta}\right)^2\frac{e^{\theta(2-v_1-v_2)}}{Q_\alpha^2(v_1,v_2)}\Bigg[\Bigg(\eta_\theta\theta Q_\alpha(v_1,v_2)e^{\theta Q(v_1,v_2)}(1+\theta Q_\alpha(v_1,v_2)) \\
&-\eta_\theta\theta Q_\alpha(v_1,v_2)e^{\theta Q_\alpha(v_1,v_2)}\Bigg)\frac{\partial Q_\alpha(v_1,v_2)}{\partial v_1}\frac{\partial Q_\alpha(v_1,v_2)}{\partial v_2} \\
&+\eta_\theta\theta Q_\alpha(v_1,v_2)e^{\theta Q_\alpha(v_1,v_2)}Q_\alpha(v_1,v_2)\frac{\partial^2 Q_\alpha(v_1,v_2)}{\partial v_1 \partial v_2}\Bigg] \\
= & (1-e^{-\theta})e^{\theta[1-v_1-v_2+Q_\alpha(v_1,v_2)]}\left(\frac{\partial Q_\alpha(v_1,v_2)}{\partial v_1}\frac{\partial Q_\alpha(v_1,v_2)}{\partial v_2}+\frac{1}{\theta}\frac{\partial^2 Q_\alpha(v_1,v_2)}{\partial v_1 \partial v_2}\right).
\end{aligned}
$$

□

Proof of Proposition 4.3.1 Since G has Fréchet marginals, by assumption (4.3.3), we have that

$$
\lim_{n\to\infty} G^n(nx,ny) = \mathcal{G}(x,y), \quad x,y \in (0,\infty),
$$

where $\mathcal{G}$ has copula Q_A and thus $\tau(\mathcal{G}) = \tau(Q_A)$. We have thus with $F_n(x,y) = \mathbb{E}\left\{G^{\Lambda_n}(x,y)\right\}$ using further (4.3.1)

$$\lim_{n\to\infty} F_n(nx_n, ny_n) = \lim_{n\to\infty} \mathbb{E}\left\{G^{n\frac{\Lambda_n}{n}}(nx_n, ny_n)\right\} = \mathcal{G}(x,y), \quad x,y \in (0,\infty) \quad (4.6.2)$$

for any x_n, y_n such that $\lim_{n\to\infty} x_n = x$ and $\lim_{n\to\infty} y_n = y$. Consequently,

$$\begin{aligned}
\tau(C_n) &= 4\int_{(0,\infty)^2} F_n(x,y)\, dF_n(x,y) - 1\\
&= 4\int_{(0,\infty)^2} F_n(nx,ny)\, dF_n(nx,ny) - 1\\
&\to 4\int_{(0,\infty)^2} \mathcal{G}(x,y)\, d\mathcal{G}(x,y) - 1, \quad n\to\infty\\
&= \tau(\mathcal{G}),
\end{aligned}$$

where the convergence above follows by Lemma 4.2 in Hashorva [58] (see also Resnick and Zeber [102] and Kulik and Soulier [76] for more general results). Next, the convergence in (4.6.2) implies

$$\lim_{n\to\infty} F_{ni}(ns_n) = \lim_{n\to\infty} \mathbb{E}\left\{G_i^{n\frac{\Lambda_n}{n}}(ns_n)\right\} = \mathcal{G}_i(s), \quad s\in(0,\infty), i=1,2$$

for any $s_n, n\ge 1$ such that $\lim_{n\to\infty} s_n = s$, where $F_{ni}, G_i, \mathcal{G}_i$ is the i^{th} marginal df of F_n, G, and $\mathcal{G}$, respectively. Hence, with similar arguments as above, we have

$$\begin{aligned}
\rho_S(C_n) &= 12\int_{(0,\infty)^2} F_n(x,y)\, dF_{n1}(x)dF_{n2}(y) - 3\\
&= 12\int_{(0,\infty)^2} F_n(nx,ny)\, dF_{n1}(nx)dF_{n2}(nx) - 3\\
&\to 12\int_{(0,\infty)^2} \mathcal{G}(x,y)\, d\mathcal{G}_1(x)d\mathcal{G}_2(x) - 3, \quad n\to\infty\\
&= \rho_S(\mathcal{G})
\end{aligned}$$

establishing the proof. □

Proof of Proposition 6.3.6 For $v = e^{-L_\Lambda^{-1}(1-u)}$ we have

$$1 - L_\Lambda(-\ln v) \sim u, \quad u\downarrow 0, \quad \lim_{u\downarrow 0} v = 1.$$

By the assumption that $\mathbb{E}\{\Lambda\}$ is finite we have

$$1 - L_\Lambda(t) \sim -L'_\Lambda(0)t = \mathbb{E}\{\Lambda\}\, t, \quad t\to 0. \quad (4.6.3)$$

Since further

$$\mu_Q = 2 - \lim_{u \downarrow 0} \frac{Q(1-u, 1-u)}{u} = 2 - \lim_{v \uparrow 1} \frac{\ln Q(v,v)}{\ln v}$$

and $\lim_{v\uparrow 1} Q(v,v) = 1$, then using (5.1.2) and (4.6.3) we obtain

$$\begin{aligned}
\mu_C &= 2 - \lim_{u \downarrow 0} u^{-1}[1 - C(1-u, 1-u))] \\
&= 2 - \lim_{u \downarrow 0} u^{-1}\Big[1 - L_\Lambda\big(-\ln Q(v,v)\big)\Big] \\
&= 2 - \lim_{u \downarrow 0} \frac{1 - L_\Lambda\big(-\ln Q(v,v)\big)}{1 - L_\Lambda(-\ln v)} \\
&= 2 - \lim_{v \uparrow 1} \frac{\ln Q(v,v)}{\ln v} \\
&= 2 - [2 - \mu_Q] = \mu_Q,
\end{aligned}$$

hence the first claim follows. Next, in view of (4.3.8) we have

$$\lim_{n\to\infty} n[1 - G(nx, ny)] = -\ln H(x,y), \quad x, y \in (0, \infty),$$

hence as $n \to \infty$

$$n[1 - G(nx, ny)] \sim \frac{1 - G(nx, ny)}{1 - G(n,n)} \sim -\ln H(x,y), \quad x, y \in (0, \infty).$$

Let $a_n, n \geq 1$ be non-negative constants such that $\lim_{n\to\infty} a_n = \infty$. By the above and (4.6.3)

$$n[1 - F(a_n x, a_n y)] = n[1 - L_\Lambda(-\ln G(a_n x, a_n y))] \sim \mathbb{E}\{\Lambda\}\, n(-\ln G(a_n x, a_n y))$$

as $n \to \infty$. Setting now $a_n = \mathbb{E}\{\Lambda\} n$ we have thus as $n \to \infty$

$$\begin{aligned}
n[1 - F(a_n x, a_n y)] &\sim a_n \frac{1 - F(a_n x, a_n y)}{\mathbb{E}\{\Lambda\}} \\
&= a_n \frac{1 - L_\Lambda(-\ln G(a_n x, a_n y))}{\mathbb{E}\{\Lambda\}} \\
&\sim a_n(-\ln G(a_n x, a_n y)) \\
&\sim a_n[1 - G(a_n x, a_n y)[\\
&\sim \mathbb{E}\{\Lambda\}\Big(-\ln H(x\mathbb{E}\{\Lambda\}, y\mathbb{E}\{\Lambda\})\Big) \\
&= -\ln H(x,y)
\end{aligned}$$

establishing the proof. □

Chapter 5

Mixture Copulas and Insurance Applications

This chapter is based on M. Tamraz: Mixture Copulas and Insurance Applications, published in *Annals of Actuarial Science*.

5.1 Introduction

In insurance applications, modeling of multivariate data is crucial for instance for pricing of dependent risks, risk management of different portfolios or reinsurance modeling of joint risks. The choice of tractable multivariate distributions for such modeling purposes is large. In this contribution, we are concerned with the joint distribution of the largest claims observed in two insurance portfolios. In this respect, we denote by X_i and Y_i the i^{th} claim observed in each portfolio respectively. We consider the classical collective model over a fixed period of time. Hence, we define N as the claim counting random variable. Clearly, when $N = 0$, no claims are reported and the largest claim observed in each portfolio is null. However, we are mainly interested in the case where $N \geq 1$. Therefore, we define $\Lambda = N|N \geq 1$. For a given bivariate distribution function G, a new class of bivariate distributions, denoted F, can be introduced in the context of the distribution of largest claims observed in a bivariate portfolio as illustrated in Hashorva et al. [60]. An instance that motivates F in practice is if X_i's model the claim sizes of an insurance portfolio and Y_i's the allocated loss adjustment expense related to the settlement of X_i's, such as legal fees, investigations of claims, etc. The dependency observed between the largest claims of X_i and Y_i is relevant when pricing an excess-of-loss reinsurance treaty in the case where the insurer and reinsurer share the settlement costs, see Cebrian et al. [20] for more details.

More specifically, if Λ is a discrete random variable with $\mathbb{P}\{\Lambda = i\} = p(i) \geq 0, i \in \mathbb{N}$, then F can be defined by its Laplace transform, see Joe [67][Chapter 4.2]

$$F(x, y) = L_\Lambda(-\ln G(x, y)), \quad x, y \geq 0, \tag{5.1.1}$$

with L_Λ the Laplace transform of Λ. Moreover, if G has continuous marginal df's G_1, G_2, we have that F has marginal df's

$$F_i(x_i) = L_\Lambda(-\ln G_i(x_i)), \quad i = 1, 2,$$

and unique copula

$$C(u_1, u_2) = L_\Lambda\big(-\ln Q(v_1, v_2)\big), \qquad u_1, u_2 \in [0, 1], \tag{5.1.2}$$

where we set $v_i = e^{-L_\Lambda^{-1}(u_i)}$ and Q the unique copula of G. Note that by differentiating (5.1.2) we get the corresponding copula density c of C as follows

$$\begin{aligned} c(u_1, u_2) = \frac{\frac{\partial v_1}{\partial u_1}\frac{\partial v_2}{\partial u_2}}{Q^2(v_1, v_2)} \Bigg(& \big(L'_\Lambda(t) + L''_\Lambda(t)\big) \frac{\partial Q(v_1, v_2)}{\partial v_1} \frac{\partial Q(v_1, v_2)}{\partial v_2} \\ & - L'_\Lambda(t) Q(v_1, v_2) q(v_1, v_2) \Bigg), \end{aligned} \tag{5.1.3}$$

where $L'_\Lambda(s) = -\mathbb{E}\left\{\Lambda e^{-s\Lambda}\right\}$, $L''_\Lambda(s) = \mathbb{E}\left\{\Lambda^2 e^{-s\Lambda}\right\}$ and q the pdf of Q given by

$$q(u_1, u_2) = \frac{g(G_1^{-1}(u_1), G_2^{-1}(u_2))}{g_1(G_1^{-1}(u_1)) g_2(G_2^{-1}(u_2))}, \quad u_1, u_2 \in [0, 1].$$

Here g is the pdf of G and g_1, g_2 its marginal pdf's.
In the aforementioned paper, three special cases for Λ were considered by transforming a discrete random variable N, namely Shifted Geometric, Shifted Poisson or Truncated Poisson. For these choices of Λ, the density function c has a very tractable form and therefore can be easily used for parametric estimation purposes. As seen from (5.1.3), it is crucial that we have a tractable formula for the Laplace transform L_Λ or that of the random variable N. Instead of the Poisson choice for N we can take for instance

$$N \stackrel{d}{=} Poisson(\lambda W), \tag{5.1.4}$$

where $\lambda > 0$ is a fixed parameter and W is a modifier, i.e., a non-negative random variable. Here $\stackrel{d}{=}$ means equality in distribution. Clearly, this idea carries over to

other parametric models for Λ.
Another interesting choice of N is motivated by the Poisson case. Clearly, we can write a Poisson random variable as a compound Poisson random variable. Thus, with motivation from the collective model, we consider N to be a compound random variable as follows

$$N = \sum_{i=1}^{Y} Z_i, \quad N = 0 \text{ if } Y = 0, \tag{5.1.5}$$

where Y is a counting random variable independent of Z_i's which are discrete random variables with values in $0, 1 \ldots$. Both constructions above are interesting and lead to new classes of mixture copulas. The drawback is that in many cases, no explicit form of the Laplace transform is available, which renders the parametric estimation difficult. In this paper, we shall focus on a tractable choice for Z_i's, namely these are independent copies of a Sibuya random variable Z with probability generating function (pgf)

$$P_Z(z) = 1 - (1-z)^\alpha,$$

with $\alpha \in (0, 1]$ a fixed parameter. For such Z and Y a Poisson random variable with parameter $\lambda > 0$, then N given by (5.1.5) has a discrete-stable distribution with parameters $\lambda > 0$ and $\alpha \in (0, 1]$. Discrete-stable distributions have been discussed in Steutel and Van Harn [112]. These so-called discrete distributions satisfy many interesting properties. Specifically, in view of Steutel and Van Harn [112] the probability generating function (pgf) P of a discrete-stable distribution N with parameters (α, λ) is of the form

$$P_N(z) = e^{-\lambda(1-z)^\alpha}, \tag{5.1.6}$$

where $\lambda > 0$, $\alpha \in (0, 1]$ and $|z| \leq 1$. By setting $z = e^{-t}$ in (5.1.6), we can define the distribution of N via its Laplace transform

$$L_N(t) = \mathbb{E}\left\{e^{-tN}\right\} = e^{-\lambda[1-e^{-t}]^\alpha}, \quad t \geq 0, \lambda > 0, \alpha \in (0, 1]. \tag{5.1.7}$$

We have the following explicit formulas

$$\mathbb{P}\{N = 0\} = e^{-\lambda}, \tag{5.1.8}$$

$$\mathbb{P}\{N = 1\} = \alpha\lambda e^{-\lambda}, \tag{5.1.9}$$

$$\mathbb{P}\{N = 2\} = \frac{\alpha\lambda}{2} e^{-\lambda}(1 - \alpha + \alpha\lambda). \tag{5.1.10}$$

Clearly, we obtain the Poisson distribution for $\alpha = 1$ in (5.1.7).
The case $\alpha < 1$ is substantially different from the Poisson case of $\alpha = 1$. Indeed, if $\alpha < 1$, then $\mathbb{E}(N) = \infty$ and for such α the discrete-stable distribution is heavy-tailed.
Hereafter, we shall discuss two different models for Λ based on N as above, namely
Model A: (Shifted discrete-stable) Setting $\Lambda = N + 1$, we have that

$$L_\Lambda(t) \;=\; \mathbb{E}\left\{e^{-t(1+N)}\right\} = e^{-t}e^{-\lambda[1-e^{-t}]^\alpha}, \quad t \geq 0, \tag{5.1.11}$$

hence

$$F(x, y) \;=\; G(x, y)e^{-\lambda[1-G(x,y)]^\alpha}, \quad x, y \geq 0. \tag{5.1.12}$$

Model B: (Truncated discrete-stable) We define $\Lambda = N|N \geq 1$. Hence $p(i) = p(i)/(1 - p(0)), i \geq 1$. This implies that

$$\begin{aligned} \mathbb{E}\left\{e^{-t\Lambda}\right\} &= \sum_{i=1}^{\infty} \frac{p(i)}{1-p(0)}e^{-ti} + \frac{p(0)}{1-p(0)} - \frac{p(0)}{1-p(0)} \\ &= \frac{e^{-\lambda(1-e^{-t})^\alpha} - e^{-\lambda}}{1-e^{-\lambda}}, \end{aligned} \tag{5.1.13}$$

leading to (set $b_\lambda := \frac{e^{-\lambda}}{1-e^{-\lambda}}$)

$$F(x, y) \;=\; \frac{1}{1-e^{-\lambda}}[e^{-\lambda[1-G(x,y)]^\alpha} - e^{-\lambda}] = b_\lambda[e^{\lambda(1-[1-G(x,y)]^\alpha)} - 1], \quad t \geq 0. \tag{5.1.14}$$

All the models introduced above lead to distribution functions F which depend on two additional parameters α and λ. The dependence introduced by the choice of G and (α, λ) is interesting. Even if we have the product case

$$G(x, y) = G_1(x)G_2(y), \quad x, y \in \mathbb{R}$$

the distribution function F is not a product distribution.
In this paper, we are interested in the main properties of F for Λ specified as above and the possible applications of such F for modeling dependent insurance data. This paper is structured as follows. In Section 2, we study some dependence properties of F by means of Monte Carlo simulations for **Model A**. Section 3 discusses various methods for estimating the parameters of the new constructed copula as well as goodness of fit. Finally, Section 4 is dedicated to applications of this copula to

concrete insurance data sets.

5.2 Dependence properties of F

5.2.1 Dependence measures

We investigate the dependence properties of F and its corresponding copula C for a given joint distribution function G with copula Q and Λ as in **Model A**. The dependence between the largest claim amounts observed in two insurance portfolios, i.e., $(X_{\Lambda:\Lambda}, Y_{\Lambda:\Lambda})$, is evaluated with respect to the parameter α of the shifted discrete-stable distribution. In this respect, we define below the most commonly used non-parametric methods. They measure different aspects of the dependence structure governed by C, see Fredricks and Nelsen [42].
For a given copula C, Kendall's τ is defined as

$$\tau(C) = 4 \int_0^1 \int_0^1 C(u, v) dC(u, v) - 1,$$

whereas Spearman's rank correlation coefficient ρ_S is given by

$$\rho_S(C) = 12 \int_0^1 \int_0^1 C(u, v) du dv - 3 = 12 \int_0^1 \int_0^1 uv dC(u, v) - 3.$$

Remarks 5.2.1. *We show in the next section by simulation that for a given copula Q with parameter θ, the level of dependence governed by C decreases as the parameter α of the shifted discrete-stable distribution increases.*

5.2.2 Monte Carlo Simulation

We study the dependence property of C by means of Monte Carlo simulations. In order to do so, we consider two types of copula for Q, namely the Gumbel and the Clayton copula both with parameter $\theta = 10$ and Λ as defined in **Model A** with $\lambda = 10$. We compare the level of dependence governed by copula C with respect to the parameter α of the discrete-stable distribution. The simulation procedure follows **Step 1- Step 4** described below and is repeated $10'000$ times.
Step 1: Generate a random value n from the discrete-stable distribution Λ.
Step 2: Generate n random vectors (U_{11}, U_{21}), ...,(U_{1n}, U_{2n}) from copula Q.

Step 3: Calculate $Z_1 = \max(U_{11}, \ldots, U_{1n})$ and $Z_2 = \max(U_{21}, \ldots, U_{2n})$.
Step 4: Return the vector (V_1, V_2) with

$$V_1 = Z_1 e^{-\lambda(1-Z_1)^\alpha}, \quad V_2 = Z_2 e^{-\lambda(1-Z_2)^\alpha}.$$

In view of **Step 1**, simulating a random number from the discrete-stable distribution is not straightforward as it does not have a closed form for its probability mass function. Therefore, Devroye [30] developed the following result.

Lemma 5.2.2. *If a random variable X follows a discrete-stable distribution with parameters (α, λ), then X follows a Poisson distribution with parameter $\lambda^{1/\alpha} S_{\alpha,1}$ where $S_{\alpha,1}$ a positive stable random variate with parameter α.*

Several methods were discussed in the literature for the choice of $S_{\alpha,1}$. In the sequel, we refer to the method described in Kanter [69] where the expression of $S_{\alpha,1}$ is given by

$$S_{\alpha,1} = \left(\frac{\sin((1-\alpha)\pi U)}{E \sin(\alpha \pi U)} \right)^{\frac{(1-\alpha)}{\alpha}} \left(\frac{\sin(\alpha \pi U)}{\sin(\pi U)} \right)^{\frac{1}{\alpha}}$$

with $\alpha \in (0,1)$ and $U \sim Uniform(0,1)$ being independent of the unit exponential random variable $E \sim Exp(1)$.
In Table 5.2.1 below, we compute the empirical dependence measures relative to copula C.

	Q: Gumbel copula with $\theta = 10$		Q: Clayton copula with $\theta = 10$	
α	$\tau(C)$	$\rho_S(C)$	$\tau(C)$	$\rho_S(C)$
1.00	0.9034	0.9863	0.3511	0.5030
0.90	0.9104	0.9882	0.3549	0.5034
0.80	0.9221	0.9911	0.3771	0.5321
0.75	0.9254	0.9919	0.4034	0.5645
0.60	0.9393	0.9946	0.4656	0.6396

Table 5.2.1: Empirical Kendall's Tau and Spearman's rho with respect to α.

Table 5.2.1 shows that as the parameter α of the discrete-stable distribution decreases, the level of dependence governed by copula C increases. Note that $\alpha = 1$, i.e. the Shifted Poisson case, yields the lowest level of dependence between the maximal claim amounts observed in the two portfolios.

Figure 5.2.2 below highlights the results observed in Table 5.2.1. However, we reduce the number of simulations to $1'000$. We consider the Gumbel copula as a model for Q with parameter $\theta = 10$. It is clear that when α tends to 0, the level of dependence is high and V_1 and V_2 are in some sense proportional.

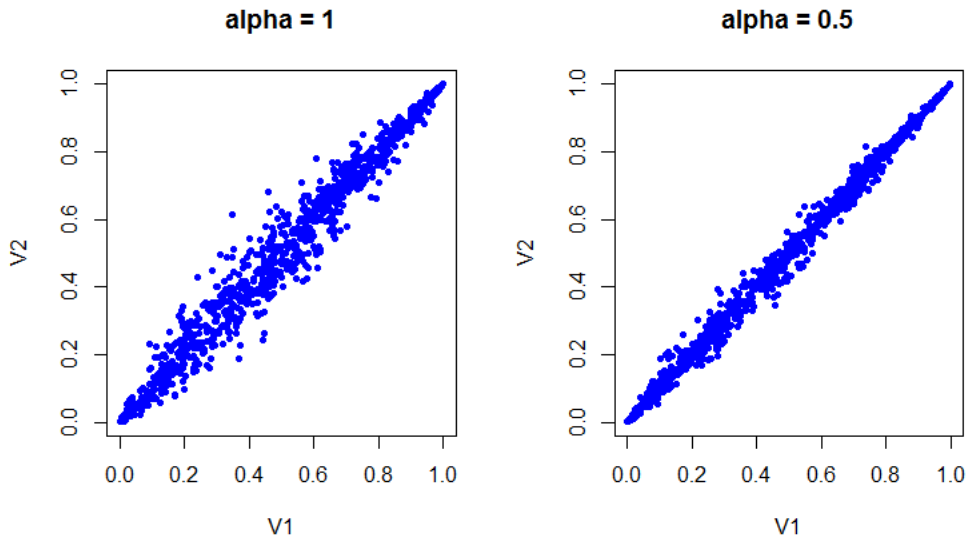


Figure 5.2.1: Scatterplot of the simulated vector (V_1, V_2) - Gumbel copula.

For the Clayton copula with parameter $\theta = 10$, we observe almost the same phenomenon as in Figure 5.2.2. For $\alpha = 1$, the observations are widely spread around the identity line and there isn't a clear dependence structure between V_1 and V_2. However, in the second graph, for $\alpha = 0.5$, the observations are more condensed around the identity line and shows some kind of dependency close to the Gumbel case, especially in the tails.

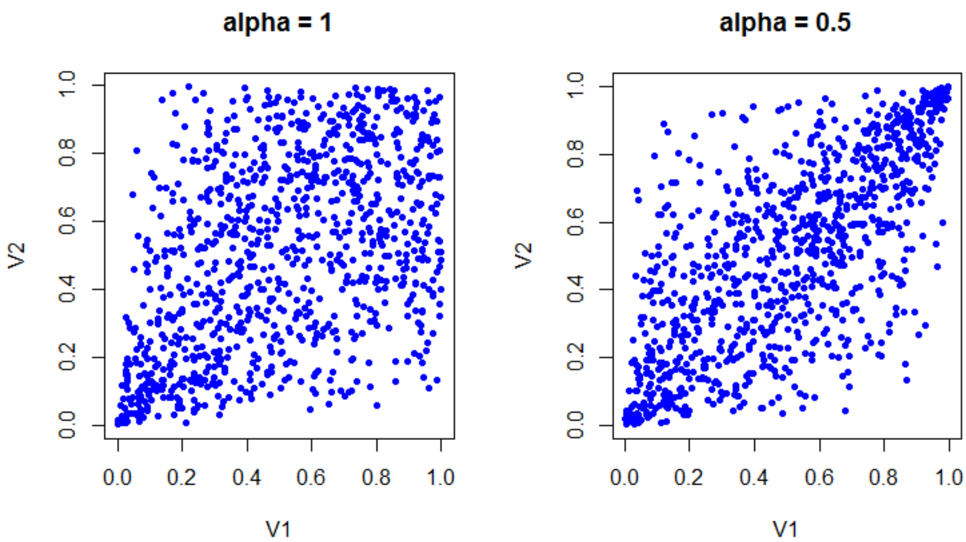


Figure 5.2.2: Scatterplot of the simulated vector (V_1, V_2) - Clayton copula.

5.3 Parameter estimation and Goodness of fit

5.3.1 Parameter Estimation

In this section, we discuss different methods to estimate the parameters of the new copula C, principally the parameter θ of the original copula Q and the parameters (α, λ) of the discrete-stable distribution Λ. In the sequel, for the sake of simplicity, we denote by $\Theta = (\theta, \alpha, \lambda)$ the parameters of the new copula. The different methods for estimating the parameters of a copula is widely discussed in the literature. We count two parametric methods and one semi-parametric. The choice of one of these methods depends on the willingness of the user to make assumptions or not about the unknown margins.

Typically, when marginal distributions are known, parametric methods are more frequently employed. The most popular method discussed in the literature is the Maximum Likelihood Estimation (MLE). It is a fully parametric method well known for its optimality properties. The parameters of the copula and of the marginal distributions are estimated simultaneously by maximising the log-likelihood function. However, this method is computationally intensive especially when estimating multiple parameters. An alternative method that requires less computations is the Inference Functions for Margins (IFM) proposed by Joe [67]. It is a 2-steps estimation method. In the first step, the parameters of the marginal distributions are estimated separately by maximising the corresponding log-likelihood functions. Next, by replacing the marginal parameters by their first stage estimators, the maximisation of the log-likelihood solves for Θ in the second step. However, both methods rely on the choice of the marginals. Kim et al. [70] shows that a misspecification of the marginals may lead to discrepancies in the performance of the estimators.

In practice, the marginal distributions are unknown and are thus estimated non-parametrically. Genest et al. [45] described a new method for estimating the dependence parameter Θ of the copula C which is a semi-parametric one known as the Pseudo-Maximum Likelihood (PML) method. It is solely based on the ranks of the observations. In the first stage, the marginals are replaced by their empirical counterparts in the pseudo log-likelihood function. Then, in the second stage, the maximisation of the latter returns the estimators of Θ of the new copula C. In the sequel, we shall utilise the PML method.
Let $X \sim G_1$ and $Y \sim G_2$ where G_1 and G_2 are the marginals of X and Y respectively. In light of the PML method, G_1 and G_2 are estimated by their empirical counterparts denoted hereafter by $\widetilde{G_1}$ and $\widetilde{G_2}$ and defined as follows

$$\widetilde{G_1}(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{X_i \leq x\}, \quad \widetilde{G_2}(y) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{Y_i \leq y\}.$$

The method consists in finding Θ that maximises the pseudo log-likelihood function given by

$$l(\Theta) = \sum_{i=1}^{n} \log c_\Theta(U_i, V_i), \tag{5.3.1}$$

where c_Θ is the copula density defined in (5.1.3), $U_i = \frac{n}{n+1}\widetilde{G_1}(x_i)$ and $V_i = \frac{n}{n+1}\widetilde{G_2}(y_i)$ are the pseudo-observations. The rescaling factor $\frac{n}{n+1}$ is introduced to avoid numerical difficulties arising at the boundaries $[0,1]^2$, see Genest et al. [45]. Kim et al. [70] shows that the PML methods performs better than the IFM and MLE methods. Moreover, Genest et al. [45] and Shih and Louis [110] (in the presence of censorship) show that under suitable conditions, the resulting estimator of Θ is consistent and asymptotically normally distributed.
Below, we give the expression of the copula density for both **Model A** and **Model B** along with the corresponding pseudo log-likelihood functions.

- **Model A:** Λ follows a shifted discrete-stable distribution.

The pdf of the shifted discrete-stable copula is given by

$$c_\Theta(u_1,u_2) = W(v_1,v_2)\frac{e^{\lambda\left[(1-v_1)^\alpha+(1-v_2)^\alpha-(1-Q(v_1,v_2))^\alpha\right]}}{(1+\lambda\alpha v_1(1-v_1)^{\alpha-1})(1+\lambda\alpha v_2(1-v_2)^{\alpha-1})}, \quad (5.3.2)$$

where

$$\begin{aligned} W(v_1,v_2) &= \lambda\alpha(1-Q(v_1,v_2))^{\alpha-2}\Big[2-(\alpha+1)Q(v_1,v_2)+\lambda\alpha Q(v_1,v_2)(1-Q(v_1,v_2))^\alpha\Big] \\ &\times\frac{\partial Q(v_1,v_2)}{\partial v_1}\frac{\partial Q(v_1,v_2)}{\partial v_2}+\Big[1+\lambda\alpha Q(v_1,v_2)(1-Q(v_1,v_2))^{\alpha-1}\Big]\frac{\partial^2 Q(v_1,v_2)}{\partial v_1\partial v_2}, \end{aligned}$$

$$\text{and } v_j = f^{-1}(u_j) \text{ such that } f(v_j) = v_j e^{-\lambda(1-v_j)^\alpha}, j=1,2.$$

The resulting pseudo log-likelihood function of the above copula can be written as follows

$$\begin{aligned} l(\Theta) &= \sum_{i=1}^n\Big[\lambda[(1-v_{1i})^\alpha+(1-v_{2i})^\alpha-(1-Q(v_{1i},v_{2i}))^\alpha]-\ln(1+\lambda\alpha v_{1i}(1-v_{1i})^{\alpha-1}) \\ &-\ln(1+\lambda\alpha v_{2i}(1-v_{2i})^{\alpha-1})+\ln W(v_{1i},v_{2i})\Big]. \end{aligned}$$

- **Model B:** Λ follows a truncated discrete-stable distribution.
The joint density of the truncated discrete-stable copula is of the form

$$\begin{aligned} c(u_1,u_2) &= W(v_1,v_2)\Big(\frac{1-e^{-\lambda}}{\lambda\alpha}\frac{e^{\lambda[(1-v_1)^\alpha+(1-v_2)^\alpha-(1-Q(v_1,v_2))^\alpha]}}{(1-v_1)^{\alpha-1}(1-v_2)^{\alpha-1}} \\ &\times(1-Q(v_1,v_2))^{\alpha-2}\Big), \end{aligned} \quad (5.3.3)$$

where

$$\begin{aligned} W(v_1,v_2) &= (1-\alpha+\lambda\alpha(1-Q(v_1,v_2))^\alpha)\frac{\partial Q(v_1,v_2)}{\partial v_1}\frac{\partial Q(v_1,v_2)}{\partial v_2} \\ &+(1-Q(v_1,v_2))\frac{\partial^2 Q(v_1,v_2)}{\partial v_1\partial v_2}, \end{aligned}$$

$$\text{and } v_j = 1-\left[-\frac{\ln(e^{-\lambda}+u_j(1-e^{-\lambda}))}{\lambda}\right]^{\frac{1}{\alpha}}, j=1,2.$$

The resulting pseudo log-likelihood function of the above copula is given by

$$l(\Theta) = \sum_{i=1}^n\Big[\ln(1-e^{-\lambda})-\ln(\lambda\alpha)+\lambda[(1-v_{1i})^\alpha+(1-v_{2i})^\alpha-(1-Q(v_{1i},v_{2i}))^\alpha]$$

$$-(\alpha-1)\ln(1-v_{1i})-(\alpha-1)\ln(1-v_{2i})+(\alpha-2)\ln(1-Q(v_{1i},v_{2i}))$$
$$+\ln W(v_{1i},v_{2i})\Big]. \tag{5.3.4}$$

5.3.2 Goodness of fit

Following the estimation of the parameter Θ, one need to assess the fit of the parametric copula C_Θ to a given data set. In this respect, we consider the hypothesis tests

$$H_0 : C_\Theta \in \mathcal{C}_0 \quad \text{against} \quad H_1 : C_\Theta \notin \mathcal{C}_0,$$

where $\mathcal{C}_0 = \{C_\Theta : \Theta \in \mathcal{O}\}$ is a class of some known parametric copulas and $\mathcal{O}$ an open subset of $\mathbb{R}^p$ for some integer $p \geq 1$. We refer to Genest et al. [51] for a review of the different methods used to assess the goodness of fit of a parametric copula.
In the sequel, we shall use the Cramer-von Mises test statistic, denoted hereafter by CVM. The corresponding statistic of this test is denoted by S_n and is defined as follows (let $\mathbb{C}_n(u,v) = \sqrt{n}\,(C_n(u,v) - C_\Theta(u,v)))$

$$S_n = \int_0^1 \int_0^1 \mathbb{C}_n(u,v)^2 dC_n(u,v),$$

where $C_n(u,v) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}\{U_i \leq u, V_i \leq v\}$ is the empirical copula and C_Θ the fitted copula with parameter Θ. It is worth mentioning that C_n depends solely on the pseudo-observations U_i and V_i, see Deheuvels [26], Genest and Favre [47]. Large values for this test lead to the rejection of H_0.
Moreover, one might be interested to compute the P-values associated to S_n. The larger the P-value the less likely is the rejection of the hypothesis H_0 at a significance level p. Genest et al. [51] described a parametric bootstrap procedure for the computation of the P-values corresponding to the goodness of fit using the CVM test statistic. The procedure is summarized under the following steps

- **Step 1:** Calculate the MLE $\widehat{\Theta}$ of Θ using the PML method.
- **Step 2:** Compute the value of the test S_n with $S_n = \sum_{i=1}^n \left(C_n(U_i,V_i) - C_{\widehat{\Theta}}(U_i,V_i)\right)^2$.
- **Step 3:** Let K denotes the number of bootstrap replications. Repeat Steps 4-6 for $k \in \{1,\ldots,K\}$:
  - **Step 4:** Generate a random sample $(\hat{X}_{i,k}, \hat{Y}_{i,k})$ for $i \in \{1,\ldots,n\}$ from

$C_{\hat{\Theta}}$ as described in Section 5.2.2 and compute their corresponding pseudo-observations, i.e, $\hat{U}_{i,k} = \frac{n}{n+1}\hat{X}_{i,k}$ and $\hat{V}_{i,k} = \frac{n}{n+1}\hat{Y}_{i,k}$.

– **Step 5:** Compute the empirical copula $C_{n,k}(u,v) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left\{\hat{U}_{i,k} \leq u, \hat{V}_{i,k} \leq v\right\}$ and estimate the MLE of Θ by $\widehat{\Theta}_k$ at each iteration based on $(\hat{U}_{i,k}, \hat{V}_{i,k})$ for $i \in \{1, \ldots, n\}$.

– **Step 6:** Calculate $S_{n,k} = \sum_{i=1}^{n}\left(C_{n,k}(\hat{U}_{i,k}, \hat{V}_{i,k}) - C_{\hat{\Theta}_k}(\hat{U}_{i,k}, \hat{V}_{i,k})\right)^2$.

An approximate P-value for this test is given by

$$p = \frac{1}{K}\sum_{k=1}^{K} \mathbf{1}\left(S_{n,k} \geq S_n\right).$$

Note that the largest the sample size, the more accurate the bootstrap procedure is, see Genest et al. [51].

5.4 Insurance Applications

For illustration purposes, we consider real insurance data set applications. The original copula Q can be one of the following copulas: Gumbel, Frank, Student and Joe, see Appendix for more details on the copulas. Also, Λ with parameters (α, λ) follows one of the two distributions: shifted discrete-stable and truncated discrete-stable. We construct a new copula based on Q and Λ and assess the fit of this new family of copula to insurance data sets. We use the AIC criteria to assess the quality of each model relative to each of the other models. It is defined as

$$AIC = -2(\widehat{\Theta}) + 2p$$

where $p = 3$ is the number of parameters to estimate and $(\widehat{\Theta})$ the pseudo log-likelihood function as in (5.3.1) evaluated at $\widehat{\Theta}$, estimator of Θ. Moreover, we use the CVM test to assess the goodness of fit of the copula to the data sets and compute the corresponding P-values relative to each copula model. Additionally, we include the root mean square error to measure the differences between the observed values and the ones predicted by the model. It is denoted hereafter by RMSE and is defined as follows, see Vandenberghe et al. [116] for instance

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(C_{\Theta}(U_i, V_i) - C_n(U_i, V_i))^2}$$

where C_n is the empirical copula based on the observed values and C_Θ the fitted one. Both models are then compared to other families of copulas already considered in Hashorva et al. [60].

Remarks 5.4.1. *In light of the bootstrap procedure described in Section 5.3.2 for the computation of the P-value, we set $K = 1000$ that is K random samples of size n are generated where n corresponds to the size of each data set. Generally, K should be taken larger than the size of the data set used, see Genest et al. [51]. However, this is computationally intensive for most of the data sets considered.*

5.4.1 Loss ALAE from medical insurance

In this section, we consider the SOA Medical Group Insurance data sets describing the medical claims observed over the years 1991-1992. These data sets can be found on the Society of Actuaries website under the following path:
https://www.soa.org/Research/Experience-Study/group-health/91-92-group-medical-claims.aspx. The $171'236$ claims recorded in both data sets are part of a larger database that includes the losses of 26 insurers over the period 1991-1992, see Grazier et al. [54] for more description on the data. We shall investigate the dependence between the hospital charges, corresponding to the loss variable X_i, and the Other charges corresponding to the ALAE variable Y_i associated to the settlement of X_i. The same data set was explored in Cebrian et al. [20] where claims occurring during accident year 1991 were considered. For our study, we work with claims occurring during accident year 1992. The sample comprises of $75'789$ claims. There are 4 different medical group plan types. Each policyholder belongs to one of these medical groups.
In the sequel, we consider the 1992 records relating to Plan type 4, the loss variable X_i exceeding 25'000 in order to have a positive dependence between the loss and the ALAE, and strictly positive ALAE. Some statistics on the data are presented in Table 5.4.1.

	Loss	ALAE
Min	25'003	5
Q1	30'859	7'775
Q2	40'985	14'111
Q3	64'067	23'547
Max	1'404'432	409'586
No. Obs.	5'106	5'106
Mean	62'589	20'001
Std. Dev.	69'539	24'130

Table 5.4.1: Statistics for Loss ALAE data from medical insurance.

The scatterplot (ALAE, loss) on a log scale is depicted in Figure 5.4.1.

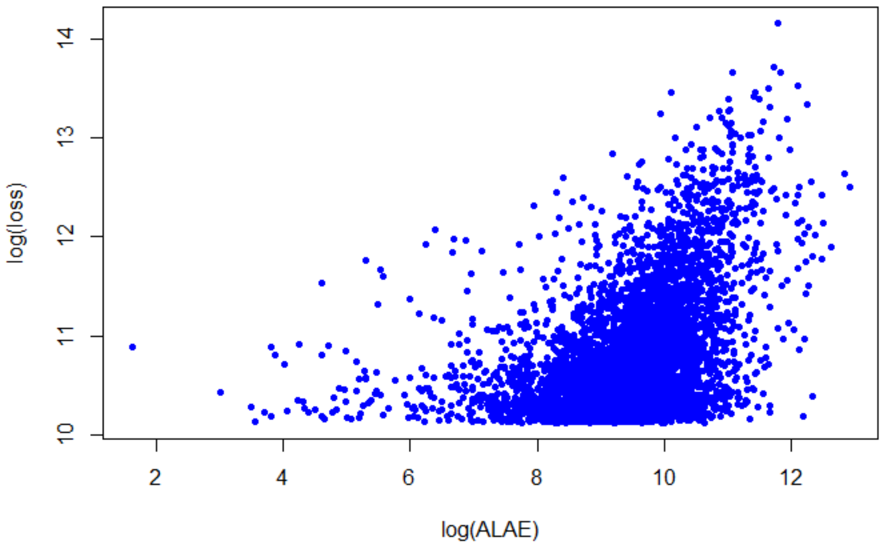


Figure 5.4.1: Scatterplot for log ALAE and log Loss.

Furthermore, we compute the empirical dependence measures between the losses and their respective ALAE as shown in Table 5.4.2 below. The latter indicates a positive dependence between these two variables with an empirical upper tail dependence of 0.3806.

Dependence measures	Values
Pearson's Correlation	0.4442
Spearman's Rho	0.4442
Kendall's Tau	0.3088
Upper tail dependence	0.3806

Table 5.4.2: Dependence measures for Loss ALAE data from medical insurance.

By maximizing (5.3.1), we get the estimators of the parameters Θ of the copula models summarized hereunder in Table 5.4.3. The table below includes as well the estimation of the parameters when Λ is either Geometric, Shifted Poisson or Truncated Poisson already considered in Hashorva et al. [60].

Original copula Q	**Distribution for N**	$\boldsymbol{\lambda}$	$\boldsymbol{\alpha}$	$\boldsymbol{\theta}$	$\boldsymbol{m}$
Gumbel	None	-	-	1.4328	-
	Geometric	0.9751	-	1.4262	-
	Truncated Poisson	0.0380	-	1.4278	-
	Shifted Poisson	23.751	-	1.4035	-
	Truncated discrete-stable	1.0243	0.6742	1.000	-
	Shifted discrete-stable	1	0.8261	1.2078	-
Frank	None	-	-	3.0482	-
	Geometric	0.9999	-	3.0481	-
	Truncated Poisson	0.0001	-	3.0481	-
	Shifted Poisson	0.0001	-	3.0481	-
	Truncated discrete-stable	1.0242	0.6742	0.0001	-
	Shifted discrete-stable	1.1190	0.6102	0.0013	-
Student	None	-	-	0.4576	11.893
	Geometric	0.9999	-	0.4576	11.892
	Truncated Poisson	0.0001	-	0.4576	11.890
	Shifted Poisson	0.0001	-	0.4576	11.890
	Truncated discrete-stable	1.1751	0.6899	0.0001	13.999
	Shifted discrete-stable	0.3598	0.6404	0.2912	9.9874
Joe	None	-	-	1.6440	-
	Geometric	0.5804	-	1.4477	-
	Truncated Poisson	1.0243	-	1.4832	-
	Shifted Poisson	0.5626	-	1.4945	-
	Truncated discrete-stable	1.0243	0.8787	1.3033	-
	Shifted discrete-stable	1.0000	0.8268	1.2890	-

Table 5.4.3: Parameter estimation for the different copula models

Following the estimation of the parameters, one is interested to assess the fit of these new copula models to the data set. The table below illustrates the results.

Original copula Q	Distribution for N	P-value	RMSE	AIC
Gumbel	None	0.755	0.0039	-1,371.21
	Geometric	0.740	0.0039	-1,369.28
	Truncated Poisson	0.745	0.0039	-1,369.27
	Shifted Poisson	0.527	0.0047	-1,328.14
	Truncated discrete-stable	-	0.0032	-1,384.87
	Shifted discrete-stable	0.267	0.0059	-1,326.77
Frank	None	0.021	0.0096	-1,137.12
	Geometric	0.026	0.0096	-1,135.09
	Truncated Poisson	0.021	0.0096	-1,135.10
	Shifted Poisson	0.017	0.0096	-1,135.12
	Truncated discrete-stable	-	0.0032	-1,384.86
	Shifted discrete-stable	-	0.0039	-1,282.69
Student	None	0.046	0.0090	-1,195.83
	Geometric	0.063	0.0090	-1,193.82
	Truncated Poisson	0.024	0.0090	-1,193.82
	Shifted Poisson	0.045	0.0090	-1,193.82
	Truncated discrete-stable	-	0.0031	-1,377.86
	Shifted discrete-stable	0.030	0.0065	-1,288.69
Joe	None	0.055	0.0039	-1,371.21
	Geometric	0.986	0.0027	-1,393.23
	Truncated Poisson	0.919	0.0032	-1,386.87
	Shifted Poisson	0.892	0.0034	-1,384.34
	Truncated discrete-stable	-	0.0032	-1,384.87
	Shifted discrete-stable	0.799	0.0740	-1,370.37

Table 5.4.4: P-Values, RMSE and AIC values for the different copula models

Table 5.4.4 shows that

- based on the P-values, the family of Gumbel and Joe copula are accepted at a significance level of 10% with the exception of Joe copula having a P-value of 5.5%. Clearly, the family of Frank and Student copula do not represent a good fit to the data due to a P-value smaller than 5%, however the Student Geometric copula model is accepted at a significance level of 5%,

- based on the RMSE, the Joe Geometric copula outperforms the other models having the smallest RMSE,

- and finally based on the AIC criteria, the Joe Geometric copula is the model that best fits the data followed by the Joe Truncated Poisson, Joe truncated discrete-stable and Gumbel truncated discrete-stable copulas.

5.4.2 Worker's compensation insurance data

This data set examines the losses due to permanent and partial disability of the worker's compensation line of business. In this data, we model the dependence between the pure premium P, defined as the loss due to partial and permanent disability per dollar of payroll, and the payroll $PayR$. The same data was used in Zhang and Lin [127] and Frees et al. [43]. In order to reproduce the fit of the Geometric mixture copula developed by Zhang and Lin [127] we use the same estimation procedure. Therefore, the losses and payrolls are transformed to a logarithmic scale such that $X = \ln P$ and $Y = \ln PayR$. Also, before replacing the marginals by their empirical distributions for the PML estimation approach, we smooth them by using the Gaussian non parametric kernel smoothing method defined for both components as follows (see e.g., Hansen [57])

$$\widehat{F_X}(x_i) = \frac{1}{n}\sum_{i=1}^{n}\Phi\Big(\frac{x_i - x_j}{h}\Big) \text{ with } h = 0.2605, \quad \widehat{F_Y}(y_i) = \frac{1}{n}\sum_{i=1}^{n}\Phi\Big(\frac{y_i - y_j}{h}\Big) \text{ with } h = 0.1290.$$

The empirical dependence measures are summarized in Table 5.4.5 below.

Dependence measures	Values
Pearson's Correlation	0.8194
Spearman's Rho	0.8181
Kendall's Tau	0.6306
Upper tail dependence	0.6627

Table 5.4.5: Dependence measures between P and $PayR$.

Table 5.4.6 gathers the estimation results from maximising (5.3.1). The parameters are estimated for all copula models including the ones considered in Hashorva et al. [60] namely the case where Λ is either Geometric, Poisson or Truncated Poisson.

Original copula Q	**Distribution for** N	$\boldsymbol{\lambda}$	$\boldsymbol{\alpha}$	$\boldsymbol{\theta}$	$\boldsymbol{m}$
Gumbel	None	-	-	2.4238	-
	Geometric	0.3013	-	2.0356	-
	Truncated Poisson	1.5530	-	2.2109	-
	Shifted Poisson	4.1778	-	2.2193	-
	Truncated discrete-stable	1.6537	0.9999	2.1778	-
	Shifted discrete-stable	0.9458	0.9749	2.1801	-
Frank	None	-	-	8.7858	-
	Geometric	0.9267	-	10.0930	-
	Truncated Poisson	0.0001	-	8.7857	-
	Shifted Poisson	0.0001	-	8.6998	-
	Truncated discrete-stable	1.6553	0.4523	3.1376	-
	Shifted discrete-stable	0.0001	0.7991	8.6998	-
Student	None	-	-	0.8117	12.825
	Geometric	0.9999	-	0.8117	12.826
	Truncated Poisson	1.0000	-	0.8015	12.999
	Shifted Poisson	1.0000	-	0.8026	12.924
	Truncated discrete-stable	0.0887	0.7980	0.7673	14.999
	Shifted discrete-stable	0.0003	0.8000	0.8081	8.6620
Joe	None	-	-	2.8596	-
	Geometric	0.1299	-	2.0409	-
	Truncated Poisson	10.1113	-	2.4120	-
	Shifted Poisson	6.7874	-	2.3976	-
	Truncated discrete-stable	8.3546	0.9999	2.3752	-
	Shifted discrete-stable	7.2456	0.9467	2.2790	-

Table 5.4.6: Parameter estimation for the different copula models

Following the estimation of the parameters, Table 5.4.7 highlights the results from the Goodness of fit test along with the RMSE and AIC criteria for each copula model.

Original copula Q	Distribution for N	P-value	RMSE	AIC
Gumbel	None	0.474	0.0127	-784.96
	Geometric	0.389	0.0134	-808.94
	Truncated Poisson	0.423	0.0130	-798.90
	Shifted Poisson	0.393	0.0140	-779.87
	Truncated discrete-stable	0.410	0.0135	-788.39
	Shifted discrete-stable	0.403	0.0134	-786.29
Frank	None	0.229	0.0158	-846.57
	Geometric	0.184	0.0174	-883.23
	Truncated Poisson	0.261	0.0158	-846.56
	Shifted Poisson	0.239	0.0158	-839.04
	Truncated discrete-stable	0.154	0.0152	-774.30
	Shifted discrete-stable	0.248	0.0158	-837.04
Student	None	0.313	0.0150	-803.80
	Geometric	0.294	0.0150	-801.80
	Truncated Poisson	0.215	0.0164	-789.33
	Shifted Poisson	0.181	0.0175	-775.52
	Truncated discrete-stable	-	0.0143	-800.15
	Shifted discrete-stable	0.281	0.0152	-795.50
Joe	None	0.051	0.0241	-640.19
	Geometric	0.416	0.0132	-811.03
	Truncated Poisson	0.469	0.0128	-783.24
	Shifted Poisson	0.445	0.0133	-772.27
	Truncated discrete-stable	0.430	0.0134	-771.43
	Shifted discrete-stable	0.432	0.1384	-771.99

Table 5.4.7: P-Values, RMSE and AIC values for the different copula models

Table 5.4.7 shows that

- based on the P-values, all models are accepted at a significance level of 10% with the exception of Joe copula having a P-value of 5.1%,
- based on the RMSE, the Gumbel copula followed by the Joe Truncated Poisson copula represent the best fit for the data as they have the smallest RMSE,
- and finally based on the AIC criteria, the family of Frank mixture copulas outperforms the others as for the majority of these copula models, the AIC is the smallest.

Joe copula should not be used to model this data as it has the lowest P-value (5.1%) and the greatest RMSE (0.0241) and AIC (-640.19) among all copula models. Moreover, Table 5.4.7 shows that for the majority of the models an increase in the P-value is associated with a decrease in the RMSE.

5.4.3 Danish fire insurance data

In this section, we shall consider the Danish data set collected from the Copenhagen Reinsurance Company which describes the fire insurance claims observed over the period 1980-1990. This data set is available on the following website `www.ma.hw.ac.uk/~mcneil/`. It comprises of $n = 2'167$ fire losses based on three components: buildings, content and profit. However, in the sequel, we shall analyze the dependency between the first two components. Let X_i, Y_i be the i^{th} loss observed for both components respectively. For more description on the data, we refer to Haug et al. [62]. Table 5.4.8 below displays the estimated parameters for each family of copula obtained when maximising the pseudo log-likelihood function defined in (5.3.1).

	Original		Shifted discrete-stable (λ, α)				Truncated discrete-stable (λ, α)			
	θ	m	λ	α	θ	m	λ	α	θ	m
Gumbel (θ)	1.1762	-	0.0001	0.9395	1.1762	-	10.000	0.8508	1	-
Frank (θ)	0.8807	-	0.2879	0.5000	0.0001	-	10.000	0.8508	0.0001	-
Student (θ, m)	0.1574	9.59	0.3498	0.6079	0.0001	3.99	10.000	0.9761	0.0001	1.99
Joe (θ)	1.3585	-	0.0001	0.9916	1.3581	-	10.000	0.9093	1.0687	-

Table 5.4.8: Parameter estimation for the different copula models

The table below summarizes the relevant measures relative to each copula model.

Original copula Q	Distribution for N	P-value	RMSE	AIC
Gumbel	None	0.000	0.0248	-133.18
	Geometric	0.003	0.0248	-131.17
	Truncated Poisson	0.001	0.0248	-131.18
	Shifted Poisson	0.001	0.0248	-131.17
	Truncated discrete-stable	-	0.0282	-63.64
	Shifted discrete-stable	0.000	0.0248	-129.28
Frank	None	0.000	0.0264	-29.12
	Geometric	0.001	0.0264	-27.12
	Truncated Poisson	0.000	0.0264	-27.12
	Shifted Poisson	0.000	0.0264	-27.12
	Truncated discrete-stable	-	0.0282	-63.64
	Shifted discrete-stable	-	0.0216	-175.81
Student	None	0.000	0.0266	-47.86
	Geometric	0.000	0.0266	-45.84
	Truncated Poisson	0.000	0.0266	-45.81
	Shifted Poisson	0.000	0.0266	-45.42
	Truncated discrete-stable	-	0.0310	-288.44
	Shifted discrete-stable	0.009	0.0221	-172.98
Joe	None	0.002	0.0224	-204.85
	Geometric	0.003	0.0224	-202.83
	Truncated Poisson	0.001	0.0224	-202.84
	Shifted Poisson	0.007	0.0224	-202.83
	Truncated discrete-stable	-	0.0282	-63.64
	Shifted discrete-stable	0.007	0.0224	-200.84

Table 5.4.9: P-Values, RMSE and AIC values for the different copula models

For this data set, Table 5.4.9 shows that

- based on the P-values, all models are rejected at a significance level of 1%,
- based on the RMSE, the Frank shifted discrete-stable copula models best the data with the lowest error,
- and finally based on the AIC, the copulas that best fit the data are the Student truncated discrete-stable copula followed by the Joe copula.

However, it is clear that, for this data set, the above models do not describe well the dependence of the maximum claim amounts and this is mainly explained by a low dependence level between the two components, see Hashorva et al. [60].

5.4.4 Loss ALAE from general liability insurance

We use the data set available in R collected by the Insurance Services Office that examines the losses and their respective ALAE of a general liability insurance portfolio. For more description on the data, we refer to Denuit et al. [28]. The data set comprises of $n = 1'500$ claims from which 34 claims were censored. In the sequel, X_i represents the i^{th} loss observed and Y_i the corresponding ALAE. Each loss X_i is associated with a policy limit ℓ_i. Typically, if X_i exceeds the policy limit ℓ_i, the observed loss corresponds to ℓ_i, i.e., the exact amount of the loss is unknown. In this respect, we define the indicator function δ_i as follows

$$\delta_i = \begin{cases} 1 & \text{if} \quad X_i \leqslant \ell_i, \\ 0 & \text{if} \quad X_i > \ell_i, i = 1, \ldots, n. \end{cases}$$

To estimate the parameters of the new copula, we shall maximise the pseudo log-likelihood function $l(\Theta)$ defined in (5.3.1). Typically, U_i and V_i are the pseudo-observations of the variables X_i and Y_i respectively as defined in Section 5.3.1. However, given that this data set is right-censored for the loss variable X_i, the marginal of the latter, i.e. $\widetilde{G_1}(x)$, shall be approximated by the Kaplan Meier estimator. Thus, the resulting pseudo log-likelihood function is given by

$$l(\Theta) = \sum_{i=1}^{n} \left(\delta_i \ln c_\Theta(u_i, v_i) + (1 - \delta_i) \ln \left(1 - \frac{\partial C_\Theta(u_i, v_i)}{\partial v} \right) \right). \tag{5.4.1}$$

We shall estimate the parameter Θ of the new copula by maximizing (5.4.1). The table below describes the estimated parameters for the different families of copulas.

	Original		Shifted discrete-stable (λ, α)				Truncated discrete-stable (λ, α)			
	θ	m	λ	α	θ	m	λ	α	θ	m
Gumbel (θ)	1.4284	-	0.2117	0.8916	1.3613	-	0.7733	0.8427	1.1950	-
Frank (θ)	3.0440	-	0.7146	0.7000	1.2887	-	0.7206	0.7002	0.6865	-
Student (θ, m)	0.4642	10.00	0.5950	0.6502	0.2212	6.99	1.5783	0.7141	0.0001	8.99
Joe (θ)	1.6183	-	0.7333	0.9999	1.4530	-	1.3592	0.7881	1.1294	-

Table 5.4.10: Parameter estimation for the different copula models

Following the estimation of the parameters, one is interested in assessing the fit of those models to the general liability data set. The table below highlights the P-values, RMSE and AIC criteria for each model. These models are then compared to the ones observed in Hashorva et al. [60], i.e., the case where Λ is either Geometric, Shifted Poisson or Truncated Poisson.

Original copula Q	Distribution for N	P-value	RMSE	AIC
Gumbel	None	0.967	0.0055	-210.18
	Geometric	0.482	0.0088	-278.23
	Truncated Poisson	0.939	0.0059	-360.49
	Shifted Poisson	0.908	0.0062	-361.20
	Truncated discrete-stable	-	0.0057	-207.28
	Shifted discrete-stable	0.946	0.0058	-206.92
Frank	None	0.304	0.0106	-321.44
	Geometric	0.185	0.0119	-174.40
	Truncated Poisson	0.276	0.0105	-306.40
	Shifted Poisson	0.297	0.0105	-306.41
	Truncated discrete-stable	-	0.0059	-206.99
	Shifted discrete-stable	0.721	0.0076	-194.84
Student	None	0.456	0.0089	-180.99
	Geometric	0.328	0.0101	-228.82
	Truncated Poisson	0.279	0.0107	-271.40
	Shifted Poisson	0.223	0.0115	-295.42
	Truncated discrete-stable	-	0.0060	-209.58
	Shifted discrete-stable	0.489	0.0088	-181.51
Joe	None	0.565	0.0055	-179.00
	Geometric	0.643	0.0080	-292.41
	Truncated Poisson	0.702	0.0077	-331.21
	Shifted Poisson	0.936	0.0058	-361.76
	Truncated discrete-stable	-	0.0059	-206.87
	Shifted discrete-stable	0.975	0.0434	-206.01

Table 5.4.11: P-Values, RMSE and AIC values for the different copula models

Table 5.4.11 shows that

- based on the P-values, all models are accepted at a significance level of 10%,
- based on the RMSE, the Joe Shifted discrete-stable copula outperforms the others having the smallest RMSE followed by the Gumbel and Joe shifted Poisson copulas,
- and finally based on the AIC criteria, the Frank shifted discrete-stable copula is the model that best fits the data, having the smallest AIC among all other models.

5.5 Proofs

5.5.1 Derivation of (5.3.2)-(5.3.3)

We derive first (5.3.2). Λ follows a shifted discrete-stable distribution with Laplace transform defined in (5.1.11). In light of (5.1.3), we compute the 1^{st} and 2^{nd} derivatives of (5.1.11) with respect to t

$$\begin{aligned} L_\Lambda^{'}(t) &= -e^{-t}e^{-\lambda(1-e^{-t})^\alpha}\Big(1+\lambda\alpha e^{-t}(1-e^{-t})^{\alpha-1}\Big), \qquad (5.5.1)\\ L_\Lambda^{''}(t) &= e^{-t}e^{-\lambda(1-e^{-t})^\alpha}\Big(1+3\lambda\alpha e^{-t}(1-e^{-t})^{\alpha-1}+\lambda^2\alpha^2e^{-2t}(1-e^{-t})^{2\alpha-2}+\lambda\alpha e^{-2t}(1-e^{-t})^{\alpha-2}\\ &\quad -\lambda\alpha^2e^{-2t}(1-e^{-t})^{\alpha-2}\Big). \end{aligned}$$

By setting $t=-\ln Q(v_1,v_2)$ in (5.5.1) with $v_i=e^{-L_\Lambda^{-1}(u_i)}$ for $i=1,2$, $c_\Theta(u_1,u_2)$ defined in (5.1.3) is given by

$$\begin{aligned} c_\Theta(u_1,u_2) &= \frac{e^{\lambda[(1-v_1)^\alpha+(1-v_2)^\alpha-(1-Q(v_1,v_2))^\alpha]}}{(1+\lambda\alpha v_1(1-v_1)^{\alpha-1})(1+\lambda\alpha v_2(1-v_2)^{\alpha-1})}\Big[(2\lambda\alpha(1-Q(v_1,v_2))^{\alpha-1}\\ &\quad +\lambda^2\alpha^2Q(v_1,v_2)(1-Q(v_1,v_2))^{2\alpha-2}+\lambda\alpha Q(v_1,v_2)(1-Q(v_1,v_2))^{\alpha-2}\\ &\quad -\lambda\alpha^2Q(v_1,v_2)(1-Q(v_1,v_2))^{\alpha-2})\frac{\partial Q(v_1,v_2)}{\partial v_1}\frac{\partial Q(v_1,v_2)}{\partial v_2}\\ &\quad +(1+\lambda\alpha Q(v_1,v_2)(1-Q(v_1,v_2))^{\alpha-1})\frac{\partial^2Q(v_1,v_2)}{\partial v_1\partial v_2}\Big]\\ &= \frac{e^{\lambda[(1-v_1)^\alpha+(1-v_2)^\alpha-(1-Q(v_1,v_2))^\alpha]}}{(1+\lambda\alpha v_1(1-v_1)^{\alpha-1})(1+\lambda\alpha v_2(1-v_2)^{\alpha-1})}\Big[\lambda\alpha(1-Q(v_1,v_2))^{\alpha-2}\\ &\quad \times(2-(\alpha+1)Q(v_1,v_2)+\lambda\alpha Q(v_1,v_2)(1-Q(v_1,v_2))^\alpha)\frac{\partial Q(v_1,v_2)}{\partial v_1}\frac{\partial Q(v_1,v_2)}{\partial v_2}\\ &\quad +(1+\lambda\alpha Q(v_1,v_2)(1-Q(v_1,v_2))^{\alpha-1})\frac{\partial^2Q(v_1,v_2)}{\partial v_1\partial v_2}\Big], \end{aligned}$$

where for $i=1,2$

$$u_i=v_ie^{-\lambda(1-v_i)^\alpha} \quad \text{and} \quad \frac{\partial v_i}{\partial u_i}=\frac{e^{\lambda(1-v_i)^\alpha}}{1+\lambda\alpha v_i(1-v_i)^{\alpha-1}}.$$

Next, we show (5.3.3). Hereafter, Λ follows a truncated discrete-stable distribution. Its Laplace transform is defined in (5.1.13) and the corresponding first and second derivatives of (5.1.13) are given by

$$
\begin{aligned}
L_{\Lambda}^{'}(t) &= \frac{-\lambda\alpha e^{-\lambda(1-e^{-t})^{\alpha}} e^{-t}(1-e^{-t})^{\alpha-1}}{1-e^{-\lambda}}, \\
L_{\Lambda}^{''}(t) &= \frac{\lambda\alpha e^{-\lambda(1-e^{-t})^{\alpha}} e^{-t}(1-e^{-t})^{\alpha-2}}{1-e^{-\lambda}}\Big(1-\alpha e^{-t}+\lambda\alpha e^{-t}(1-e^{-t})^{\alpha}\Big).
\end{aligned} \tag{5.5.2}
$$

By replacing t in (5.5.2) with $-\ln Q(v_1, v_2)$, we show that (5.1.3) is given by

$$
\begin{aligned}
c_{\Theta}(u_1, u_2) &= \Big(\frac{1-e^{-\lambda}}{\lambda\alpha}\Big)^2 \frac{e^{\lambda[(1-v_1)^{\alpha}+(1-v_2)^{\alpha}]}}{(1-v_1)^{\alpha-1}(1-v_2)^{\alpha-1}}\Big(\frac{\lambda\alpha}{1-e^{-\lambda}}\Big)e^{-\lambda(1-Q(v_1,v_2))^{\alpha}}(1-Q(v_1,v_2))^{\alpha-2} \\
&\quad \times\Big[(1-\alpha+\lambda\alpha(1-Q(v_1,v_2))^{\alpha})\frac{\partial Q(v_1,v_2)}{\partial v_1}\frac{\partial Q(v_1,v_2)}{\partial v_2} \\
&\quad +(1-Q(v_1,v_2))\frac{\partial^2 Q(v_1,v_2)}{\partial v_1 \partial v_2}\Big] \\
&= \frac{1-e^{-\lambda}}{\lambda\alpha}\frac{e^{\lambda[(1-v_1)^{\alpha}+(1-v_2)^{\alpha}-(1-Q(v_1,v_2))^{\alpha}]}}{(1-v_1)^{\alpha-1}(1-v_2)^{\alpha-1}}(1-Q(v_1,v_2))^{\alpha-2} \\
&\quad \Big[(1-\alpha+\lambda\alpha(1-Q(v_1,v_2))^{\alpha})\frac{\partial Q(v_1,v_2)}{\partial v_1}\frac{\partial Q(v_1,v_2)}{\partial v_2} \\
&\quad +(1-Q(v_1,v_2))\frac{\partial^2 Q(v_1,v_2)}{\partial v_1 \partial v_2}\Big],
\end{aligned}
$$

where for $i = 1, 2$

$$
v_i = 1 - \Big(\frac{-\ln[e^{-\lambda}+u_i(1-e^{-\lambda})]}{\lambda}\Big)^{\frac{1}{\alpha}} \quad \text{and} \quad \frac{\partial v_i}{\partial u_i} = \frac{1-e^{-\lambda}}{\lambda\alpha}\frac{e^{\lambda(1-v_i)^{\alpha}}}{(1-v_i)^{\alpha-1}}.
$$

Chapter 6

On some multivariate Sarmanov mixed Erlang reinsurance risks: Aggregation and Capital allocation

This chapter is based on G. Ratovomirija, M. Tamraz and R. Vernic: On some multivariate Sarmanov mixed Erlang reinsurance risks: Aggregation and Capital allocation, published in *Insurance Mathematics and Economics*, 74:197-209, 2017.

6.1 Introduction

Modern risk management usually involves complex dependent risk factors. In this respect, several regulations were put in place in order to assess the minimum capital requirement, namely the Economic Capital (EC) that insurance and reinsurance companies are constrained to hold according to their risk exposures. In practice, the EC is evaluated by means of risk measures on the aggregated risk, so that the companies will be covered from unexpected large losses.
For instance, the EC under the Solvency II framework for EU countries focuses on a Value-at-Risk (VaR) approach at a tolerance level of 99.5% of the aggregated risk over a one year period, while in Switzerland, the EC under the Swiss Solvency Test (SST) is based on the Tail-Value-at-Risk (TVaR) approach at a 99% confidence level of the aggregated risk over a one year period. Since the EC quantified in the latter reflects the aggregate capital needed to cover the entire loss of a company, it is also of interest to study how this capital should be allocated among the different risk factors (e.g., lines of business) in the insurance and reinsurance companies, in other words, how much amount of capital each individual risk contributes to the aggregated EC. This allows the risk managers to identify and monitor conveniently their

risks. An extensive literature has been developed on capital allocation techniques from which we shall restrict to the TVaR method; see Dhaene et al. [32], Tasche [114] and the references therein for more details.

Therefore, the main task of actuaries is to choose an appropriate model for the multivariate risk factors, namely the dependence structure model and the distributions of the marginals. The aim of this contribution is to address risk aggregation and TVaR capital allocation for insurance and reinsurance mixed Erlang risks whose dependency is governed by the Sarmanov distribution with a certain expression of the kernel functions. This study comes along the lines of some recent contributions: Vernic [117] considered capital allocation based on the TVaR rule for the Sarmanov distribution with exponential marginals; Cossette et al. [23] used the Farlie-Gumbel-Morgenstern (FGM) distribution to model the dependency between mixed Erlang distributed risks and applied it to capital allocation and risk aggregation; Hashorva and Ratovomirija [59] and Ratovomirija [100] presented aggregation and capital allocation in insurance and reinsurance for mixed Erlang distributed risks joined by the Sarmanov distribution with a specific kernel function different from the one considered in this study. Note that the choice of the Sarmanov and mixed Erlang distributions is not incidental, these distributions gained a lot of interest in the actuarial literature lately: for the Sarmanov distribution, see e.g., Yang and Hashorva [125], Hernández-Bastida and Fernández-Sánchez [63], Abdallah et al. [1], Vernic [118], while for the mixed Erlang distribution we refer to Lee and Lin [78], Willmot and Lin [121], Lee and Lin [79] or Willmot and Woo [122]. One key advantage of the Sarmanov distribution is its flexibility to join different types of marginals and its allowance to obtain exact results. An interesting property of the mixed Erlang distributions is the fact that many risk related quantities, such as TVaR, have an analytical form.

This paper is organized as follows: in the second section, we present some preliminaries on the Sarmanov distribution, on the TVaR capital allocation problem and on the mixed Erlang distribution, supplemented with several lemmas on this last distribution that will be needed for the proofs of the main results. Section 3 contains the main results on risk aggregation and capital allocation for the stop-loss reinsurance, which are also particularized in the case without stop-loss reinsurance; the main formulas of this section are illustrated with some numerical examples. The paper ends with two appendices: the first one discusses and compares the dependence structure of the bivariate Sarmanov distribution with mixed Erlang marginals and different kernel functions, providing upper and lower bounds for the corresponding Pearson correlation coefficient, while the second appendix contains all the proofs of the theoretical results.

6.2 Preliminaries

Throughout the paper, by convention, we assume that all the involved quantities exist (i.e., expectations, variances, covariances etc.).

6.2.1 Multivariate Sarmanov distribution

The Sarmanov distribution caught the interest of many researchers in different fields. It was first introduced by Sarmanov [107] in the bivariate case, then extended by Lee [77] to the multivariate case. Its applications in many insurance contexts show its flexible structure when modeling the dependence between multivariate risks given the distribution of the marginals, see the references cited above.
According to Sarmanov [107], the joint probability density function (pdf) of a bivariate Sarmanov distribution is defined as follows

$$h(x_1, x_2) = f_1(x_1)f_2(x_2)(1 + \alpha_{1,2}\phi_1(x_1)\phi_2(x_2)), x_1, x_2 \in \mathbb{R}, \tag{6.2.1}$$

where for $i = 1, 2, f_i$ are the densities of the marginals, ϕ_i are kernel functions assumed to be bounded, non-constant, and $\alpha_{1,2}$ is a real number satisfying the following conditions

$$\mathbb{E}(\phi_i(X_i)) = 0, i = 1, 2, \quad 1 + \alpha_{1,2}\phi_1(x_1)\phi_2(x_2) \geq 0, \forall x_1, x_2 \in \mathbb{R}.$$

Lee [77] introduced general methods for the choice of ϕ. Yang and Hashorva [125] considered the case where ϕ depends on some function g, being expressed as follows

$$\phi(x) = g(x) - \mathbb{E}(g(X)), \text{ where } \mathbb{E}(g(X)) < \infty.$$

In the context of risk aggregation and capital allocation, Hashorva and Ratovomirija [59] assumed that $g(x) = e^{-x}$, Vernic [117] studied the case where the marginals are exponentially distributed, while Cossette et al. [23] used the FGM distribution with mixed Erlang marginals (the FGM is a special case of the Sarmanov distribution for $g(x) = 2(1 - F(x))$, with F denoting the distribution function of the marginal).

Thus, in the sequel, we consider the following kernel function

$$\phi_i(x_i) = f_i(x_i) - \mathbb{E}(f_i(X_i)), \tag{6.2.2}$$

where f_i are such that $\mathbb{E}(f_i(X_i)) < \infty, \forall i$. In this case, the range of $\alpha_{1,2}$ is given by

$$\frac{-1}{\max\{\gamma_1\gamma_2, (M_1-\gamma_1)(M_2-\gamma_2)\}} \leqslant \alpha_{1,2} \leqslant \frac{1}{\max\{\gamma_1(M_2-\gamma_2), (M_1-\gamma_1)\gamma_2\}}, \tag{6.2.3}$$

where $\gamma_i = \mathbb{E}(f_i(X_i))$ and $M_i = \max_{x\in\mathbb{R}} f_i(x), i = 1, 2$. Moreover, we shall work with a generalization of the above distribution to the multivariate case, see Lee [77]. Hereafter we let $\boldsymbol{X} = (X_1, \ldots, X_n)$ denote an n-dimensional random vector, $\boldsymbol{x} = (x_1, \ldots, x_n)$ (e.g., the observations on $\boldsymbol{X}$) and we let $I_n = 1, \ldots, n$. Therefore, we shall model the dependency between the risks X_i having pdf $f_i, i \in I_n$, via the multivariate Sarmanov distribution having the following pdf

$$h(\boldsymbol{x}) = \prod_{i=1}^{n} f_i(x_i)\left(1 + \sum_{1\leq j<l\leq n} \alpha_{j,l}\phi_j(x_j)\phi_l(x_l)\right), \boldsymbol{x} \in \mathbb{R}^n, \tag{6.2.4}$$

where ϕ_i are the non-constant kernel functions defined in (6.2.2) and $\alpha_{j,l}$ are real numbers satisfying the condition

$$1 + \sum_{1\leq j<l\leq n} \alpha_{j,l}\phi_j(x_j)\phi_l(x_l) \geq 0. \tag{6.2.5}$$

Remarks 6.2.1. *It should be noted that a more general expression of the Sarmanov density for the multivariate case can be written as follows*

$$h(\boldsymbol{x}) = \prod_{i=1}^{n} f_i(x_i)\left(1 + \sum_{l=2}^{n} \sum_{1\leq j_1<j_2<\ldots<j_l\leq n} \alpha_{j_1,\ldots,j_l} \prod_{k=1}^{l} \phi_{j_k}(x_{j_k})\right), \boldsymbol{x} \in \mathbb{R}^n, \tag{6.2.6}$$

such that $\mathbb{E}(\phi_i(X_i)) = 0$ *and* $1 + \sum_{l=2}^{n}\sum_{1\leq j_1<j_2<\ldots<j_l\leq n} \alpha_{j_1,\ldots,j_l}\prod_{k=1}^{l}\phi_{j_k}(x_{j_k}) \geq 0$. *However, (6.2.6) requires the estimation of all the dependence parameters, which is in general very complex. Thus, it is often assumed that* $\alpha_{j_1,\ldots,j_l} = 0$ *for* $l \geq 3$, *see Mari and Kotz [83]. For simplicity, in this paper, we consider the Sarmanov density defined in (6.2.4).*

6.2.2 Mixed Erlang distributions

The mixed Erlang distribution has many attractive distributional properties when modeling the claim sizes of an insurance portfolio (see, e.g., Willmot and Lin [121]) and the dependence between multivariate insurance risks, see Lee and Lin [79]. Actually, during these past few years, modeling the dependence of multivariate mixed Erlang risks raised the interest of many researchers, see also Cossette et al. [23], Hashorva and Ratovomirija [59] or Ratovomirija [100].

In this regard, we define the pdf of a mixed Erlang distribution denoted $ME(\beta, \underline{Q})$ by

$$f(x, \beta, \underline{Q}) = \sum_{k=1}^{\infty} q_k w_k(x, \beta), x \geq 0, \tag{6.2.7}$$

where $w_k(x, \beta) = \frac{\beta^k x^{k-1} e^{-\beta x}}{(k-1)!}$ is the pdf of an Erlang distribution with $\beta > 0$ the scale parameter, $k \in \mathbb{N}^*$ the shape parameter and $\underline{Q} = (q_1, q_2, \ldots)$ is a vector of non-negative mixing probabilities such that $\sum_{k=1}^{\infty} q_k = 1$. We denote by W_k the distribution function (df) of the Erlang distribution and by $\overline{W}_k$ its corresponding survival (tail) function given, respectively, by

$$W_k(x, \beta) = 1 - e^{-\beta x} \sum_{j=0}^{k-1} \frac{(\beta x)^j}{j!}, \quad \overline{W}_k(x, \beta) = e^{-\beta x} \sum_{j=0}^{k-1} \frac{(\beta x)^j}{j!}, x \geq 0.$$

Thus, the mixed Erlang df can be expressed in terms of the Erlang df as follows

$$F(x, \beta, \underline{Q}) = \sum_{k=1}^{\infty} q_k W_k(x, \beta) = 1 - e^{-\beta x} \sum_{k=1}^{\infty} q_k \sum_{j=0}^{k-1} \frac{(\beta x)^j}{j!}, x \geq 0. \tag{6.2.8}$$

Moreover, the expected value of this distribution is $\mu = \frac{1}{\beta} \sum_{k=1}^{\infty} k q_k$.
In addition, we present some distributional properties and useful results for the mixed Erlang distributions.

Lemma 6.2.2. *Let $X \sim ME(\beta, \underline{Q})$ with pdf $f(x, \beta, \underline{Q})$ and $\mathbb{E}(f(X, \beta, \underline{Q})) < \infty$. Then $c(x, \beta, \underline{Q}) := \frac{f(x,\beta,\underline{Q})^2}{\mathbb{E}\left(f(X,\beta,\underline{Q})\right)}$ is again a pdf of a mixed Erlang distribution with mixing probabilities $V(\underline{Q}) = (v_1, v_2, \ldots)$ and scale parameter 2β, i.e., we have*

$$c(x, \beta, \underline{Q}) = \sum_{k=1}^{\infty} v_k w_k(x, 2\beta) = f\left(x, 2\beta, V(\underline{Q})\right),$$

where

$$v_k = \frac{\sum_{i=1}^{k}\begin{pmatrix}k-1\\ i-1\end{pmatrix}\frac{q_i q_{k+1-i}}{2^k}}{\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\begin{pmatrix}i+j-2\\ i-1\end{pmatrix}\frac{q_i q_j}{2^{i+j-1}}}. \tag{6.2.9}$$

The proof of the above lemma is given in the Appendix 6.5. We shall use the notation $\tilde{\mu}$ for the expected value corresponding to the pdf $c(\cdot, \beta, \underline{Q})$ defined in this lemma, i.e.,

$$\tilde{\mu} = \frac{1}{2\beta}\sum_{k=1}^{\infty} k v_k. \tag{6.2.10}$$

The following results have already been developed in Cossette et al. [23] and Ratovomirija [100].

Lemma 6.2.3. *Let $X \sim ME(\beta, \underline{Q})$ with pdf $f(x, \beta, \underline{Q})$ and $\mathbb{E}(X) < \infty$. Then $f^G(x, \beta, \underline{Q}) := \frac{xf(x,\beta,\underline{Q})}{\mathbb{E}(X)}$ is equal to the pdf $f(x, \beta, G(\underline{Q}))$ of a mixed Erlang distribution with mixing probabilities $G(\underline{Q}) = (g_1, g_2, \ldots)$ given by*

$$g_k = \begin{cases} 0 & \text{if } k = 1 \\ \frac{(k-1)q_{k-1}}{\sum_{j=1}^{\infty} j q_j} & \text{if } k = 2, 3, \ldots \end{cases}.$$

Lemma 6.2.4. *Let $X \sim ME(\beta_1, \underline{Q})$. Then it follows that for any positive constant β_2 such that $\beta_2 \geq \beta_1$, we have $X \sim ME(\beta_2, \Psi(\underline{Q}))$, where the elements of $\Psi(\underline{Q}) = (\psi_1, \psi_2, \ldots)$ are given by*

$$\psi_k = \sum_{i=1}^{k} q_i \begin{pmatrix}k-1\\ k-i\end{pmatrix}\left(\frac{\beta_1}{\beta_2}\right)^i\left(1-\frac{\beta_1}{\beta_2}\right)^{k-i}, k \geq 1.$$

Lemma 6.2.5. *Let X_1, X_2 be two independent mixed Erlang random variables (r.v.s) such that $X_i \sim ME(\beta, \underline{Q}_i), i = 1, 2$. Then $S_2 := X_1 + X_2 \sim ME\left(\beta, \underline{\Pi}\left(\underline{Q}_1, \underline{Q}_2\right)\right)$ with the components of $\underline{\Pi}\left(\underline{Q}_1, \underline{Q}_2\right)$ given by*

$$\pi_l\left(\underline{Q}_1, \underline{Q}_2\right) = \begin{cases} 0 & \text{for } l = 1 \\ \sum_{j=1}^{l-1} q_{1,j} q_{2,l-j} & \text{for } l > 1 \end{cases}.$$

Remarks 6.2.6. *According to Remark 2.1 in Cossette et al. [22], the result in Lemma 6.2.5 can be extended to* $S_n := \sum_{i=1}^n X_i$*, given that* $X_1, \ldots, X_n$ *are independent r.v.s and* $X_i \sim ME(\beta, \underline{Q}_i)$ *for* $i \in I_n$*. Thus,* $S_n \sim ME\left(\beta, \underline{\Pi}\left(\underline{Q}_1, \ldots, \underline{Q}_n\right)\right)$*, where the mixing weights are determined iteratively as follows*

$$\pi_l\left(\underline{Q}_1, \ldots, \underline{Q}_{n+1}\right) = \begin{cases} 0 & \text{for} \quad l = 1, \ldots, n \\ \sum_{j=n}^{l-1} \pi_j\left(\underline{Q}_1, \ldots, \underline{Q}_n\right) q_{n+1,l-j} & \text{for} \quad l = n+1, n+2, \ldots \end{cases}.$$

Lemma 6.2.7. *Given* $d > 0$ *and the r.v.* $X \sim ME(\beta, \underline{Q})$ *, the df of* $Y := (X - d)_+$ *can be expressed as*

$$F_Y(y) = F_X(y + d) = F_X(d) + H_X(y, d), y \geq 0, \tag{6.2.11}$$

where

$$H_X(y, d) := \mathbb{P}(0 < Y \leq y) = \sum_{k=0}^{\infty} \Delta_k\left(d, \beta, \underline{Q}\right) W_{k+1}(y, \beta),$$

with

$$\Delta_k\left(d, \beta, \underline{Q}\right) = \beta^{-1} \sum_{j=0}^{\infty} q_{j+k+1} w_{j+1}(d, \beta).$$

Moreover, defining $U_X(y, d) := \int_y^\infty u \frac{\partial}{\partial u} H_X(u, d)\, du$*, it also holds that*

$$U_X(y, d) = \frac{1}{\beta} \sum_{k=0}^{\infty} (k+1) \Delta_k(d, \beta, \underline{Q}) \overline{W}_{k+2}(y, \beta), y > 0.$$

Note that with some straightforward calculation it can be proved that $U_X(y, d) < \infty$. The following result is proved in Appendix 6.5. We introduce the convention that an empty product equals 1.

Lemma 6.2.8. *Consider the independent r.v.s* $X_i \sim ME\left(\beta, \underline{Q}_i\right)$*, let* $d_i > 0$ *and* $Y_i = (X_i - d_i)_+, i \in I_n$*. Then the df of* $R_n = \sum_{i=1}^n Y_i$ *can be written as*

$$\begin{aligned} F_{R_n}(y) &= \prod_{i=1}^n F_{X_i}(d_i) + \sum_{k=1}^n \sum_{1 \leq j_1 < \ldots < j_k \leq n} H_{X_{j_1} + \ldots + X_{j_k}}(y, d_{j_1}, \ldots, d_{j_k}) \\ &\quad \times \prod_{i \in I_n \setminus \{j_1, \ldots, j_k\}} F_{X_i}(d_i), y \geq 0, \end{aligned} \tag{6.2.12}$$

where, for $k \geq 1$,

$$H_{\sum_{i=1}^k X_i}(y, d_1, \ldots, d_k) := \mathbb{P}\left(\bigcap_{i=1}^k (X_i > d_i), \sum_{i=1}^k (X_i - d_i) \leq y\right)$$

$$= \sum_{h_1=0}^{\infty} ... \sum_{h_k=0}^{\infty} \Delta_{h_1}\left(d_1, \beta, \underline{Q}_1\right) \cdot ... \cdot \Delta_{h_k}\left(d_k, \beta, \underline{Q}_k\right) W_{\sum_{i=1}^{k} h_i + k}\left(y, \beta\right).$$

Moreover, if

$$U_{\sum_{i=1}^{k} X_i, X_{k+1}}(y, d_1, \ldots, d_{k+1}) := \int_y^{\infty} \int_0^s u \frac{\partial}{\partial u} H_{X_{k+1}}(u, d_{k+1}) \left[\frac{\partial}{\partial v} H_{\sum_{i=1}^{k} X_i}\left(v, d_1, ..., d_k\right)\right]_{v=s-u} duds,$$

then

$$U_{\sum_{i=1}^{k} X_i, X_{k+1}}(y, d_1, \ldots, d_{k+1}) = \frac{1}{\beta} \sum_{h_1=0}^{\infty} ... \sum_{h_{k+1}=0}^{\infty} \left(h_{k+1}+1\right) \Delta_{h_1}(d_1, \beta, \underline{Q}_1) \cdot ... \cdot \Delta_{h_{k+1}}(d_{k+1}, \beta, \underline{Q}_{k+1}) \times \overline{W}_{\sum_{i=1}^{k+1} h_i + k + 2}(y, \beta).$$

6.2.3 TVaR capital allocation

As mentioned in the introduction, it is of great interest for insurance and reinsurance companies to quantify the total capital required for the safety of the company, and also to determine the part of this capital to be allocated to each risk/portfolio in order to cover its loss. Among the capital allocation techniques discussed in the literature, we shall consider the TVaR rule. In order to present the allocation formulas, we recall the definitions of the VaR and TVaR risk measures for a risk X and tolerance level $p \in (0,1)$, i.e.,

$$VaR_p\left(X\right) = \min\left\{x \left| F_X\left(x\right) \geq p\right.\right\}, TVaR_p\left(X\right) = \mathbb{E}\left(X \left| X > VaR_p\left(X\right)\right.\right).$$

Let X_i denote the ith risk r.v. of an insurance portfolio and let $S = \sum_{i=1}^{n} X_i$ represent the aggregate risk of the portfolio. Then, if the total risk capital is evaluated as $TVaR_p\left(S\right)$, the TVaR capital allocation rule naturally allocates to the ith risk

$$C_i(p) = TVaR_p\left(X_i, S\right) := \mathbb{E}\left(X_i \left| S > VaR_p\left(S\right)\right.\right),$$

which can be rewritten as

$$C_i(p) = \frac{1}{1-p} \mathbb{E}\left(X_i \mathbb{1}_{\{S > VaR_p(S)\}}\right), \tag{6.2.13}$$

where $\mathbb{1}_A$ denotes the indicator function of the set A. Clearly, $TVaR_p(S) = \sum_{i=1}^{n} C_i(p)$.

6.3 Main results

6.3.1 Joint distribution of aggregate Sarmanov risks

We consider n insurance portfolios where each portfolio consists of $k_1, \ldots, k_n$ risks, respectively. We denote by $S_i = \sum_{j=1}^{k_i} X_j^{(i)}$ the aggregate risk of portfolio i, where $X_j^{(i)}$ is the jth individual risk from the ith portfolio having pdf $f_j^{(i)}, j = 1, \ldots, k_i, i \in I_n$. We assume that the joint distribution of $\boldsymbol{X} := \left(X_1^{(1)}, \ldots, X_{k_1}^{(1)}; \ldots; X_1^{(n)}, \ldots, X_{k_n}^{(n)}\right)$ is governed by Sarmanov's distribution with the pdf as defined in (6.2.4) and fulfilling (6.2.2) and (6.2.5) for the kernel functions ϕ, i.e., in this case,

$$\begin{aligned} h(\boldsymbol{x}) &= \prod_{i=1}^{n} \prod_{j=1}^{k_i} f_j^{(i)}\left(x_j^{(i)}\right) \left[1 + \sum_{1\le a<b\le n} \sum_{s=1}^{k_a} \sum_{t=1}^{k_b} \alpha_{s,t}^{(a,b)} \phi_s^{(a)}\left(x_s^{(a)}\right) \phi_t^{(b)}\left(x_t^{(b)}\right) \right. \\ & \left. + \sum_{a=1}^{n} \sum_{1\le s<t\le k_a} \alpha_{s,t}^{(a)} \phi_s^{(a)}\left(x_s^{(a)}\right) \phi_t^{(a)}\left(x_t^{(a)}\right) \right], \end{aligned} \tag{6.3.1}$$

where $\boldsymbol{x} = \left(x_1^{(1)}, \ldots, x_{k_1}^{(1)}, \ldots, x_1^{(n)}, \ldots, x_{k_n}^{(n)}\right)$.
Next, we present the joint density of $\boldsymbol{S} = (S_1, \ldots, S_n)$ under these assumptions.

Theorem 6.3.1. *The joint pdf of $\boldsymbol{S}$ can be expressed as follows*

$$f_{\boldsymbol{S}}(s_1, \ldots, s_n) = \prod_{i=1}^{n} f_{S_i}(s_i) + \sum_{1\le a\le b\le n} \sum_{s=1}^{k_a} \sum_{t=T(s,a,b)}^{k_b} \alpha_{s,t}^{(a,b)} \prod_{i=1}^{n} \tilde{f}_{S_{i;s,t}^{(a,b)}}(s_i),$$

where $T(s,a,b) = \max\left\{1, (s+1)\mathbb{1}_{(a=b)}\right\}, \alpha_{s,t}^{(a,a)} = \alpha_{s,t}^{(a)}$,

$$f_{S_i} = f_1^{(i)} * \ldots * f_{k_i}^{(i)}, \tilde{f}_{S_{i;s,t}^{(a,b)}} = \tilde{f}_{1;s,t}^{(i;a,b)} * \ldots * \tilde{f}_{k_i;s,t}^{(i;a,b)}, i \in I_n, \tag{6.3.2}$$

and, for $i \in I_n, j = 1, \ldots, k_i$,

$$\tilde{f}_{j;s,t}^{(i;a,b)}(x) = \begin{cases} \left(\phi_j^{(i)} f_j^{(i)}\right)(x) & if\ (i,j) \in \{(a,s),(b,t)\} \\ f_j^{(i)}(x) & otherwise \end{cases}.$$

Remarks 6.3.2. *It should be noted that Ratovomirija [100] provided a general expression for the joint density of $\boldsymbol{S}$ in the particular case when $k_1 = \ldots = k_n = k$.*

Next, we derive a special case of Theorem 6.3.1 where we assume that all the marginals are mixed Erlang distributed, i.e., $X_j^{(i)} \sim ME(\beta_j^{(i)}, \underline{Q}_j^{(i)})$ with $\underline{Q}_j^{(i)} = (q_{j,1}^{(i)}, q_{j,2}^{(i)}, \ldots), j = 1, \ldots, k_i, i \in I_n$. Moreover, individual risks within and across the portfolios are considered to be joined by Sarmanov's distribution with the joint pdf specified in (6.3.1) and kernel functions $\phi_j^{(i)}(x) = f_j^{(i)}(x) - \mathbb{E}\left(f_j^{(i)}\left(X_j^{(i)}\right)\right)$. We denote

$$\boldsymbol{X} = \left(X_1^{(1)}, \ldots, X_{k_1}^{(1)}; \ldots; X_1^{(n)}, \ldots, X_{k_n}^{(n)}\right) \sim SME_\zeta(\boldsymbol{\beta}, \underline{\boldsymbol{Q}}, \boldsymbol{\alpha}),$$

where $\zeta = \sum_{i=1}^n k_i, \boldsymbol{\beta} = \left(\beta_1^{(1)}, \ldots, \beta_{k_1}^{(1)}; \ldots; \beta_1^{(n)}, \ldots, \beta_{k_n}^{(n)}\right)$,
$\underline{\boldsymbol{Q}} = \left(\underline{Q}_1^{(1)}, \ldots, \underline{Q}_{k_1}^{(1)}; \ldots; \underline{Q}_1^{(n)}, \ldots, \underline{Q}_{k_n}^{(n)}\right)$ and $\boldsymbol{\alpha}$ consists of all the $\alpha-$coefficients of the Sarmanov pdf (6.3.1). In the following, for simplicity, we also denote $\gamma_j^{(i)} = \mathbb{E}\left(f_j^{(i)}\left(X_j^{(i)}\right)\right)$ assuming it exists.

Proposition 6.3.3. *If $\boldsymbol{X} \sim SME_\zeta(\boldsymbol{\beta}, \underline{\boldsymbol{Q}}, \boldsymbol{\alpha})$ with $\beta_{k_n}^{(n)} \geq \beta_j^{(i)}$, for $j = 1, \ldots, k_i, i \in I_n$, then the df of $\boldsymbol{S}$ is given by*

$$\begin{aligned} F_{\boldsymbol{S}}(\boldsymbol{s}) &= \xi_n \prod_{j=1}^n F_{S_j^{(1)}}(s_j) - \sum_{1\leq a\leq b\leq n} \sum_{s=1}^{k_a} \sum_{t=T(s,a,b)}^{k_b} \alpha_{s,t}^{(a,b)} \gamma_s^{(a)} \gamma_t^{(b)} \Bigg(\prod_{j=1}^n F_{S_{j;s}^{(2;a)}}(s_j) \\ &\quad + \prod_{j=1}^n F_{S_{j;t}^{(2;b)}}(s_j) - \prod_{j=1}^n F_{S_{j;s,t}^{(3;a,b)}}(s_j) \Bigg), \end{aligned}$$

where

$$\xi_n = 1 + \sum_{1\leq a\leq b\leq n} \sum_{s=1}^{k_a} \sum_{t=T(s,a,b)}^{k_b} \alpha_{s,t}^{(a,b)} \gamma_s^{(a)} \gamma_t^{(b)}, \tag{6.3.3}$$

$$S_j^{(1)} \sim ME\left(2\beta_{k_n}^{(n)}, \underline{\Pi}\left(\Psi(\underline{Q}_1^{(j)}), \ldots, \Psi(\underline{Q}_{k_j}^{(j)})\right)\right), \tag{6.3.4}$$

$$S_{j;s,t}^{(3;a,b)} \sim ME\left(2\beta_{k_n}^{(n)}, \underline{\Pi}\left(\Psi\left(M_{s,t}^{(a,b)}(\underline{Q}_1^{(j)})\right), \ldots, \Psi\left(M_{s,t}^{(a,b)}(\underline{Q}_{k_j}^{(j)})\right)\right)\right), \tag{6.3.5}$$

$$S_{j;s}^{(2;a)} = S_{j;s,s}^{(3;a,a)}, \tag{6.3.6}$$

and

$$M_{s,t}^{(a,b)}\left(\underline{Q}_l^{(j)}\right) = \begin{cases} V\left(\underline{Q}_l^{(j)}\right) & \textit{if} \quad (j,l) \in \{(a,s),(b,t)\} \\ \underline{Q}_l^{(j)} & \textit{otherwise} \end{cases}.$$

The components of V are defined in Lemma 6.2.2, the elements of Ψ in Lemma

6.2.4 and the ones of $\underline{\Pi}$ are given in Remark 6.2.6.

6.3.2 Stop-loss mixed Erlang reinsurance risks with Sarmanov dependence

In this section, we study the effect of mixed Erlang distributed risks on reinsurance. In order to mitigate their risks, insurers enter into reinsurance agreements. There are several types of reinsurance contracts. However, we shall only consider the stop-loss reinsurance. In a stop-loss reinsurance contract, the reinsurer pays the part of the loss that is greater than a certain positive amount d (the deductible). In the following, we shall provide the distribution of the aggregated loss of several reinsurance portfolios in the stop-loss framework, and determine the amount of capital to be allocated to each reinsurance portfolio under the TVaR allocation principle.

In this respect, we consider n insurance portfolios as defined in the last section with aggregated losses $(S_1, \ldots, S_n)$ subject to the deductibles $\boldsymbol{d} = (d_1, \ldots, d_n)$ on the reinsured amounts $(T_1, \ldots, T_n)$, where the T_i's, $i \in I_n$, are defined as follows

$$T_i = (S_i - d_i)_+ = \begin{cases} 0 & \text{if } S_i \leq d_i \\ S_i - d_i & \text{if } S_i > d_i \end{cases}.$$

Hereafter, we shall denote by $R_n = \sum_{i=1}^{n} T_i$ the aggregated reinsurance stop-loss risk.

Proposition 6.3.4. *Let $(X_1^{(1)}, \ldots, X_{k_1}^{(1)}; \ldots; X_1^{(n)}, \ldots, X_{k_n}^{(n)}) \sim SME_\zeta(\boldsymbol{\beta}, \underline{\boldsymbol{Q}}, \boldsymbol{\alpha})$ with $\gamma_j^{(i)} < \infty$ and $\beta_{k_n}^{(n)} \geq \beta_j^{(i)}, j = 1, \ldots, k_i$, and let $d_i > 0$ for $i \in I_n$. Then the df of R_n is given by*

$$\begin{aligned} F_{R_n}(y) &= F_{\mathbf{S}}(\mathbf{d}) + \sum_{k=1}^{n} \sum_{1 \leq j_1 < \ldots < j_k \leq n} \Bigg[\xi_n H_{S_{j_1}^{(1)} + \ldots + S_{j_k}^{(1)}}(y, d_{j_1}, \ldots, d_{j_k}) \prod_{i \in I_n \setminus \{j_1, \ldots, j_k\}} F_{S_i^{(1)}}(d_i) \\ &- \sum_{1 \leq a \leq b \leq n} \sum_{s=1}^{k_a} \sum_{t=T(s,a,b)}^{k_b} \alpha_{s,t}^{(a,b)} \gamma_s^{(a)} \gamma_t^{(b)} \Bigg(H_{S_{j_1;s}^{(2;a)} + \ldots + S_{j_k;s}^{(2;a)}}(y, d_{j_1}, \ldots, d_{j_k}) \prod_{i \in I_n \setminus \{j_1, \ldots, j_k\}} F_{S_{i;s}^{(2;a)}}(d_i) \\ &+ H_{S_{j_1;t}^{(2;b)} + \ldots + S_{j_k;t}^{(2;b)}}(y, d_{j_1}, \ldots, d_{j_k}) \prod_{i \in I_n \setminus \{j_1, \ldots, j_k\}} F_{S_{i;t}^{(2;b)}}(d_i) \\ &- H_{S_{j_1;s,t}^{(3;a,b)} + \ldots + S_{j_k;s,t}^{(3;a,b)}}(y, d_{j_1}, \ldots, d_{j_k}) \prod_{i \in I_n \setminus \{j_1, \ldots, j_k\}} F_{S_{i;s,t}^{(3;a,b)}}(d_i) \Bigg) \Bigg], y \geq 0, \end{aligned} \tag{6.3.7}$$

with H defined in Lemmas 6.2.7- 6.2.8.

Next, we shall consider capital allocation under the TVaR principle for the reinsurance risks corresponding to the n portfolios defined above. Let $C_i(p)$ be the amount of capital to be allocated to portfolio $i, i \in I_n$, as defined in (6.2.13). The following result holds.

Proposition 6.3.5. *Let* $(X_1^{(1)}, \dots, X_{k_1}^{(1)}; \dots; X_1^{(n)}, \dots, X_{k_n}^{(n)}) \sim SME_\zeta(\boldsymbol{\beta}, \underline{\boldsymbol{Q}}, \boldsymbol{\alpha})$ *such that* $\gamma_j^{(i)} < \infty$ *and* $\beta_{k_n}^{(n)} \geq \beta_j^{(i)}$, *for* $i \in I_n$, $j = 1, \dots, k_i$. *Let* $d_i > 0, i \in I_n$ *and set* $x_p := VaR_p(R_n)$. *Then the capital allocated to portfolio* l *under the TVaR rule is*

$$\begin{aligned} C_l(p) &= \frac{1}{1-p} \sum_{k=0}^{n-1} \sum_{\substack{1 \leq j_1 < \ldots < j_k \leq n \\ l \notin \{j_1, \ldots, j_k\}}} \Bigg[\xi_n U_{\sum_{i=1}^k S_{j_i}^{(1)}, S_l^{(1)}} (x_p, d_{j_1}, .., d_{j_k}, d_l) \prod_{i \in I_n \setminus \{j_1, \ldots, j_k, l\}} F_{S_i^{(1)}}(d_i) \\ &- \sum_{1 \leq a \leq b \leq n} \sum_{s=1}^{k_a} \sum_{t=T(s,a,b)}^{k_b} \alpha_{s,t}^{(a,b)} \gamma_s^{(a)} \gamma_t^{(b)} \Bigg(U_{\sum_{i=1}^k S_{j_i;s}^{(2;a)}, S_{l;s}^{(2;a)}} (x_p, d_{j_1}, .., d_{j_k}, d_l) \prod_{i \in I_n \setminus \{j_1, \ldots, j_k, l\}} F_{S_{i;s}^{(2;a)}}(d_i) \\ &+ U_{\sum_{i=1}^k S_{j_i;t}^{(2;b)}, S_{l;t}^{(2;b)}} (x_p, d_{j_1}, .., d_{j_k}, d_l) \prod_{i \in I_n \setminus \{j_1, \ldots, j_k, l\}} F_{S_{i;t}^{(2;b)}}(d_i) \\ &- U_{\sum_{i=1}^k S_{j_i;s,t}^{(3;a,b)}, S_{l;s,t}^{(3;a,b)}} (x_p, d_{j_1}, .., d_{j_k}, d_l) \prod_{i \in I_n \setminus \{j_1, \ldots, j_k, l\}} F_{S_{i;s,t}^{(3;a,b)}}(d_i) \Bigg) \Bigg], \end{aligned}$$

where, by convention, when $k = 0$, *we consider only one term in the sum* $\sum_{\substack{1 \leq j_1 < \ldots < j_k \leq n \\ l \notin \{j_1, \ldots, j_k\}}}$ *in which each component of the type* $U_{\sum_{i=1}^k S_{j_i}, S_l}$ *is replaced with* U_{S_l}.

Example 6.3.1. *Let* S_1 *and* S_2 *be the aggregate risks of two insurance portfolios consisting of* $k_1 = 2$ *and* $k_2 = 3$ *mixed Erlang distributed risks, respectively, with* $\beta_3^{(2)} > \beta_j^{(i)}, j = 1, \dots, k_i$ *and* $i = 1, 2$. *Hence,* $S_1 = X_1^{(1)} + X_2^{(1)}$ *and* $S_2 = X_1^{(2)} + X_2^{(2)} + X_3^{(2)}$. *Following Propositions 6.3.3-6.3.4, the distribution of the aggregate stop-loss reinsurance risk* $R_2 = T_1 + T_2$, *where* $T_i = (S_i - d_i)_+, i = 1, 2$, *is given by*

$$\begin{aligned} F_{R_2}(y) &= \xi_2 \Big(F_{S_1^{(1)}}(d_1) F_{S_2^{(1)}}(d_2) + H_{S_1^{(1)}}(y, d_1) F_{S_2^{(1)}}(d_2) + H_{S_2^{(1)}}(y, d_2) F_{S_1^{(1)}}(d_1) + H_{S_1^{(1)} + S_2^{(1)}}(y, d_1, d_2) \Big) \\ &- \sum_{1 \leq a \leq b \leq 2} \sum_{s=1}^{k_a} \sum_{t=T(s,a,b)}^{k_b} \alpha_{s,t}^{(a,b)} \gamma_s^{(a)} \gamma_t^{(b)} \Bigg[\sum_{(i,j) \in \{(a,s),(b,t)\}} \Big(F_{S_{1;j}^{(2;i)}}(d_1) F_{S_{2;j}^{(2;i)}}(d_2) + H_{S_{1;j}^{(2;i)}}(y, d_1) F_{S_{2;j}^{(2;i)}}(d_2) \\ &+ H_{S_{2;j}^{(2;i)}}(y, d_2) F_{S_{1;j}^{(2;i)}}(d_1) + H_{S_{1;j}^{(2;i)} + S_{2;j}^{(2;i)}}(y, d_1, d_2) \Big) - F_{S_{1;s,t}^{(3;a,b)}}(d_1) F_{S_{2;s,t}^{(3;a,b)}}(d_2) \\ &- H_{S_{1;s,t}^{(3;a,b)}}(y, d_1) F_{S_{2;s,t}^{(3;a,b)}}(d_2) - H_{S_{2;s,t}^{(3;a,b)}}(y, d_2) F_{S_{1;s,t}^{(3;a,b)}}(d_1) - H_{S_{1;s,t}^{(3;a,b)} + S_{2;s,t}^{(3;a,b)}}(y, d_1, d_2) \Bigg]. \end{aligned}$$

Furthermore, if $TVaR_p(R_2)$ *is the total risk capital needed to cover* R_2, *in light of Proposition 6.3.5, the contribution of* T_i *to this capital is expressed as follows*

$$C_i(p) = \frac{1}{1-p} \Bigg\{ \xi_2 \Big(U_{S_i^{(1)}}(x_p, d_i) F_{S_j^{(1)}}(d_j) + U_{S_j^{(1)}, S_i^{(1)}}(x_p, d_j, d_i) \Big) - \sum_{1 \leq a \leq b \leq 2} \sum_{s=1}^{k_a} \sum_{t=T(s,a,b)}^{k_b} \alpha_{s,t}^{(a,b)} \gamma_s^{(a)} \gamma_t^{(b)}$$

$$\times \left[\sum_{(k,l)\in\{(a,s),(b,t)\}} \left(U_{S_{i;l}^{(2,k)}}(x_p, d_i) F_{S_{j;l}^{(2;k)}}(d_j) + U_{S_{j;l}^{(2;k)}, S_{i;l}^{(2,k)}}(x_p, d_j, d_i) \right) \right.$$
$$\left. -U_{S_{i;s,t}^{(3;a,b)}}(x_p, d_i) F_{S_{j;s,t}^{(3;a,b)}}(d_j) - U_{S_{j;s,t}^{(3;a,b)}, S_{i;s,t}^{(3;a,b)}}(x_p, d_j, d_i) \right] \Bigg\}, i \neq j \in \{1, 2\},$$

where ξ_2 *is defined in (6.3.3) for* $n = 2$*, while* H *and* U *are given in Lemmas 6.2.7-6.2.8.*

Numerical illustration. *To numerically illustrate the just mentioned formulas, in Table 6.3.1 we present the parameters of the individual risks* $X_j^{(i)}$ *of the two portfolios, where* $j = 1, \ldots, k_i$ *and* $i = 1, 2$*, together with some related statistical measures (for simplicity, only two decimal places were retained).*

	$X_j^{(i)}$	$\beta_j^{(i)}$	$\underline{Q}_j^{(i)}$	Mean	Variance	Skewness	Kurtosis
Portfolio I	$X_1^{(1)}$	0.12	(0.4,0.6)	13.33	127.78	1.55	6.50
	$X_2^{(1)}$	0.14	(0.3,0.7)	12.14	97.45	1.49	6.28
Portfolio II	$X_1^{(2)}$	0.15	(0.5,0.5)	10.00	77.78	1.62	6.80
	$X_2^{(2)}$	0.16	(0.8,0.2)	7.50	53.13	1.88	8.16
	$X_3^{(2)}$	0.18	(0.55,0.45)	8.06	52.39	1.66	6.97

Table 6.3.1: Statistical measures for the individual risks $X_j^{(i)}, j = 1, \ldots, k_i, i = 1, 2$ (Example 6.3.1).

Moreover, we assume that the Sarmanov parameters $\alpha_{i,j}$ *are as follows*

$$\alpha_{1,2}^{(1)} = 16, \quad \alpha_{1,1}^{(1,2)} = 8, \quad \alpha_{1,2}^{(1,2)} = 5, \quad \alpha_{1,3}^{(1,2)} = 2, \quad \alpha_{2,1}^{(1,2)} = 8,$$
$$\alpha_{2,2}^{(1,2)} = 5, \quad \alpha_{2,3}^{(1,2)} = 2, \quad \alpha_{1,2}^{(2)} = 15, \quad \alpha_{1,3}^{(2)} = 17, \quad \alpha_{2,3}^{(2)} = 16.$$

Under the stop-loss reinsurance framework, we considered the values $d_1 = 50$ *and* $d_2 = 45$ *for the deductibles of Portfolios I and II, respectively. Table 6.3.2 describes the allocated capitals* $C_i(p), i = 1, 2$*, required to cover the losses of both portfolios after application of the deductibles, as well as the capital needed to cover the loss* R_2 *of the whole reinsured portfolio. We considered several values for the tolerance level* p*.*

p(%)	$VaR_p(R_2)$	$C_1 = TVaR_p(T_1, R_2)$	$C_2 = TVaR_p(T_2, R_2)$	$TVaR_p(R_2)$
90.00	5.03	7.33	8.37	15.70
92.50	8.24	8.85	9.90	18.75
95.00	12.65	11.07	11.90	22.97
97.50	19.96	15.08	14.96	30.04
99.00	29.31	20.82	18.34	39.16
99.90	51.88	37.35	24.05	61.40

Table 6.3.2: Capital allocated to Portfolios I and II (Example 6.3.1).

Table 6.3.2 shows that for a tolerance level $p \geq 97.5\%$, Portfolio I is riskier than Portfolio II as more capital is needed to cover the losses (this can be explained by the fact that both risks in Portfolio I have higher expected values and variances than the risks in Portfolio II); however, for $p \leq 95\%$, more capital is allocated to Portfolio II.

6.3.3 Particular case: mixed Erlang risks with Sarmanov dependency

We shall now consider the same setting as before, but in the particular case with only one insurance portfolio, no reinsurance and no deductible. For simplicity, we denote by $\boldsymbol{X} = (X_1, \ldots, X_k)$ the k individual risks with joint distribution governed by the k-variate Sarmanov distribution with kernel functions $\phi_j(x_j) = f_j(x_j) - \gamma_j$, where $\gamma_j = \mathbb{E}(f_j(X_j))$, $X_j \sim ME(\beta_j, \underline{Q}_j)$, $j \in I_k$, and we denote by $S = \sum_{j=1}^k X_j$ the aggregate risk of the portfolio. Hence, $\boldsymbol{X} \sim SME_k(\boldsymbol{\beta}, \underline{\boldsymbol{Q}}, \boldsymbol{\alpha})$, where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)$, $\underline{\boldsymbol{Q}} = (\underline{Q}_1, \ldots, \underline{Q}_k)$ and $\boldsymbol{\alpha} = (\alpha_{i,j})_{1 \leq i < j \leq k}$. Next, we are going to present the distribution of the aggregate risk S that can easily be derived from Proposition 6.3.3.

Proposition 6.3.6. *Let $\boldsymbol{X} \sim SME_k(\boldsymbol{\beta}, \underline{\boldsymbol{Q}}, \boldsymbol{\alpha})$, where $\beta_j \leq \beta_k$ for $j = 1, \ldots, k-1$. Then the distribution of the aggregate risk S is given by*

$$F_S(u) = \xi_k F_{S^{(1)}}(u) - \sum_{1 \leq s < t \leq k} \alpha_{s,t} \gamma_s \gamma_t \left(F_{S_s^{(2)}}(u) + F_{S_t^{(2)}}(u) - F_{S_{s,t}^{(3)}}(u) \right),$$

where $\xi_k = 1 + \sum_{1 \leq s < t \leq k} \alpha_{s,t} \gamma_s \gamma_t$, while

$$S^{(1)} \sim ME\Big(2\beta_k, \underline{\Pi}\left(\Psi(\underline{Q}_1), \ldots, \Psi(\underline{Q}_k)\right)\Big),$$

$$S_{s,t}^{(3)} \sim ME\left(2\beta_k, \underline{\Pi}\left(\Psi\left(M_{s,t}(\underline{Q}_1)\right), \ldots, \Psi\left(M_{s,t}(\underline{Q}_k)\right)\right)\right), \; S_s^{(2)} = S_{s,s}^{(3)},$$

and

$$M_{s,t}\left(\underline{Q}_l\right) = \begin{cases} V(\underline{Q}_l) & \text{if} \quad l \in \{s,t\} \\ \underline{Q}_l & \text{otherwise} \end{cases}.$$

Corollary 6.3.7. *Under the assumptions of Proposition 6.3.6 it follows that* $S \sim ME(2\beta_k, \underline{P})$*, where the components of the vector of mixing weights* $\underline{P} = (p_1, p_2, \ldots)$ *are defined by*

$$\begin{aligned} p_i &= \xi_k \pi_i\left(\Psi(\underline{Q}_1), \ldots, \Psi(\underline{Q}_k)\right) - \sum_{1 \le s < t \le k} \alpha_{s,t}\gamma_s\gamma_t \Big[\pi_i\left(\Psi\left(M_s(\underline{Q}_1)\right), \ldots, \Psi\left(M_s(\underline{Q}_k)\right)\right) \\ &\quad + \pi_i\left(\Psi\left(M_t(\underline{Q}_1)\right), \ldots, \Psi\left(M_t(\underline{Q}_k)\right)\right) - \pi_i\left(\Psi\left(M_{s,t}(\underline{Q}_1)\right), \ldots, \Psi\left(M_{s,t}(\underline{Q}_k)\right)\right)\Big], \end{aligned}$$

where $M_s = M_{s,s}$ *and* π_i *are the components of* $\underline{\Pi}$ *defined in Remark 6.2.6.*

Example 6.3.2. *Bivariate mixed Erlang risks joined by Sarmanov's distribution. Let* $(X_1, X_2) \sim SME_2\left(\boldsymbol{\beta} = (\beta_1, \beta_2), (\underline{Q}_1, \underline{Q}_2), \alpha_{1,2}\right)$ *with* $\beta_1 < \beta_2$*. It follows that* $S = X_1 + X_2 \sim ME(2\beta_2, \underline{P})$*, where the components of the vector* $\underline{P}$ *are given below*

$$\begin{aligned} p_i &= (1 + \alpha_{1,2}\gamma_1\gamma_2)\, \pi_i\left(\Psi(\underline{Q}_1), \Psi(\underline{Q}_2)\right) - \alpha_{1,2}\gamma_1\gamma_2 \Big[\pi_i\left(\Psi(V(\underline{Q}_1)), \Psi(\underline{Q}_2)\right) \\ &\quad + \pi_i\left(\Psi(\underline{Q}_1), \Psi(V(\underline{Q}_2))\right) - \pi_i\left(\Psi(V(\underline{Q}_1)), \Psi(V(\underline{Q}_2))\right)\Big], \end{aligned}$$

such that $\sum_{i=1}^{\infty} p_i = 1$.

Numerical illustration. *As a numerical illustration, we considered a bivariate vector* (X_1, X_2) *such that*

$$(X_1, X_2) \sim SME_2\Big(\boldsymbol{\beta} = (0.9, 0.95), \underline{Q}_1 = (0.4, 0.6), \underline{Q}_2 = (0.8, 0.2), \alpha_{1,2} = 2.5\Big).$$

Thus, the densities of X_1 *and* X_2 *can be, respectively, written as follows:*

$$f_1(x) = 0.4w_1(x, 0.9) + 0.6w_2(x, 0.9), \quad f_2(x) = 0.8w_1(x, 0.95) + 0.2w_2(x, 0.95).$$

Moreover, from formula (6.5.2) we have

$$\mathbb{E}\left(f_i(X_i)\right) = \beta_i \sum_{l=1}^{2}\sum_{j=1}^{2} \binom{l+j-2}{l-1} \frac{q_{i,l}q_{i,j}}{2^{l+j-1}}, \quad i = 1, 2,$$

yielding $\gamma_1 = 0.261, \gamma_2 = 0.3895$*. In the following, we restrict to only two decimal* *places. Then the joint pdf of* (X_1, X_2) *is given by*

$$h(x_1, x_2) = f_1(x_1)f_2(x_2)\left(1.25 + 2.5f_1(x_1)f_2(x_2) - 0.97f_1(x_1) - 0.65f_2(x_2)\right).$$

Table 6.3.3 summarizes some quantitative measures related to the marginals X_1 *and* X_2.

	Expected value	Variance	Skewness	Kurtosis
X_1	1.78	2.27	1.55	6.50
X_2	1.26	1.51	1.88	8.16

Table 6.3.3: Quantitative measures for X_1 and X_2 (Example 6.3.2).

As stated above, the distribution of the aggregate risk S *is again mixed Erlang with scale parameter* $2\beta_2 = 1.9$ *and the mixing probabilities given in Table 6.3.4.*

i	p_i	i	p_i	i	p_i	i	p_i
1	0.0000	11	0.0262	21	0.0002	31	8.635E-07
2	0.0827	12	0.0173	22	0.0001	32	4.873E-07
3	0.1547	13	0.0112	23	7.443E-05	33	2.743E-07
4	0.1709	14	0.0071	24	4.326E-05	34	1.540E-07
5	0.1390	15	0.0045	25	2.502E-05	35	8.625E-08
6	0.1162	16	0.0028	26	1.441E-05	36	4.821E-08
7	0.0956	17	0.0017	27	8.263E-06	37	2.689E-08
8	0.0744	18	0.0010	28	4.722E-06	38	1.497E-08
9	0.0547	19	0.0006	29	2.689E-06	39	8.319E-09
10	0.0385	20	0.0004	30	1.526E-06	40	4.615E-09

Table 6.3.4: Mixing probabilities of S (Example 6.3.2).

We are now interested in quantifying the amount of capital $C_j(p)$ to be allocated to each risk $X_j, j \in I_k$.

Proposition 6.3.8. *Let* $\boldsymbol{X} \sim SME_k(\boldsymbol{\beta}, \underline{\boldsymbol{Q}}, \boldsymbol{\alpha})$ *with* $\beta_j \leq \beta_k, j \in I_k$*, and let* $s_p = VaR_p(S)$*. Then the amount of capital* C_j *allocated to each risk* X_j *under the TVaR* *allocation principle as defined in* (6.2.13) *can be expressed as*

$$C_j(p) = \frac{1}{1-p}\sum_{i=1}^{\infty} z_{i,j}\overline{W}_i(s_p, 2\beta_k), \tag{6.3.8}$$

where the mixing coefficients $z_{i,j}$ *are given by (here the transform* Ψ *is needed to obtain the common scale parameter* $2\beta_k$*)*

$$\begin{aligned} z_{i,j} &= \xi_k\mu_j\pi_i\left(\Psi\left(\widetilde{M}_j(\underline{Q}_1)\right),\ldots,\Psi\left(\widetilde{M}_j(\underline{Q}_k)\right)\right) - \sum_{1\le a<b\le k}\alpha_{a,b}\gamma_a\gamma_b\Big[\varphi_{j;b}\pi_i\left(\Psi\left(\widetilde{M}_{j;b}(\underline{Q}_1)\right),\ldots,\Psi\left(\widetilde{M}_{j;b}(\underline{Q}_k)\right)\right) \\ &\quad +\varphi_{j;a}\pi_i\left(\Psi\left(\widetilde{M}_{j;a}(\underline{Q}_1)\right),\ldots,\Psi\left(\widetilde{M}_{j;a}(\underline{Q}_k)\right)\right) - \varphi_{j;a,b}\pi_i\left(\Psi\left(\widetilde{M}_{j;a,b}(\underline{Q}_1)\right),\ldots,\Psi\left(\widetilde{M}_{j;a,b}(\underline{Q}_k)\right)\right)\Big], \end{aligned} \tag{6.3.9}$$

with $\xi_k = 1 + \sum_{1\le a<b\le k}\alpha_{a,b}\gamma_a\gamma_b$, $\mu_i = \mathbb{E}(X_i) = \frac{1}{\beta_i}\sum_{k=1}^{\infty} kq_{i,k}$, $\tilde{\mu}_i = \frac{1}{2\beta_i}\sum_{k=1}^{\infty} kv_{i,k}$ *as defined in formula* (6.2.10),

$$\varphi_{j;a,b} = \begin{cases} \mu_j & \textit{if } \; j\notin\{a,b\} \\ \tilde{\mu}_j & \textit{if } \; j\in\{a,b\}\end{cases}, \; \varphi_{j;a} = \varphi_{j;a,a}, \tag{6.3.10}$$

and

$$\widetilde{M}_j(\underline{Q}_i) = \begin{cases} \underline{Q}_i & \textit{if } i\neq j \\ G(\underline{Q}_i) & \textit{if } i=j \end{cases}, \; \widetilde{M}_{j;a,b}(\underline{Q}_i) = \begin{cases} \underline{Q}_i & \textit{if } i\notin\{j,a,b\} \\ V(\underline{Q}_i) & \textit{if } i=a \textit{ and } i\notin\{j,b\} \textit{ or} \\ & \textit{if } i=b \textit{ and } i\notin\{j,a\} \\ G(\underline{Q}_i) & \textit{if } i=j \textit{ and } i\notin\{a,b\} \\ G(V(\underline{Q}_i)) & \textit{if } i=j=a \textit{ and } i\neq b \textit{ or} \\ & \textit{if } i=j=b \textit{ and } i\neq a \end{cases},$$

$$\widetilde{M}_{j;a} = \widetilde{M}_{j;a,a}. \tag{6.3.11}$$

Example 6.3.3. *Capital allocation for bivariate mixed Erlang risks joined by Sarmanov's distribution.*

In the bivariate case, with the above notation, $S_2 = X_1 + X_2$ *is the aggregate risk of the portfolio and we consider* $TVaR_p(S_2)$ *to be the total capital needed to cover it, whereas* C_i *is the part of this capital allocated to cover* $X_i, i = 1,2$*. For a numerical illustration, we consider the bivariate vector used in Example 6.3.2, but this time we vary the value of* $\alpha_{1,2}$*. Table 6.3.5 summarizes the results under the TVaR capital allocation principle assuming a tolerance level* $p = 99\%$ *(the second column shows the variance of* S_2 *denoted* $\sigma^2_{S_2}$*).*

α_{12}	$\sigma^2_{S_2}$	$C_1(99\%)$	$C_2(99\%)$	$TVaR_{99\%}(S_2)$
3.4	4.0509	6.3920	4.3958	10.7878
2.5	3.9788	6.3703	4.3556	10.7259
1.5	3.8987	6.3458	4.3086	10.6544
0.5	3.8186	6.3209	4.2589	10.5798
0	3.7785	6.3083	4.2330	10.5413
-0.5	3.7385	6.2956	4.2063	10.5019
-1.5	3.6584	6.2698	4.1505	10.4203
-2.1	3.6103	6.2542	4.1154	10.3696

Table 6.3.5: $TVaR_{99\%}(S_2)$ and capital allocated to each risk $X_i, i = 1, 2$ (Example 6.3.3).

It can be seen from Table 6.3.5 that the total capital needed to cover S_2 is dependent on $\alpha_{1,2}$. Actually, a larger $\alpha_{1,2}$ implies a riskier portfolio (see the corresponding variance, $\sigma^2_{S_2}$) and thus, more capital is needed to cover each risk. Also, it can be seen that X_1 accounts for a larger capital than X_2 as it is riskier (having larger variance and expected value, see Table 6.3.3).

6.4 Appendix: Dependence structure

In this section, we discuss the dependence structure of two mixed Erlang distributed r.v.s (X_1, X_2) joined by the Sarmanov distribution with different kernel functions (in the insurance context, X_1, X_2 are dependent insurance risks). As before, the kernel functions are written in the form $\phi(x) = g(x) - \mathbb{E}(g(X))$, with g properly chosen. To model the dependence between the two r.v.s X_1 and X_2, we shall use Pearson's correlation coefficient denoted by $\rho_{1,2}$, which in the case of Sarmanov's distribution can be rewritten as

$$\rho_{1,2}(X_1, X_2) = \frac{\alpha_{1,2}\mathbb{E}(X_1\phi_1(X_1))\mathbb{E}(X_2\phi_2(X_2))}{\sigma_1\sigma_2}, \tag{6.4.1}$$

where $\sigma_i = \sqrt{Var(X_i)} > 0, i = 1, 2$. Based on (6.4.1), we hereafter present this correlation coefficient for different kernel functions along with its maximal and minimal values, in the particular case of mixed Erlang marginals.

Case 1: Let $g(x) = f(x)$ (i.e., the marginal pdf), which leads to the kernel function $\phi(x) = f(x) - \mathbb{E}(f(X))$. Under the assumption $X_i \sim ME(\beta_i, \underline{Q}_i), i = 1, 2,$ using

Lemmas 6.2.2 and 6.2.3, we obtain

$$\mathbb{E}(X_i f_i(X_i)) \;=\; \int_0^\infty x f(x,\beta_i,\underline{Q}_i)^2 \mathrm{d}x = \gamma_i \int_0^\infty x \frac{f(x,\beta_i,\underline{Q}_i)^2}{\gamma_i}\mathrm{d}x = \gamma_i \int_0^\infty x f(x, 2\beta_i, V(\underline{Q}_i))\mathrm{d}x = \gamma_i \widetilde{\mu}_i,$$

hence $\mathbb{E}(X_i\phi_i(X_i)) = \gamma_i\widetilde{\mu}_i - \gamma_i\mu_i$ and Pearson's coefficient is now given by

$$\rho_{1,2}(X_1, X_2) = \frac{\alpha_{1,2}\gamma_1\gamma_2(\widetilde{\mu}_1 - \mu_1)(\widetilde{\mu}_2 - \mu_2)}{\sigma_1\sigma_2}. \tag{6.4.2}$$

Thus, from (6.2.3), the maximal and the minimal values of Pearson's correlation are respectively given by

$$\rho_{1,2}^{max}(X_1, X_2) = \frac{\gamma_1\gamma_2(\widetilde{\mu}_1 - \mu_1)(\widetilde{\mu}_2 - \mu_2)}{\max\{\gamma_1(M_2 - \gamma_2), (M_1 - \gamma_1)\gamma_2\}\sigma_1\sigma_2}, \tag{6.4.3}$$

$$\rho_{1,2}^{min}(X_1, X_2) = \frac{-\gamma_1\gamma_2(\widetilde{\mu}_1 - \mu_1)(\widetilde{\mu}_2 - \mu_2)}{\max\{\gamma_1\gamma_2, (M_1 - \gamma_1)(M_2 - \gamma_2)\}\sigma_1\sigma_2}, \tag{6.4.4}$$

where we recall $M_i = \max_{x\in\mathbb{R}} f_i(x), i = 1, 2$.

Case 2: We consider $g(x) = e^{-tx}$ for $t = 1$, hence the corresponding kernel function is given by $\phi(x) = e^{-x} - \mathbb{E}\left(e^{-X}\right)$. Pearson's correlation coefficient along with its lower and upper bounds can be found in Hashorva and Ratovomirija [59], where the particular case of mixed Erlang marginals is emphasized.

Case 3: Let $g(x) = x^t$, in which case the kernel function is given by $\phi(x) = x^t - \mathbb{E}(X^t)$. A usual choice here is $t = 1$, which leads to the kernel $\phi(x) = x - \mathbb{E}(X)$ and to the correlation $\rho_{1,2}(X_1, X_2) = \alpha_{1,2}\sigma_1\sigma_2$. In this case, to fulfill the condition $1+\alpha_{1,2}\phi_1(x_1)\phi_2(x_2) \geq 0$, upper truncated distributions can be considered for X_1, X_2, which is not the object of our study. However, if we denote by $T_i, i = 1, 2$, the corresponding upper truncation points and consider that the marginal pdf's are defined only for non-negative values, then the maximal and the minimal values of the correlation coefficient are, respectively, given by

$$\rho_{1,2}^{max}(X_1, X_2) = \frac{\sigma_1\sigma_2}{\max\{\mu_1(T_2 - \mu_2), (T_1 - \mu_1)\mu_2\}},$$

$$\rho_{1,2}^{min}(X_1, X_2) = \frac{-\sigma_1\sigma_2}{\max\{\mu_1\mu_2, (T_1 - \mu_1)(T_2 - \mu_2)\}}.$$

Case 4: We consider the FGM distribution already studied in Cossette et al. [23], obtained for $g(x) = 2\overline{F}(x)$, with the corresponding kernel function $\phi(x) = 1 - 2F(x)$. Its Pearson's correlation coefficient is given by $\rho_{1,2} = \frac{1}{3}\alpha_{1,2}$. The minimal and maximal values of $\rho_{1,2}$ are $-\frac{1}{3}$ and $\frac{1}{3}$, respectively, which is an important drawback of the FGM distribution. Moreover, in the particular case of mixed Erlang marginals joined by the FGM distribution, the Pearson correlation coefficient can be found in Cossette et al. [23].

Example 6.4.1. *Comparison of the dependency between two mixed Erlang r.v.s joined by the Sarmanov distribution.*
a) We consider the bivariate random vector

$$(X_1, X_2) \sim SME_2\Big(\boldsymbol{\beta} = (2, 2.5), \underline{Q}_1 = (0.45, 0.55), \underline{Q}_2 = (0.5, 0.5), \alpha_{1,2}\Big).$$

*In this application, we would like to compare the dependency between X_1 and X_2 based on the four different kernel functions described above. Therefore, we compute the upper and lower bounds of the Pearson correlation coefficients for each kernel, together with the corresponding parameter α, as summarized in the table below. As discussed above, in Case 3 we considered upper truncated distributions with $T_1 = T_2 = 15$ such that the tail functions of X_1, X_2 in this truncation point are very small, hence making this case comparable with the other not-truncated ones. It can be seen that the largest range of dependence corresponds to the kernel considered in **Case 1** (and studied in this paper) and the smallest to the truncated **Case 3**.*

	Kernel	α_{max}	ρ_{max}	α_{min}	ρ_{min}
Case 1	$f(x) - \mathbb{E}(f(X))$	3.2100	0.3023	-2.1289	-0.2005
Case 2	$e^{-x} - \mathbb{E}(e^{-X})$	3.5854	0.1921	-3.000	-0.1607
Case 3	$x - \mathbb{E}(X)$	0.0896	0.0318	-0.0049	-0.0017
Case 4	$1 - 2F(x)$	1.0000	0.2711	-1.0000	-0.2711

Table 6.4.1: Upper and lower bounds of Pearson's correlation coefficient for different kernel functions (Example 6.4.1.a).

b) In the sequel, we assumed a common scale parameter β for both marginals and we plotted the upper and lower bounds of the correlation coefficient as a function of β for the four kernel functions.
The figure below shows that in Case 1, the dependency increases with β, in contrast

with Case 3 (considered with $T_1 = T_2 = 15$), where the dependency decreases with β quite rapidly from the maximum correlation coefficient to approximately 0. Case 2 and Case 4 show an almost constant dependency structure with respect to β.

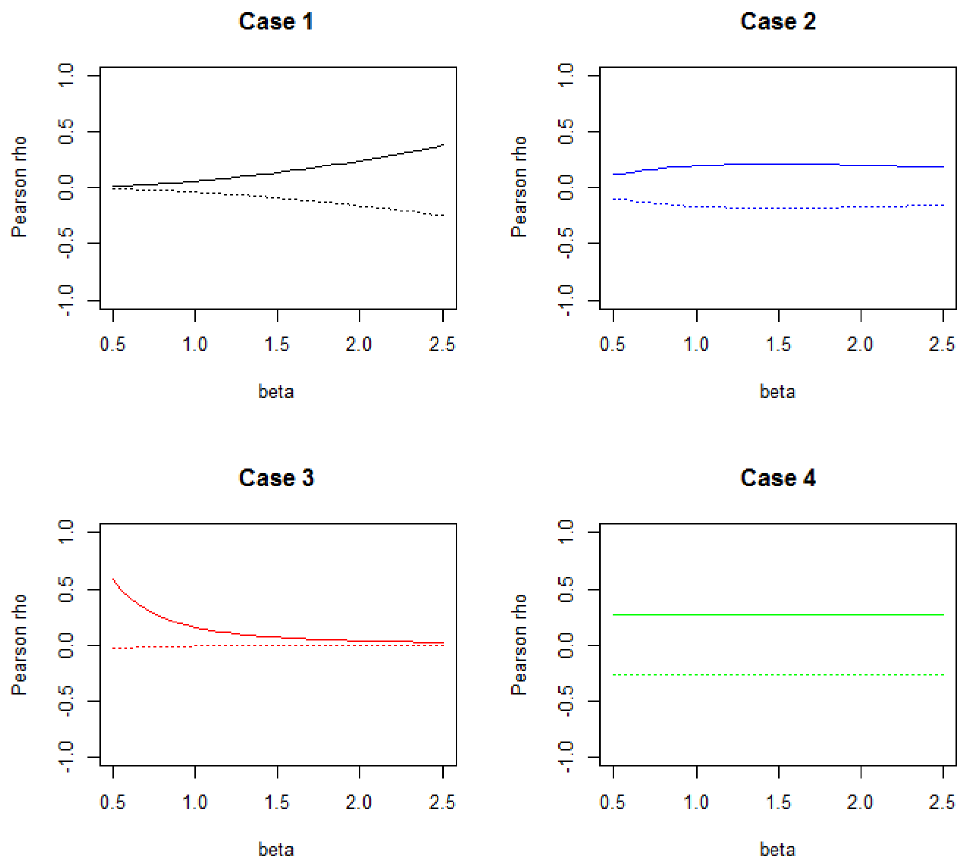


Figure 6.4.1: Peason's correlation coefficient for the same β and different kernel functions (Example 6.4.1.b).

6.5 Proofs

Proof of Lemma 6.2.2 We have

$$\begin{aligned} f(x,\beta,\underline{Q})^2 &= \left(\sum_{i=1}^{\infty} q_i w_i(x,\beta)\right)\left(\sum_{j=1}^{\infty} q_j w_j(x,\beta)\right) \\ &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} q_i q_j \frac{\beta^{i+j} x^{i+j-2} e^{-2\beta x}}{(i-1)!(j-1)!} \\ &= \sum_{i=1}^{\infty}\sum_{k=i}^{\infty} q_i q_{k+1-i} \frac{\beta^{k+1} x^{k-1} e^{-2\beta x}}{(i-1)!(k-i)!} \end{aligned}$$

$$= \beta \sum_{k=1}^{\infty} \sum_{i=1}^{k} \binom{k-1}{i-1} \frac{q_i q_{k+1-i}}{2^k} w_k(x, 2\beta). \qquad (6.5.1)$$

Also,

$$\begin{aligned} \mathbb{E}\left(f(X,\beta,\underline{Q})\right) &= \int_0^\infty f(x,\beta,\underline{Q}) f(x,\beta,\underline{Q}) \mathrm{d}x \\ &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} q_i q_j \frac{\beta^{i+j}}{(i-1)!(j-1)!} \int_0^\infty x^{i+j-2} e^{-2\beta x} \mathrm{d}x \\ &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} q_i q_j \frac{\beta^{i+j}}{(i-1)!(j-1)!} \frac{(i+j-2)!}{(2\beta)^{i+j-1}} \\ &= \beta \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \binom{i+j-2}{i-1} \frac{q_i q_j}{2^{i+j-1}}. \qquad (6.5.2) \end{aligned}$$

Therefore, dividing (6.5.1) by (6.5.2), we obtain

$$c(x,\beta,\underline{Q}) = \frac{\beta \sum_{k=1}^{\infty} \sum_{i=1}^{k} \binom{k-1}{i-1} \frac{q_i q_{k+1-i}}{2^k} w_k(x,2\beta)}{\beta \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \binom{i+j-2}{i-1} \frac{q_i q_j}{2^{i+j-1}}} = \sum_{k=1}^{\infty} v_k w_k(x,2\beta),$$

where the coefficients v_k are defined in (6.2.9). □

Proof of Lemma 6.2.8 We prove the result by induction. When $n = 1$, from Lemma 6.2.7 we have $F_{R_1}(y) = F_{Y_1}(y) = F_{X_1}(d_1) + H_{X_1}(y, d_1)$, i.e., formula (6.2.12) for $n = 1$. Assuming now that the formula (6.2.12) holds for $n-1$, for n we obtain

$$F_{R_n}(y) = \mathbb{P}(R_n \le y) = \mathbb{P}(R_n = 0) + \mathbb{P}(0 < R_{n-1} + Y_n \le y). \qquad (6.5.3)$$

But

$$\mathbb{P}(R_n = 0) = \mathbb{P}(Y_i = 0, i \in I_n) = \mathbb{P}(X_i \le d_i, i \in I_n) = \prod_{i=1}^{n} F_{X_i}(d_i),$$

while, using the induction hypothesis,

$$\begin{aligned} \mathbb{P}(0 < R_{n-1} + Y_n \le y) &= \mathbb{P}(Y_n = 0, 0 < R_{n-1} \le y) + \mathbb{P}(R_{n-1} = 0, 0 < Y_n \le y) \\ &\quad + \mathbb{P}(0 < R_{n-1}, 0 < Y_n, 0 < R_{n-1} + Y_n \le y) \\ &= F_{X_n}(d_n) \sum_{k=1}^{n-1} \sum_{1 \le j_1 < \ldots < j_k \le n-1} H_{X_{j_1} + \ldots + X_{j_k}}(y, d_{j_1}, \ldots, d_{j_k}) \end{aligned}$$

$$\times \prod_{i \in I_{n-1} \setminus \{j_1,\ldots,j_k\}} F_{X_i}(d_i) + \left(\prod_{i=1}^{n-1} F_{X_i}(d_i) \right) H_{X_n}(y, d_n) + I_{R_n},$$

where, with f_{Y_n} denoting the pdf of the r.v. Y_n,

$$\begin{aligned}
I_{R_n} &= \mathbb{P}\left(0 < R_{n-1}, 0 < Y_n, 0 < R_{n-1} + Y_n \leq y\right) \\
&= \int_0^y \left[\sum_{k=1}^{n-1} \sum_{1 \leq j_1 < \ldots < j_k \leq n-1} H_{X_{j_1}+\ldots+X_{j_k}}\left(y-u, d_{j_1}, \ldots, d_{j_k}\right) \prod_{i \in I_{n-1} \setminus \{j_1,\ldots,j_k\}} F_{X_i}(d_i) \right] f_{Y_n}(u)\, du \\
&= \sum_{k=1}^{n-1} \sum_{1 \leq j_1 < \ldots < j_k \leq n-1} \left(\prod_{i \in I_{n-1} \setminus \{j_1,\ldots,j_k\}} F_{X_i}(d_i) \right) \int_0^y H_{X_{j_1}+\ldots+X_{j_k}}\left(y-u, d_{j_1}, \ldots, d_{j_k}\right) H'_{X_n}(u, d_n)\, du \\
&= \sum_{k=1}^{n-1} \sum_{1 \leq j_1 < \ldots < j_k \leq n-1} \left(\prod_{i \in I_{n-1} \setminus \{j_1,\ldots,j_k\}} F_{X_i}(d_i) \right) \sum_{h_{j_1}=0}^{\infty} \ldots \sum_{h_{j_k}=0}^{\infty} \Delta_{h_{j_1}}\left(d_{j_1}, \beta, \underline{Q}_{j_1}\right) \cdot \ldots \cdot \Delta_{h_{j_k}}\left(d_{j_k}, \beta, \underline{Q}_{j_k}\right) \\
&\times \int_0^y W_{\sum_{i=1}^k h_{j_i}+k}\left(y-u, \beta\right) \sum_{h_n=0}^{\infty} \Delta_{h_n}\left(d_n, \beta, \underline{Q}_n\right) w_{h_n+1}(u, \beta) du \\
&= \sum_{k=1}^{n-1} \sum_{1 \leq j_1 < \ldots < j_k \leq n-1} \left(\prod_{i \in I_{n-1} \setminus \{j_1,\ldots,j_k\}} F_{X_i}(d_i) \right) H_{X_{j_1}+\ldots+X_{j_k}+X_n}\left(y, d_{j_1}, \ldots, d_{j_k}, d_n\right). \qquad (6.5.4)
\end{aligned}$$

For the last equality, apart the definition of H, we also used the fact that the convolution of two Erlang distributions having the same scale parameter is again an Erlang distribution with the same scale parameter, while its shape parameter equals the sum of the shape parameters of the convoluted distributions. Inserting now (6.5.4) into (6.5.4) and the result into (6.5.3) yields (6.2.12). To obtain the formula of U, we use

$$U_{\sum_{i=1}^k X_i, X_{k+1}}(y, d_1, \ldots, d_{k+1}) = \sum_{h_1=0}^{\infty} \ldots \sum_{h_{k+1}=0}^{\infty} \Delta_{h_1}\left(d_1, \beta, \underline{Q}_1\right) \cdot \ldots \cdot \Delta_{h_{k+1}}\left(d_{k+1}, \beta, \underline{Q}_{k+1}\right) J,$$

where

$$\begin{aligned}
J &= \int_y^{\infty} \int_0^s u w_{h_{k+1}+1}(u, \beta) w_{\sum_{i=1}^k h_i+k}\left(s-u, \beta\right) du ds \\
&= \frac{h_{k+1}+1}{\beta} \int_y^{\infty} \int_0^s w_{h_{k+1}+2}(u, \beta) w_{\sum_{i=1}^k h_i+k}\left(s-u, \beta\right) du ds \\
&= \frac{h_{k+1}+1}{\beta} \int_y^{\infty} w_{\sum_{i=1}^{k+1} h_i+k+2}\left(s, \beta\right) ds,
\end{aligned}$$

which inserted into the above formula of U immediately yields the result. □

Proof of Theorem 6.3.1 The joint density of $\boldsymbol{S} = (S_1, \ldots, S_n)$ is determined in terms of the joint density h of $\left(X_1^{(1)},\ldots,X_{k_1}^{(1)};\ldots;X_1^{(n)},\ldots,X_{k_n}^{(n)}\right)$ as follows

$$f_{\boldsymbol{S}}(s_1,\ldots,s_n)=\int\cdots\int_{\left\{\boldsymbol{x}=(x_1^{(1)},\ldots,x_{k_1}^{(1)},\ldots,x_1^{(n)},\ldots,x_{k_n}^{(n)})\middle|\sum_{j=1}^{k_i}x_j^{(i)}=s_i,i\in I_n\right\}} h(\boldsymbol{x})d\boldsymbol{x}. \qquad (6.5.5)$$

Based on (6.3.1),

$$\begin{aligned} h(\boldsymbol{x}) &= \prod_{i=1}^{n}\prod_{j=1}^{k_i} f_j^{(i)}\left(x_j^{(i)}\right)+\sum_{1\le a<b\le n}\sum_{s=1}^{k_a}\sum_{t=1}^{k_b}\alpha_{s,t}^{(a,b)}\left(\phi_s^{(a)}f_s^{(a)}\right)\left(x_s^{(a)}\right)\left(\phi_t^{(b)}f_t^{(b)}\right)\left(x_t^{(b)}\right)\prod_{\substack{i=1\\(i,j)\notin\{(a,s),(b,t)\}}}^{n}\prod_{j=1}^{k_i} f_j^{(i)}\left(x_j^{(i)}\right)\\ &+\sum_{a=1}^{n}\sum_{1\le s<t\le k_a}\alpha_{s,t}^{(a)}\left(\phi_s^{(a)}f_s^{(a)}\right)\left(x_s^{(a)}\right)\left(\phi_t^{(a)}f_t^{(a)}\right)\left(x_t^{(a)}\right)\prod_{\substack{i=1\\(i,j)\notin\{(a,s),(a,t)\}}}^{n}\prod_{j=1}^{k_i} f_j^{(i)}\left(x_j^{(i)}\right)\\ &= \prod_{i=1}^{n}\prod_{j=1}^{k_i} f_j^{(i)}\left(x_j^{(i)}\right)+\sum_{1\le a\le b\le n}\sum_{s=1}^{k_a}\sum_{t=T(s,a,b)}^{k_b}\alpha_{s,t}^{(a,b)}\prod_{i=1}^{n}\prod_{j=1}^{k_i}\tilde{f}_{j;s,t}^{(i;a,b)}\left(x_j^{(i)}\right), \end{aligned}$$

where $T(s,a,b)=\max\left\{1,(s+1)\mathbb{1}_{(a=b)}\right\}$, $\alpha_{s,t}^{(a,a)}=\alpha_{s,t}^{(a)}$ and, for $i\in I_n, j=1,\ldots,k_i$,

$$\tilde{f}_{j;s,t}^{(i;a,b)}(x)=\begin{cases}\left(\phi_j^{(i)}f_j^{(i)}\right)(x) & if\ (i,j)\in\{(a,s),(b,t)\}\\ f_j^{(i)}(x) & otherwise.\end{cases}.$$

Therefore, we can express (6.5.5) as

$$\begin{aligned} f_{\boldsymbol{S}}(\boldsymbol{s}) &= \prod_{i=1}^{n}\int\cdots\int_{\mathbb{R}^{k_i-1}}\left(\prod_{j=1}^{k_i-1}f_j^{(i)}\left(x_j^{(i)}\right)\right)f_{k_i}^{(i)}\left(s_i-\sum_{j=1}^{k_i-1}x_j^{(i)}\right)dx_1^{(i)}\ldots dx_{k_i-1}^{(i)}\\ &+\sum_{1\le a\le b\le n}\sum_{s=1}^{k_a}\sum_{t=T(s,a,b)}^{k_b}\alpha_{s,t}^{(a,b)}\prod_{i=1}^{n}\int\cdots\int_{\mathbb{R}^{k_i-1}}\left(\prod_{j=1}^{k_i-1}\tilde{f}_{j;s,t}^{(i;a,b)}\left(x_j^{(i)}\right)\right)\tilde{f}_{k_i;s,t}^{(i;a,b)}\left(s_i-\sum_{j=1}^{k_i-1}x_j^{(i)}\right)dx_1^{(i)}\ldots dx_{k_i-1}^{(i)}\\ &= \prod_{i=1}^{n}f_{S_i}(s_i)+\sum_{1\le a\le b\le n}\sum_{s=1}^{k_a}\sum_{t=T(s,a,b)}^{k_b}\alpha_{s,t}^{(a,b)}\prod_{i=1}^{n}\tilde{f}_{S_{i;s,t}^{(a,b)}}(s_i), \end{aligned}$$

with f_{S_i} and $\tilde{f}_{S_{i;s,t}^{(a,b)}}$ defined in (6.3.2). This completes the proof. □

Proof of Proposition 6.3.3 The df of $\boldsymbol{S}$ is determined in terms of the joint pdf h of $\boldsymbol{X}$ as follows

$$F_{\boldsymbol{S}}(\boldsymbol{s})=\mathbb{P}(S_1\leqslant s_1,\ldots,S_n\leqslant s_n)=\int\cdots\int_{\left\{\boldsymbol{x}=(x_1^{(1)},\ldots,x_{k_1}^{(1)},\ldots,x_1^{(n)},\ldots,x_{k_n}^{(n)})\middle|\sum_{j=1}^{k_i}x_j^{(i)}\le s_i,i\in I_n\right\}} h(\boldsymbol{x})d\boldsymbol{x}. \quad (6.5.6)$$

Starting from (6.3.1), the joint density of $\boldsymbol{X}$ is now given by

$$\begin{aligned} h(\boldsymbol{x}) &= \prod_{i=1}^{n}\prod_{j=1}^{k_i} f_j^{(i)}\left(x_j^{(i)}\right)\left[1+\sum_{1\le a\le b\le n}\sum_{s=1}^{k_a}\sum_{t=T(s,a,b)}^{k_b}\alpha_{s,t}^{(a,b)}\left(f_s^{(a)}\left(x_s^{(a)}\right)-\gamma_s^{(a)}\right)\left(f_t^{(b)}\left(x_t^{(b)}\right)-\gamma_t^{(b)}\right)\right] \\ &= \prod_{i=1}^{n}\prod_{j=1}^{k_i} f_j^{(i)}\left(x_j^{(i)}\right)\left(1+\sum_{1\le a\le b\le n}\sum_{s=1}^{k_a}\sum_{t=T(s,a,b)}^{k_b}\alpha_{s,t}^{(a,b)}\gamma_s^{(a)}\gamma_t^{(b)}\right) \\ &\quad -\sum_{1\le a\le b\le n}\sum_{s=1}^{k_a}\sum_{t=T(s,a,b)}^{k_b}\alpha_{s,t}^{(a,b)}\gamma_s^{(a)}\gamma_t^{(b)}\prod_{i=1}^{n}\prod_{j=1}^{k_i}\left(f_{j;s}^{(i;a)}\left(x_j^{(i)}\right)+f_{j;t}^{(i;b)}\left(x_j^{(i)}\right)\right) \\ &\quad +\sum_{1\le a\le b\le n}\sum_{s=1}^{k_a}\sum_{t=T(s,a,b)}^{k_b}\alpha_{s,t}^{(a,b)}\gamma_s^{(a)}\gamma_t^{(b)}\prod_{i=1}^{n}\prod_{j=1}^{k_i} f_{j;s,t}^{(i;a,b)}\left(x_j^{(i)}\right), \end{aligned}$$

where, for $i\in I_n, j=1,...,k_i$,

$$\begin{aligned} f_{j;s,t}^{(i;a,b)}(x) &= \begin{cases} \left(f_j^{(i)}(x)\right)^2 \Big/ \gamma_j^{(i)} & if\ \ (i,j)\in\{(a,s),(b,t)\} \\ f_j^{(i)}(x) & otherwise \end{cases}, \\ f_{j;s}^{(i;a)}(x) &= f_{j;s,s}^{(i;a,a)}(x). \end{aligned}$$

By Lemma 6.2.2, $\frac{\left(f_j^{(i)}\right)^2}{\gamma_j^{(i)}}$ is the pdf of a mixed Erlang distribution (with twice the scale parameter), therefore, using also the notation ξ_n from (6.3.3), one can write (6.5.6) as a sum-product of convolutions of mixed Erlang distributions as follows

$$\begin{aligned} F_{\boldsymbol{S}}(\boldsymbol{s}) &= \xi_n\prod_{i=1}^{n}\int_0^{s_i}\int_0^{s_i-x_1^{(i)}}...\int_0^{s_i-\sum_{j=1}^{k_i-2}x_j^{(i)}}\prod_{j=1}^{k_i-1} f_j^{(i)}\left(x_j^{(i)}\right)F_{k_i}^{(i)}\left(s_i-\sum_{j=1}^{k_i-1}x_j^{(i)}\right)dx_{k_i-1}^{(i)}...dx_1^{(i)} \\ &\quad -\sum_{1\le a\le b\le n}\sum_{s=1}^{k_a}\sum_{t=T(s,a,b)}^{k_b}\alpha_{s,t}^{(a,b)}\gamma_s^{(a)}\gamma_t^{(b)} \\ &\quad \times\prod_{i=1}^{n}\int_0^{s_i}\int_0^{s_i-x_1^{(i)}}...\int_0^{s_i-\sum_{j=1}^{k_i-2}x_j^{(i)}}\left[\prod_{j=1}^{k_i-1} f_{j;s}^{(i;a)}\left(x_j^{(i)}\right)F_{k_i;s}^{(i;a)}\left(s_i-\sum_{j=1}^{k_i-1}x_j^{(i)}\right)\right. \\ &\quad \left.+\prod_{j=1}^{k_i-1} f_{j;t}^{(i;b)}\left(x_j^{(i)}\right)F_{k_i;t}^{(i;b)}\left(s_i-\sum_{j=1}^{k_i-1}x_j^{(i)}\right)\right]dx_{k_i-1}^{(i)}...dx_1^{(i)}+\sum_{1\le a\le b\le n}\sum_{s=1}^{k_a}\sum_{t=T(s,a,b)}^{k_b}\alpha_{s,t}^{(a,b)}\gamma_s^{(a)}\gamma_t^{(b)} \\ &\quad \times\prod_{i=1}^{n}\int_0^{s_i}\int_0^{s_i-x_1^{(i)}}...\int_0^{s_i-\sum_{j=1}^{k_i-2}x_j^{(i)}}\prod_{j=1}^{k_i-1} f_{j;s,t}^{(i;a,b)}\left(x_j^{(i)}\right)F_{k_i;s,t}^{(i;a,b)}\left(s_i-\sum_{j=1}^{k_i-1}x_j^{(i)}\right)dx_{k_i-1}^{(i)}...dx_1^{(i)}. \end{aligned} \tag{6.5.7}$$

Since $\beta_{k_n}^{(n)}\ge\beta_j^{(i)},\forall j=1,\ldots,k_i, i\in I_n$, by Lemma 6.2.4 each ith mixed Erlang component of (6.5.7) can be transformed into a new mixed Erlang distribution with a common scale parameter $2\beta_{k_n}^{(n)}$. In addition, according to Remark 6.2.6, the convolution of mixed Erlang distributions belongs to the class of mixed Erlang distributions. Therefore, (6.5.7) can be expressed as a sum-product of mixed Erlang df's as follows

$$\begin{aligned} F_{\boldsymbol{S}}(s) &= \xi_n \prod_{j=1}^{n} F_{S_j^{(1)}}(s_j) - \sum_{1\le a\le b\le n} \sum_{s=1}^{k_a} \sum_{t=T(s,a,b)}^{k_b} \alpha_{s,t}^{(a,b)} \gamma_s^{(a)} \gamma_t^{(b)} \\ &\quad \times \left(\prod_{j=1}^{n} F_{S_{j;s}^{(2;a)}}(s_j) + \prod_{j=1}^{n} F_{S_{j;t}^{(2;b)}}(s_j) - \prod_{j=1}^{n} F_{S_{j;s,t}^{(3;a,b)}}(s_j) \right), \end{aligned}$$

where $S_j^{(1)}, S_{j;s}^{(2;a)}, S_{j;s,t}^{(3;a,b)}$ are defined by (6.3.4)-(6.3.6). Thus the proof is complete. □

Proof of Proposition 6.3.4 The distribution of R_n can be expressed in terms of the distribution of $\boldsymbol{S}$ as follows

$$\begin{aligned} F_{R_n}(y) &= \mathbb{P}\left(\bigcap_{i=1}^{n}(T_i=0)\right) + \sum_{k=1}^{n} \sum_{1\le j_1<\ldots<j_k\le n} \mathbb{P}\left(\bigcap_{i\in I_n\setminus\{j_1,\ldots,j_k\}} (T_i=0) \bigcap_{i=1}^{k} (T_{j_i}>0) \bigcap \left(\sum_{i=1}^{k} T_{j_i} \le y \right) \right) \\ &= \mathbb{P}\left(\bigcap_{i=1}^{n}(S_i\le d_i)\right) + \sum_{k=1}^{n} \sum_{1\le j_1<\ldots<j_k\le n} \mathbb{P}\left(\bigcap_{i\in I_n\setminus\{j_1,\ldots,j_k\}} (S_i\le d_i) \bigcap_{i=1}^{k} (S_{j_i}>d_{j_i}) \bigcap \left(\sum_{i=1}^{k} (S_{j_i}-d_{j_i}) \le y \right) \right). \end{aligned}$$

Considering now the df of $\boldsymbol{S}$ given in Proposition 6.3.3 and the definition of H from Lemma 6.2.8, we can rewrite F_{R_n} in the form (6.3.7), which completes the proof. □

Proof of Proposition 6.3.5 For simplicity, we shall prove the case $l=n$, the proof for a general l being similar, but with a notation more complicated. To use (6.2.13), we must evaluate

$$\mathbb{E}\left(T_n \mathbb{1}_{\{R_n>x_p\}}\right) = J_1 + J_2, \tag{6.5.8}$$

where

$$\begin{aligned} J_1 &= \int_{x_p}^{\infty} u f_{T_n,\{R_{n-1}=0\}}(u)\, du \text{ with } f_{T_n,\{R_{n-1}=0\}}(u) = \frac{\partial}{\partial u}\mathbb{P}(0<T_n\le u, R_{n-1}=0), \\ J_2 &= \int_{x_p}^{\infty}\int_0^s u f_{T_n,R_{n-1}}(u,s-u)\, du ds \text{ with } f_{T_n,R_{n-1}}(u,v) = \frac{\partial^2}{\partial u\partial v}\mathbb{P}(0<T_n\le u, 0<R_{n-1}\le v). \end{aligned}$$

We shall now use Proposition 6.3.4 and the notation from Lemma 6.2.7. We have

$$\begin{aligned}
J_1 &= \int_{x_p}^{\infty} u \frac{\partial}{\partial u}\left[\xi_n \left(\prod_{i=1}^{n-1} F_{S_i^{(1)}}(d_i)\right) H_{S_n^{(1)}}(u,d_n) - \sum_{1\le a\le b\le n}\sum_{s=1}^{k_a}\sum_{t=T(s,a,b)}^{k_b} \alpha_{s,t}^{(a,b)}\gamma_s^{(a)}\gamma_t^{(b)}\right.\\
&\quad \left.\times\left(H_{S_{n;s}^{(2;a)}}(u,d_n)\prod_{i=1}^{n-1}F_{S_{i;s}^{(2;a)}}(d_i) + H_{S_{n;t}^{(2;b)}}(u,d_n)\prod_{i=1}^{n-1}F_{S_{i;t}^{(2;b)}}(d_i) - H_{S_{n;s,t}^{(3;a,b)}}(u,d_n)\prod_{i=1}^{n-1}F_{S_{i;s,t}^{(3;a,b)}}(d_i)\right)\right] du\\
&= \xi_n U_{S_n^{(1)}}(x_p,d_n)\prod_{i=1}^{n-1}F_{S_i^{(1)}}(d_i) - \sum_{1\le a\le b\le n}\sum_{s=1}^{k_a}\sum_{t=T(s,a,b)}^{k_b}\alpha_{s,t}^{(a,b)}\gamma_s^{(a)}\gamma_t^{(b)}\\
&\quad\times\left(U_{S_{n;s}^{(2;a)}}(x_p,d_n)\prod_{i=1}^{n-1}F_{S_{i;s}^{(2;a)}}(d_i) + U_{S_{n;t}^{(2;b)}}(x_p,d_n)\prod_{i=1}^{n-1}F_{S_{i;t}^{(2;b)}}(d_i) - U_{S_{n;s,t}^{(3;a,b)}}(x_p,d_n)\prod_{i=1}^{n-1}F_{S_{i;s,t}^{(3;a,b)}}(d_i)\right).
\end{aligned}$$

On the other hand, a reasoning similar with the one in the proof of Proposition 6.3.4 yields

$$\begin{aligned}
f_{T_n,R_{n-1}}(u,v) &= \frac{\partial^2}{\partial u\partial v}\sum_{k=1}^{n-1}\sum_{1\le j_1<\ldots<j_k\le n-1}\mathbb{P}\left((0<T_n\le u)\bigcap_{i\in I_{n-1}\setminus\{j_1,\ldots,j_k\}}(T_i=0)\bigcap_{i=1}^{k}(T_{j_i}>0)\bigcap\left(\sum_{i=1}^{k}T_{j_i}\le v\right)\right)\\
&= \frac{\partial^2}{\partial u\partial v}\sum_{k=1}^{n-1}\sum_{1\le j_1<\ldots<j_k\le n-1}\left[\xi_n H_{S_n^{(1)}}(u,d_n)\, H_{\sum_{i=1}^{k}S_{j_i}^{(1)}}(v,d_{j_1},..,d_{j_k})\prod_{i\in I_{n-1}\setminus\{j_1,\ldots,j_k\}}F_{S_i^{(1)}}(d_i)\right.\\
&\quad - \sum_{1\le a\le b\le n}\sum_{s=1}^{k_a}\sum_{t=T(s,a,b)}^{k_b}\alpha_{s,t}^{(a,b)}\gamma_s^{(a)}\gamma_t^{(b)}\\
&\quad\times\left(H_{S_{n;s}^{(2;a)}}(u,d_n)\, H_{\sum_{i=1}^{k}S_{j_i;s}^{(2;a)}}(v,d_{j_1},..,d_{j_k})\prod_{i\in I_{n-1}\setminus\{j_1,\ldots,j_k\}}F_{S_{i;s}^{(2;a)}}(d_i)\right.\\
&\quad + H_{S_{n;t}^{(2;b)}}(u,d_n)\, H_{\sum_{i=1}^{k}S_{j_i;t}^{(2;b)}}(v,d_{j_1},..,d_{j_k})\prod_{i\in I_{n-1}\setminus\{j_1,\ldots,j_k\}}F_{S_{i;t}^{(2;b)}}(d_i)\\
&\quad \left.\left. - H_{S_{n;s,t}^{(3;a,b)}}(u,d_n)\, H_{\sum_{i=1}^{k}S_{j_i;s,t}^{(3;a,b)}}(v,d_{j_1},..,d_{j_k})\prod_{i\in I_{n-1}\setminus\{j_1,\ldots,j_k\}}F_{S_{i;s,t}^{(3;a,b)}}(d_i)\right)\right],
\end{aligned}$$

from where, using the notation introduced in Lemma 6.2.8, we obtain

$$\begin{aligned}
J_2 &= \sum_{k=1}^{n-1}\sum_{1\le j_1<\ldots<j_k\le n-1}\left[\xi_n U_{\sum_{i=1}^{k}S_{j_i}^{(1)},S_n^{(1)}}(x_p,d_{j_1},..,d_{j_k},d_n)\prod_{i\in I_{n-1}\setminus\{j_1,\ldots,j_k\}}F_{S_i^{(1)}}(d_i)\right.\\
&\quad - \sum_{1\le a\le b\le n}\sum_{s=1}^{k_a}\sum_{t=T(s,a,b)}^{k_b}\alpha_{s,t}^{(a,b)}\gamma_s^{(a)}\gamma_t^{(b)}\left(U_{\sum_{i=1}^{k}S_{j_i;s}^{(2;a)},S_{n;s}^{(2;a)}}(x_p,d_{j_1},..,d_{j_k},d_n)\right.\\
&\quad\times\prod_{i\in I_{n-1}\setminus\{j_1,\ldots,j_k\}}F_{S_{i;s}^{(2;a)}}(d_i) + U_{\sum_{i=1}^{k}S_{j_i;t}^{(2;b)},S_{n;t}^{(2;b)}}(x_p,d_{j_1},..,d_{j_k},d_n)\prod_{i\in I_{n-1}\setminus\{j_1,\ldots,j_k\}}F_{S_{i;t}^{(2;b)}}(d_i)\\
&\quad\left.\left. - U_{\sum_{i=1}^{k}S_{j_i;s,t}^{(3;a,b)},S_{n;s,t}^{(3;a,b)}}(x_p,d_{j_1},..,d_{j_k},d_n)\prod_{i\in I_{n-1}\setminus\{j_1,\ldots,j_k\}}F_{S_{i;s,t}^{(3;a,b)}}(d_i)\right)\right].
\end{aligned}$$

Inserting now the formulas of J_1, J_2 into (6.5.8) and the result into (6.2.13) yields the formula of C_n. Thus, the proof is complete. □

Proof of Proposition 6.3.8 Without loss of generality, we assume that $j < k$. To prove the stated formula of C_j, we need to find

$$\mathbb{E}\left(X_j \mathbb{1}_{\{S>s_p\}}\right) = \int_{s_p}^{\infty}\int_0^s\int_0^{s-x_1}\dots\int_0^{s-\sum_{i=1}^{k-2}x_i} x_j h\left(x_1, x_2, \dots, x_{k-1}, s-\sum_{i=1}^{k-1}x_i\right)\mathrm{d}x_{k-1}\dots\mathrm{d}x_2\mathrm{d}x_1\mathrm{d}s, \tag{6.5.9}$$

with h as defined in (6.2.4) with kernels (6.2.2). We denote $\xi_k = 1+\sum_{1\le a<b\le k}\alpha_{a,b}\gamma_a\gamma_b$ and first evaluate $x_j h(\boldsymbol{x})$ as

$$\begin{aligned} x_j h(\boldsymbol{x}) &= x_j \prod_{i=1}^k f_i(x_i)\left[1+\sum_{1\le a<b\le k}\alpha_{a,b}\left(f_a(x_a)f_b(x_b)-\gamma_a f_b(x_b)-\gamma_b f_a(x_a)+\gamma_a\gamma_b\right)\right] \\ &= \xi_k\mu_j\prod_{i=1}^k f_{i,j}(x_i) - \sum_{1\le a<b\le k}\alpha_{a,b}\gamma_a\gamma_b\left[\varphi_{j;b}\prod_{i=1}^k f_{i,j;b}(x_i)+\varphi_{j;a}\prod_{i=1}^k f_{i,j;a}(x_i)-\varphi_{j;a,b}\prod_{i=1}^k f_{i,j;a,b}(x_i)\right], \end{aligned}$$

where $\varphi_{j;a,b}$ and $\varphi_{j;a}$ are defined in (6.3.10), and, for $i \in I_k$, we define the following pdf's

$$\begin{aligned} f_{i,j}(x) &= \begin{cases} f_i(x) & \text{if} \quad i \neq j \\ \frac{x f_j(x)}{\mu_j} & \text{if} \quad i = j \end{cases}, \\ f_{i,j;a,b}(x) &= \begin{cases} f_i(x) & \text{if} \quad i \notin \{j,a,b\} \\ \frac{x f_j(x)}{\mu_j} & \text{if} \quad i = j \notin \{a,b\} \\ \frac{f_i^2(x)}{\gamma_i} & \text{if} \quad i \in \{a,b\}, i \neq j \\ \frac{x\left(f_j^2(x)/\gamma_j\right)}{\widetilde{\mu_j}} & \text{if} \quad i = j \in \{a,b\} \end{cases}, \text{ while } f_{i,j;a}(x) = f_{i,j;a,a}(x). \end{aligned}$$

According to Lemmas 6.2.2, 6.2.3 and 6.2.4, the just defined pdf's can be regarded of mixed Erlang type with parameter $2\beta_k$. Thus, (6.5.9) becomes

$$\begin{aligned} \mathbb{E}\left(X_j \mathbb{1}_{\{S>s_p\}}\right) &= \xi_k\mu_j\int_{s_p}^{\infty}\int_0^s\int_0^{s-x_1}\dots\int_0^{s-\sum_{i=1}^{k-2}x_i} f_{k,j}\left(s-\sum_{i=1}^{k-1}x_i\right)\prod_{i=1}^{k-1}f_{i,j}(x_i)\mathrm{d}x_{k-1}\dots\mathrm{d}x_2\mathrm{d}x_1\mathrm{d}s \\ &\quad - \sum_{1\le a<b\le k}\alpha_{a,b}\gamma_a\gamma_b\int_{s_p}^{\infty}\int_0^s\int_0^{s-x_1}\dots\int_0^{s-\sum_{i=1}^{k-2}x_i}\left[\varphi_{j;b}f_{k,j;b}\left(s-\sum_{i=1}^{k-1}x_i\right)\prod_{i=1}^{k-1}f_{i,j;b}(x_i)\right. \\ &\quad \left. +\varphi_{j;a}f_{k,j;a}\left(s-\sum_{i=1}^{k-1}x_i\right)\prod_{i=1}^{k-1}f_{i,j;a}(x_i)-\varphi_{j;a,b}f_{k,j;a,b}\left(s-\sum_{i=1}^{k-1}x_i\right)\prod_{i=1}^{k-1}f_{i,j;a,b}(x_i)\right] \\ &\quad \times\mathrm{d}x_{k-1}\dots\mathrm{d}x_2\mathrm{d}x_1\mathrm{d}s, \end{aligned}$$

i.e., the sum of four integrals consisting of tails of convolutions of mixed Erlang distributions, which leads to the following four mixed Erlang distributions, respectively,

$$ME\left(2\beta_k, \underline{\Pi}\left(\Psi\left(\widetilde{M}_j(\underline{Q}_1)\right), \ldots, \Psi\left(\widetilde{M}_j(\underline{Q}_k)\right)\right)\right), ME\left(2\beta_k, \underline{\Pi}\left(\Psi\left(\widetilde{M}_{j;b}(\underline{Q}_1)\right), \ldots, \Psi\left(\widetilde{M}_{j;b}(\underline{Q}_k)\right)\right)\right),$$

$$ME\left(2\beta_k, \underline{\Pi}\left(\Psi\left(\widetilde{M}_{j;a}(\underline{Q}_1)\right), \ldots, \Psi\left(\widetilde{M}_{j;a}(\underline{Q}_k)\right)\right)\right), ME\left(2\beta_k, \underline{\Pi}\left(\Psi\left(\widetilde{M}_{j;a,b}(\underline{Q}_1)\right), \ldots, \Psi\left(\widetilde{M}_{j;a,b}(\underline{Q}_k)\right)\right)\right),$$

where $\widetilde{M}_j, \widetilde{M}_{j;a}$ and $\widetilde{M}_{j;a,b}$ are defined in (6.3.11). Then formula (6.3.8) holds with the mixing coefficients $z_{i,j}$ defined in (6.3.9). This completes the proof. □

Chapter 7

On the evaluation of multivariate compound distributions with continuous severity distributions and Sarmanov's counting distribution

This chapter is based on M. Tamraz and R. Vernic: On the evaluation of multivariate compound distributions with continuous severity distributions and Sarmanov's counting distribution, published in *ASTIN Bulletin*, 1-30, 2018.

7.1 Introduction

Used to model the aggregate claims of a portfolio, the univariate collective model is represented as

$$S = \sum_{l=0}^{N} X_l, \tag{7.1.1}$$

where N is the random variable (r.v.) number of claims and $(X_l)_{l\geq 1}$ are the corresponding non-negative r.v.s claim sizes with $X_0 = 0$. The classical hypotheses that provide the tractability of this model are independent, identically distributed (i.i.d.) discrete claim sizes, also independent of N. There is a large amount of literature related to the evaluation of the compound distribution of S under these assumptions, see, e.g., Klugman et al. [73] for different methods, and Sundt and Vernic [113] for a survey of the recursive methods. However, in practice, the claim sizes (severity) distribution is rather of continuous type, hence the usual approach consists in two steps: in the first step, the severity distribution is discretized, while

in the second step, a specific method is applied to the resulting discrete compound distribution. In this case, a special attention must be paid to the choice of the discretization span: a large span can generate important errors, while a very small span can lead to a very long running time, especially in the multivariate case. In this respect, it is unfortunate that closed-type formulas for compound distributions with continuous type claim sizes are so scarce; in the univariate case, apart the Gamma severity distribution (which also includes the well-known exponential case) leading to the so-called Tweedie compound distribution (see, e.g., Dunn and Smyth [33]), we mention the recent work of Sarabia et al. [106], who went even further on by considering a Pareto type dependency between the aggregated claim sizes.
In this paper, we propose closed-type formulas for some multivariate compound distributions with Sarmanov counting distribution and Erlang severity distributions; furthermore, inspired by Sarabia et al. [106], we also include some dependency between the claim sizes. Our formulas are expressed mainly in terms of the special hypergeometric function already implemented in the existing mathematical software, hence making the related calculations numerically feasible without involving other techniques. More precisely, we deal with a multivariate extension of model (7.1.1), i.e., for $m \geq 2$, we consider

$$(S_1, ..., S_m) = \left(\sum_{l=0}^{N_1} X_{1l}, ..., \sum_{l=0}^{N_m} X_{ml} \right), \tag{7.1.2}$$

where N_j denotes the number of claims of type j and $(X_{jl})_{l\geq1}$ the corresponding claim sizes, where, by convention, $X_{j0} = 0, 0 \leq j \leq m$. This model corresponds to the situation where we have m different types of claims generated by some related events, hence the claim numbers $(N_j)_{j=1}^{m}$ are dependent. The model has been studied mostly under the assumptions that the claims of type j are i.i.d., independent of the claim numbers and independent of the claims of type $k, \forall j \neq k$ (see, e.g., Sundt and Vernic [113], Jin and Ren [65] or Robe-Voinea and Vernic [103]). We shall call by "inside-type independency" the independency assumption between the claims of same type, while by "between-types independency" we designate the independency assumptions between claims of different types. Then, similarly with Sarabia et al. [106], we shall relax the inside-type independency condition by considering a certain type of dependency in each set of claims $(X_{jl})_{l\geq1}$. Note that in the univariate case, dependency between the individual risks has already been considered especially in the individual model, see, e.g., Goovaerts and Dhaene [53], Genest et al. [50], Denuit et al. [29] and the references therein.
Regarding the claim sizes distributions, we choose the Erlang distribution in the

inside-type independency case and, when relaxing this condition, the same multivariate type II Pareto distribution as in Sarabia et al. [106]. This choice has been driven by the interest in obtaining closed formulas numerically computable with the existing software, but also by the fact that both distributions have been intensively studied in the actuarial literature lately; for the Erlang distribution see, e.g., Willmot and Lin [123] or Willmot and Woo [124], while for the multivariate type II Pareto distribution see Asimit et al. [8] or the generalizations in Asimit et al. [5], Guillén et al. [56].

Therefore, this paper is structured as follows: in Section 2, we introduce some notation and recall several special functions and distributions that will be used in the sequel. In Section 3, we present closed-type formulas for univariate compound distributions with Erlang severity distribution, while in Section 4 we extend these formulas to multivariate compound distributions with Sarmanov counting distribution; moreover, Section 4 is divided into two subsections corresponding to the cases with and without inside-type independency. In the bivariate case, a special attention is paid to the correlation coefficient of the resulting compound distribution, which is expressed in terms of the correlation coefficient of the original counting distribution, and results smaller than the last one. To illustrate the applicability, efficiency and importance of the derived formulas, in Section 5 we present a numerical example in which we compare the cumulative distribution functions (cdf's) obtained for a particular bivariate compound distribution by using the closed-type formulas and the usual recursion-discretization approach. The paper ends with some conclusions and future work, followed by an Appendix containing the proofs.

7.2 Preliminaries

7.2.1 Notation, definitions and useful formulas

In connection with the univariate collective model (7.1.1), we denote the probability mass function (pmf) of the discrete r.v. N by p, while h denotes the probability density function (pdf) of the claim sizes, which are assumed to be positive, continuous and identically distributed (i.d.), not necessarily independent. Therefore, the distribution of S is called *compound* with counting distribution p and severity distribution h; we denote it by $p \vee h$. Then, letting $h^{(n)}$ denote the pdf of the sum

$\sum_{i=0}^{n} X_i$, where $h^{(0)}(x) = \begin{cases} 1, x = 0 \\ 0, otherwise \end{cases}$, it holds that (see Sarabia et al. [106])

$$(p \vee h)(x) = \sum_{n=0}^{\infty} p(n) h^{(n)}(x), x \geq 0, \tag{7.2.1}$$

hence $(p \vee h)(0) = p(0)$. Under the classical assumption that the claim amounts are also independent, this formula reduces to the well-known one

$$(p \vee h)(x) = \sum_{n=0}^{\infty} p(n) h^{*n}(x), \tag{7.2.2}$$

where h^{*n} denotes the n-fold convolution of h.
To simplify the writing in the multivariate case, we denote $\mathbf{X} = (X_1, ..., X_m)$ or $\mathbf{x} = (x_1, ..., x_m)$, $\overline{1, m} = \{1, 2, ..., m\}$; moreover, $\mathbf{0}$ denotes the 0-vector, while $\mathbf{x} - \mathbf{y}$ and $\mathbf{x} \geq \mathbf{y}$ are considered component-wise. In what concerns the model (7.1.2), the pmf of the random vector consisting of the (dependent) claim numbers $(N_1, ..., N_m)$ is still denoted by p, while h_j denotes the pmf of the i.d. continuous positive claim amounts of type j, and $\mathbf{h} = (h_1, ..., h_m)$. With this notation, assuming that both inside-type and between-types independency conditions hold, the compound pdf $(p \vee \mathbf{h})$ of $\mathbf{S} = (S_1, ..., S_m)$ can be written as

$$(p \vee \mathbf{h})(\mathbf{x}) = \sum_{n_1=0}^{\infty} ... \sum_{n_m=0}^{\infty} p(\mathbf{n}) \prod_{j=1}^{m} h_j^{*n_j}(x_j), \ \mathbf{x} \geq \mathbf{0}. \tag{7.2.3}$$

However, relaxing the inside-type independency assumption while keeping the between-types independency condition, the distribution of $\mathbf{S}$ becomes (we omit the proof being similar with the one in Sarabia et al. [106])

$$(p \vee \mathbf{h})(\mathbf{x}) = \sum_{n_1=0}^{\infty} ... \sum_{n_m=0}^{\infty} p(\mathbf{n}) \prod_{j=1}^{m} h_j^{(n_j)}(x_j), \ \mathbf{x} \geq \mathbf{0}. \tag{7.2.4}$$

It follows that $(p \vee \mathbf{h})(\mathbf{0}) = p(\mathbf{0})$, and, marginally, e.g.,

$$\begin{aligned} (p \vee \mathbf{h})(x_1, 0, ..., 0) &= \sum_{n_1=1}^{\infty} p(n_1, 0, ..., 0) h_1^{(n_1)}(x_1), x_1 > 0, \\ (p \vee \mathbf{h})(0, x_2, ..., x_m) &= \sum_{n_2=1}^{\infty} ... \sum_{n_m=1}^{\infty} p(0, n_2, ..., n_m) \prod_{j=2}^{m} h_j^{(n_j)}(x_j), x_2, ..., x_m > 0. \end{aligned}$$

Note that in both univariate and multivariate cases, the formulas (7.2.1) and (7.2.4) can be used as definitions for a more general compounding operation involving a discrete function $p : \mathbb{N}^m \to \mathbb{R}, m \geq 1$, which is not necessarily a pmf (see also, Vernic [119]); however, we keep the assumption that the functions h and h_js are pdf-s.

We shall now recall several special functions. The Laplace transform of a r.v. X is defined by $\mathcal{L}_X(t) = \mathbb{E}\left[e^{-tX}\right]$.

Spence's function or dilogarithm, which is a particular case of the polylogarithm function, is defined for $|t| < 1$ by the power series $Li_2(t) = \sum_{k=1}^{\infty} t^k/k^2$.

We let

$$I_0(x) = \sum_{n=0}^{\infty} (x/2)^{2n} \frac{1}{(n!)^2}, I_1(x) = \sum_{n=0}^{\infty} (x/2)^{2n+1} \frac{1}{n!\,(n+1)!}$$

denote the modified Bessel functions of first and, respectively, second kind.

Based on the Pochhammer symbol $(a)_{(n)} = a(a+1)\times\ldots\times(a+n-1), n \geq 1, (a)_{(0)} = 1$, the generalized hypergeometric function ${}_rF_q$ is defined by

$$ {}_rF_q(\{a_1, ..., a_r\}, \{b_1, ..., b_q\}; z) = 1 + \sum_{n=1}^{\infty} \frac{(a_1)_{(n)} \times ... \times (a_r)_{(n)}}{(b_1)_{(n)} \times ... \times (b_q)_{(n)}} \frac{z^n}{n!}. \tag{7.2.5}$$

We shall need the following result (its proof is given in the Appendix). By convention, an empty product equals 1.

Lemma 7.2.1. *For $n \in \mathbb{N}$ and $k \in \mathbb{N}^*$, it holds that*

$$i)\ \frac{1}{((n+1)k-1)!\,(n+1)} = \frac{1}{(k-1)!k^{nk}\left(\frac{k+1}{k}\right)_{(n)}\left(\frac{k+2}{k}\right)_{(n)} \times ... \times \left(\frac{k+(k-1)}{k}\right)_{(n)} (2)_{(n)}}, \tag{7.2.6}$$

$$ii)\ \frac{1}{(nk)!} = \frac{1}{n!k^{nk}\left(\frac{1}{k}\right)_{(n)}\left(\frac{2}{k}\right)_{(n)} \times ... \times \left(\frac{k-1}{k}\right)_{(n)}}. \tag{7.2.7}$$

Note that due to the convention, when $k = 1$ formula (7.2.6) becomes $\frac{1}{n!(n+1)} = \frac{1}{(2)_{(n)}}$, while formula (7.2.7) yields the identity $\frac{1}{n!} = \frac{1}{n!}$.

7.2.2 Some distributions

We shall now recall some distributions needed in the sequel.

Univariate distributions

In the discrete case, we shall use the well known Poisson distribution $Po(\mu), \mu > 0$, the negative binomial distribution $NB(r, q), r > 0, q \in (0, 1)$, with pmf $\frac{\Gamma(r+n)}{n!\Gamma(r)} q^r (1-q)^n$, $n \in \mathbb{N}$, expected value $\frac{r(1-q)}{q}$ and variance $\frac{r(1-q)}{q^2}$, and the logarithmic distribution

$Log\left(\theta\right)$, $\theta \in (0,1)$, with pmf $\frac{-1}{\ln(1-\theta)}\frac{\theta^n}{n}, n \geq 1$, expected value $\frac{-\theta}{(1-\theta)\ln(1-\theta)}$ and variance $\frac{-\theta(\theta+\ln(1-\theta))}{(1-\theta)^2\ln^2(1-\theta)}$.

In the continuous case, we recall the Gamma distribution $Ga\left(\alpha,\beta\right), \alpha,\beta > 0$, whose pdf is given by $f\left(x\right) = \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}, x > 0$, expected value $\frac{\alpha}{\beta}$ and variance $\frac{\alpha}{\beta^2}$, where $\Gamma\left(\cdot\right)$ denotes the Gamma function; it is well-known that the n-fold convolution of $Ga\left(\alpha,\beta\right)$ is still Gamma distributed, i.e., $Ga\left(n\alpha,\beta\right)$. In particular, when $\alpha = k \in \mathbb{N}^*$, the Gamma distribution is called Erlang that we shall denote by $Erlang\left(k,\beta\right)$. Also, $Ga\left(1,\beta\right)$ is the exponential distribution denoted by $Exp\left(\beta\right)$.

Another distribution that we shall encounter in the following is the beta distribution of the second kind, also called beta prime, inverted beta or Pearson type VI distribution (for details on this distribution, see, e.g., Kleiber and Kotz [71]). The pdf of this distribution is given by $f\left(x\right) = \frac{x^{\beta-1}}{B(\alpha,\beta)(1+x)^{\alpha+\beta}}, x > 0$, where $\alpha,\beta > 0$ and $B\left(\alpha,\beta\right) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ denotes the beta function. Introducing also a scale parameter $\sigma > 0$, the pdf becomes

$$f\left(x\right) = \frac{x^{\beta-1}}{\sigma^\beta B\left(\alpha,\beta\right)\left(1+x/\sigma\right)^{\alpha+\beta}}, x > 0.$$

We denote this distribution by $B_{II}\left(\beta,\alpha,\sigma\right)$ and note that it can be obtained as the distribution of the ratio of two independent r.v.s, i.e., as $\sigma\frac{Y}{Z}$, where $Y \sim Ga\left(\beta,1\right)$ and $Z \sim Ga\left(\alpha,1\right)$.

Moreover, the ratio $\sigma\frac{Y}{Z}$ of two independent r.v.s, where $\sigma > 0, Z \sim Ga\left(\alpha,1\right)$ and $Y \sim Exp\left(1\right)$, follows a Pareto distribution $Pa\left(\alpha,\sigma\right)$ with pdf $\frac{\alpha}{\sigma}\left(1+\frac{x}{\sigma}\right)^{-\alpha-1}, x > 0$, expected value $\frac{\sigma}{\alpha-1}, \alpha > 1$, and variance $\frac{\alpha\sigma^2}{(\alpha-1)^2(\alpha-2)}, \alpha > 2$.

Multivariate type II Pareto distribution

Starting from m i.i.d. r.v.s $Y_1,...,Y_m$ exponentially $Exp\left(1\right)$ distributed and independent of the r.v. $Z \sim Ga\left(\alpha,1\right), \alpha > 0$, the random vector defined by $\mathbf{X} = \left(\sigma\frac{Y_1}{Z},...,\sigma\frac{Y_m}{Z}\right), \sigma > 0$, follows an m-variate Pareto of type II distribution with pdf

$$f\left(\mathbf{x}\right) = \frac{\Gamma\left(\alpha+m\right)}{\Gamma\left(\alpha\right)\sigma^m}\left(1+\frac{1}{\sigma}\sum_{i=1}^{m}x_i\right)^{-\alpha-m}, x_1,...,x_m > 0.$$

We denote this distribution by $PaII_m\left(\alpha,\sigma\right)$ and note that its marginals are all identically distributed $Pa\left(\alpha,\sigma\right)$, the covariance between components is $cov(X_i,X_j) = \frac{\sigma^2}{(\alpha-1)^2(\alpha-2)}, \alpha > 2, i \neq j$, and the correlation $\rho = \frac{1}{\alpha}$. Sarabia et al. [106] also showed that the sum of the components $\sum_{i=1}^{m}X_i$ follows the beta distribution of the second kind, $B_{II}\left(m,\alpha,\sigma\right)$. For more details on the Pareto distribution see Arnold [3].

Sarmanov's multivariate distribution

We recall that the random vector $\mathbf{N} = (N_1, ..., N_m)$ follows an $m-$variate discrete Sarmanov distribution with joint pmf given for $\mathbf{n} \in \mathbb{N}^m$ by (see, e.g.,Kotz et al. [75])

$$p(\mathbf{n}) = \left(\prod_{l=1}^{m} p_l(n_l)\right)\left(1 + \sum_{k=2}^{m} \sum_{1\le j_1<...<j_k\le m} \omega_{j_1...j_k}\ \phi_{j_1}(n_{j_1}) \times ... \times \phi_{j_k}(n_{j_k})\right), \tag{7.2.8}$$

where $(p_l)_{l=1}^m$ are the marginal pmf-s, $(\phi_j)_{j=1}^m$ are bounded non-constant kernel functions, and the constants $\omega_{j_1...j_k} \in \mathbb{R}$ are such that the following conditions hold

$$\sum_{n\in\mathbb{N}} \phi_j(n)\, p_j(n) = 0,\ \forall j \in \overline{1,m}, \tag{7.2.9}$$

$$1 + \sum_{k=2}^{m} \sum_{1\le j_1<...<j_k\le m} \omega_{j_1...j_k} \phi_{j_1}(n_{j_1}) \times ... \times \phi_{j_k}(n_{j_k}) \ge 0,\ \forall \mathbf{n} \in \mathbb{N}^m. \tag{7.2.10}$$

The joint distribution of any subset of marginals of $\mathbf{N}$ is of the same type. The particular forms discussed in the literature for the functions ϕ_j can be unified into a general one satisfying condition (7.2.9), i.e.,

$$\phi_j(x) = f_j(x) - \mathbb{E}[f_j(N_j)], \tag{7.2.11}$$

with the functions f_j properly chosen such that $\mathbb{E}[f_j(N_j)] < \infty$. For simplicity, we denote $E_j := \mathbb{E}[f_j(N_j)]$. In our study, we shall consider the following particular cases:

1. $f(x) = e^{-\delta x} \Rightarrow \phi(x) = e^{-\delta x} - \mathcal{L}_N(\delta)$; a frequent choice is $\delta = 1$.

2. $f = p \Rightarrow \phi(x) = p(x) - \sum_{n\in\mathbb{N}} p^2(n)$.

Remark 7.2.1. *In the bivariate case where* $p(\mathbf{n}) = \prod_{i=1}^{2} p_i(n_i)\left(1 + \omega \prod_{i=1}^{2} \phi_i(n_i)\right)$ *,*$\mathbf{n} \in \mathbb{N}^2$*, condition (7.2.10) yields the following range of* ω

$$\max\left\{-\frac{1}{m_1 m_2}, -\frac{1}{M_1 M_2}\right\} \le \omega \le \min\left\{-\frac{1}{m_1 M_2}, -\frac{1}{m_2 M_1}\right\},$$

where $m_i = \min_{n\in\mathbb{N}} \phi_i(n)$, $M_i = \max_{n\in\mathbb{N}} \phi_i(n)$. *From these limits we can also obtain the correlation range, where the correlation coefficient is given by*

$$\rho = \omega \frac{\mathbb{E}[N_1\phi_1(N_1)]\,\mathbb{E}[N_2\phi_2(N_2)]}{\sqrt{Var[N_1]\,Var[N_2]}}. \tag{7.2.12}$$

7.3 Some univariate compound distributions with Erlang severity distribution

Let us consider the univariate compound model (7.1.1) with independent claim sizes $Erlang\,(k,\beta)$ distributed, $k\in\mathbb{N}^{*},\beta>0$. Then, according to (7.2.2), $(p\vee h)(0)=p(0)$ and, for $x>0$,

$$(p\vee h)(x)=\sum_{n=1}^{\infty}p(n)\,\frac{\beta^{nk}}{(nk-1)!}x^{nk-1}e^{-\beta x}. \tag{7.3.1}$$

Regarding the choice of p, we consider the following three cases for which we express the resulting pdf in terms of the generalized hypergeometric function.

Proposition 7.3.1. *i) Poisson case: let* $N\sim Po(\mu)$. *Then* $(p\vee h)(0)=e^{-\mu}$ *and for* $x>0$,

$$(p\vee h)(x)=e^{-\mu-\beta x}\frac{\mu(\beta x)^{k}}{x(k-1)!}\,{}_0F_k\left(\{\},\left\{\frac{k+1}{k},\frac{k+2}{k},...,\frac{2k-1}{k},2\right\};\mu\left(\frac{\beta x}{k}\right)^{k}\right).$$

ii) Negative Binomial case: assuming that $N\sim NB(r,q),r>0,q\in(0,1)$, *we have* $(p\vee h)(0)=q^{r}$, *while for* $x>0$,

$$(p\vee h)(x)=\frac{rq^{r}(1-q)(\beta x)^{k}e^{-\beta x}}{x(k-1)!}\,{}_1F_k\left(\{r+1\},\left\{\frac{k+1}{k},\frac{k+2}{k},...,\frac{2k-1}{k},2\right\};(1-q)\left(\frac{\beta x}{k}\right)^{k}\right).$$

iii) Logarithmic case: assume that $N\sim Log(\theta),\theta\in(0,1)$. *Then* $(p\vee h)(0)=0$, *while for* $x>0$,

$$(p\vee h)(x)=\begin{cases}-\frac{e^{-\beta x}}{x\ln(1-\theta)}\left(e^{\theta\beta x}-1\right),k=1\\ -\frac{ke^{-\beta x}}{x\ln(1-\theta)}\left[{}_0F_k\left(\{\},\left\{\frac{1}{k},\frac{2}{k},...,\frac{k-1}{k}\right\};\theta\left(\frac{\beta x}{k}\right)^{k}\right)-1\right],k\geq 2\end{cases}.$$

Remark 7.3.1. *The above proposition can be easily extended to the case where the counting distribution is a mixture with the corresponding components being distributions considered in the proposition. For example, if* N *follows a mixture of two Poisson distributions, i.e.,*

$$p(n)=qe^{-\mu_1}\frac{\mu_1^{n}}{n!}+(1-q)\,e^{-\mu_2}\frac{\mu_2^{n}}{n!},n\in\mathbb{N},q\in(0,1),\mu_i>0,i=1,2,$$

then we easily obtain that for $x > 0$,

$$\begin{aligned}(p \vee h)(x) &= \frac{e^{-\beta x}(\beta x)^k}{x(k-1)!}\left[q\mu_1 e^{-\mu_1}\, {}_0F_k\left(\{\}, \left\{\frac{k+1}{k}, \frac{k+2}{k}, ..., \frac{2k-1}{k}, 2\right\}; \mu_1\left(\frac{\beta x}{k}\right)^k\right)\right. \\ &\left. + (1-q)\,\mu_2 e^{-\mu_2}\, {}_0F_k\left(\{\}, \left\{\frac{k+1}{k}, \frac{k+2}{k}, ..., \frac{2k-1}{k}, 2\right\}; \mu_2\left(\frac{\beta x}{k}\right)^k\right)\right].\end{aligned}$$

Remark 7.3.2. *The distribution obtained in the first case (i) of Proposition 7.3.1 belongs to the class of Tweedie distributions, which in some cases can be modeled as compound distributions with Poisson counting distribution and Gamma severity distribution, see, e.g., Jorgensen [68].*

Sarabia et al. [106] presented closed-type formulas for the univariate collective model (7.1.1) under the assumption that the claim sizes are dependent according to the multivariate type II Pareto distribution described in Section 7.2.2. In the following lemma, we recall their results for the same counting distributions considered above.

Lemma 7.3.2. *If the claim sizes are multivariate type II Pareto distributed with parameters* (α, σ), *then:*
i) Poisson case: let $N \sim Po(\mu)$. *Then* $(p \vee h)(0) = e^{-\mu}$ *and for* $x > 0$,

$$(p \vee h)(x) = \frac{\alpha\mu e^{-\mu}}{\sigma(1+x/\sigma)^{1+\alpha}}\, {}_1F_1\left(\{1+\alpha\}, \{2\}; \frac{\mu x}{\sigma + x}\right).$$

ii) Negative Binomial case: assuming that $N \sim NB(r, q)$, $r > 0, q \in (0,1)$, *we have* $(p \vee h)(0) = q^r$, *while for* $x > 0$,

$$(p \vee h)(x) = \frac{r\alpha(1-q)q^r}{\sigma(1+x/\sigma)^{1+\alpha}}\, {}_2F_1\left(\{1+r, 1+\alpha\}, \{2\}; \frac{(1-q)x}{\sigma + x}\right).$$

iii) Logarithmic case: assume that $N \sim Log(\theta)$, $\theta \in (0,1)$. *Then* $(p \vee h)(0) = 0$, *while for* $x > 0$,

$$(p \vee h)(x) = -\frac{1}{x\ln(1-\theta)}\left[\frac{1}{(1+(1-\theta)x/\sigma)^{\alpha}} - \frac{1}{(1+x/\sigma)^{\alpha}}\right].$$

7.4 Multivariate compound distributions with Sarmanov's counting distribution

In this section, we extend the above results to the multivariate case corresponding to model (7.1.2). We assume that the vector number of claims follows Sarmanov's multivariate distribution. Regarding the claim sizes, we first assume the existence of both inside-type and between-types independencies, then we relax the inside-type independency condition similarly to Sarabia et al. [106]. In both situations, the following result holds.

Proposition 7.4.1. *Consider the multivariate compound distribution (7.2.4) under the assumption that the multivariate counting distribution p is of Sarmanov type (7.2.8). Then the resulting compound distribution also belongs to Sarmanov's class, satisfying for* $\mathbf{s} \geq \mathbf{0}$,

$$\begin{aligned} (p \vee \mathbf{h})(\mathbf{s}) &= \prod_{l=1}^{m} (p_l \vee h_l)(s_l) + \sum_{k=2}^{m} \sum_{1 \leq j_1 < ... < j_k \leq m} \omega_{j_1 ... j_k} \\ &\quad \times \prod_{l=1}^{k} ((f_{j_l} p_{j_l}) \vee h_{j_l} - E_{j_l}(p_{j_l} \vee h_{j_l}))(s_{j_l}) \prod_{l=k+1}^{m} (p_{j_l} \vee h_{j_l})(s_{j_l}) \\ &= \prod_{l=1}^{m} (p_l \vee h_l)(s_l) \left(1 + \sum_{k=2}^{m} \sum_{1 \leq j_1 < ... < j_k \leq m} \omega_{j_1 ... j_k} \tilde{\phi}_{j_1}(s_{j_1}) \times ... \times \tilde{\phi}_{j_k}(s_{j_k})\right), \end{aligned} \tag{7.4.1}$$

where the indexes $\{j_{k+1}, ..., j_m\} = \{1, ..., m\} \setminus \{j_1, ..., j_k\}$ *and* $\tilde{\phi}_j = \frac{(f_j p_j) \vee h_j}{p_j \vee h_j} - E_j, j = \overline{1, m}$.

Remark 7.4.1. *From the above definition of* $\tilde{\phi}_j$ *it results that* $\tilde{\phi}_j(0) = \frac{(f_j p_j)(0)}{p_j(0)} - E_j = f_j(0) - E_j = \phi_j(0)$, *from where formula (7.4.1) yields* $(p \vee \mathbf{h})(\mathbf{0}) = p(\mathbf{0})$ *as expected, and, e.g.,*

$$\begin{aligned} (p \vee \mathbf{h})(s, 0, ..., 0) &= (p_1 \vee h_1)(s) \left(\prod_{l=2}^{m} p_l(0)\right) \left[1 + \sum_{k=1}^{m-1} \sum_{2 \leq j_1 < ... < j_k \leq m} \omega_{1 j_1 ... j_k} \tilde{\phi}_1(s) \prod_{l=1}^{k} \phi_{j_l}(0) \right. \\ &\quad \left. + \sum_{k=2}^{m-1} \sum_{2 \leq j_1 < ... < j_k \leq m} \omega_{j_1 ... j_k} \prod_{l=1}^{k} \phi_{j_l}(0)\right], s > 0. \end{aligned}$$

As a consequence, the compound distribution $p \vee \mathbf{h}$ *has now continuous and discrete parts (with pmf in* $\mathbf{0}$*).*

In what concerns the correlation, the following proposition deals with the bivariate case.

Proposition 7.4.2. *The correlation coefficient of the bivariate compound distribution defined by (7.1.2) with pdf (7.4.1) is given by*

$$\rho(S_1, S_2) = \rho(N_1, N_2) \prod_{j=1}^{2} \mathbb{E}[X_j] \sqrt{\frac{Var[N_j]}{Var[S_j]}}, \tag{7.4.2}$$

where $\rho(N_1, N_2)$ *is the correlation coefficient of the bivariate counting distribution of type (7.2.8). If, moreover, the claim sizes* $(X_{jl})_{l\geq 1}$ *are*
i) i.i.d., then

$$\mathbb{E}[X_j]\sqrt{\frac{Var[N_j]}{Var[S_j]}} = \left(\frac{\mathbb{E}[N_j]}{Var[N_j]}\frac{Var[X_j]}{(\mathbb{E}[X_j])^2}+1\right)^{-1/2};$$

ii) dependent, i.d., with equal covariances $c_j := cov(X_{ji}, X_{jl})$ *for any* $i \neq l, j = 1, 2$, *then*

$$\mathbb{E}[X_j]\sqrt{\frac{Var[N_j]}{Var[S_j]}} = \left(\frac{\mathbb{E}[N_j]}{Var[N_j]}\frac{Var[X_j]}{(\mathbb{E}[X_j])^2}+\frac{c_j\mathbb{E}[N_j(N_j-1)]}{(\mathbb{E}[X_j])^2\,Var[N_j]}+1\right)^{-1/2}.$$

The following corollary states the fact that the correlation of the compound distribution cannot exceed the one of the counting distribution (its proof is obvious, hence we omit it). Note that when the claims of some type are assumed to be correlated, it is natural to assume that their correlation is positive.

Corollary 7.4.3. *Under the assumptions of Proposition 7.4.2, in both cases (i) and (ii) with* $c_j > 0, j = 1, 2$, *it holds that* $\rho(S_1, S_2) < \rho(N_1, N_2)$.

In consequence, we need the form of $\rho(N_1, N_2)$. The following result present some particular formulas useful to the evaluation of $\rho(N_1, N_2)$ in cases that will be considered later on.

Lemma 7.4.4. *i) Let* N *be a discrete r.v. and* $\phi(x) = e^{-\delta x} - \mathcal{L}_N(\delta)$. *It holds that:*
- if $N \sim Po(\mu)$ *then* $\mathcal{L}_N(\delta) = e^{\mu\left(e^{-\delta}-1\right)}$ *and* $\frac{\mathbb{E}[N\phi(N)]}{\sqrt{Var[N]}} = \left(e^{-\delta}-1\right)\sqrt{\mu}e^{\left(e^{-\delta}-1\right)\mu}$;
- if $N \sim NB(r, q)$ *then* $\mathcal{L}_N(\delta) = \left(\frac{q}{1-(1-q)e^{-\delta}}\right)^r$ *and* $\frac{\mathbb{E}[N\phi(N)]}{\sqrt{Var[N]}} = \frac{\left(e^{-\delta}-1\right)q^r\sqrt{r(1-q)}}{\left(1-(1-q)e^{-\delta}\right)^{r+1}}$;
- if $N \sim Log(\theta)$ *then* $\mathcal{L}_N(\delta) = \frac{\ln\left(1-e^{-\delta}\theta\right)}{\ln(1-\theta)}$ *and* $\frac{\mathbb{E}[N\phi(N)]}{\sqrt{Var[N]}} = \sqrt{\frac{-\theta}{\theta+\ln(1-\theta)}}\left(\frac{e^{-\delta}(1-\theta)}{1-\theta e^{-\delta}} - \frac{\ln\left(1-\theta e^{-\delta}\right)}{\ln(1-\theta)}\right)$.
ii) Assuming now that p *denotes the pmf of* N *and* $\phi(x) = p(x) - \sum_n p^2(n)$, *hence* $E = \sum_{n\in\mathbb{N}} p^2(n)$, *we have:*
- if $N \sim Po(\mu)$ *then* $E = e^{-2\mu}I_0(2\mu)$ *and* $\frac{\mathbb{E}[N\phi(N)]}{\sqrt{Var[N]}} = \sqrt{\mu}e^{-2\mu}\left(I_1(2\mu) - I_0(2\mu)\right)$;
- if $N \sim NB(r, q)$ *then* $E = q^{2r}\,{}_2F_1\left(\{r, r\}, \{1\}; (1-q)^2\right)$ *and* $\frac{\mathbb{E}[N\phi(N)]}{\sqrt{Var[N]}} = \sqrt{r(1-q)}q^{2r}$
$\times\left[rq(1-q)\,{}_2F_1\left(\{r+1, r+1\}, \{2\}; (1-q)^2\right) - {}_2F_1\left(\{r, r\}, \{1\}; (1-q)^2\right)\right]$;
- if $N \sim Log(\theta)$ *then* $E = \frac{Li_2\left(\theta^2\right)}{\ln^2(1-\theta)}$ *and* $\frac{\mathbb{E}[N\phi(N)]}{\sqrt{Var[N]}} = \frac{(1-\theta)\ln(1-\theta)\ln\left(1-\theta^2\right) - \theta Li_2\left(\theta^2\right)}{\ln^2(1-\theta)\sqrt{-\theta(\theta+\ln(1-\theta))}}$.

We shall now have a look at some particular choices for the counting and severity distributions.

7.4.1 Inside-type independency case

In this section, we assume that the claim sizes of type j, i.e., $(X_{jl})_{l\geq 1}$, are independent for any fixed j, and also independent of all the claim sizes of other types, i.e., of $(X_{kl})_{l\geq 1}, \forall k \neq j$. Under these assumptions, Vernic [119] presented recursive formulas for the evaluation of multivariate compound distributions when the claim sizes distributions are of discrete type. We shall now see how the components of the compound pdf look like in the particular continuous case of independent Erlang distributed claim sizes, i.e., when $X_{jl} \sim Erlang\,(k_j, \beta_j)\,, l \geq 1, j = \overline{1,m}$. We also assume that each marginal counting distribution, p_j, is of Poisson, negative binomial or logarithmic type. Then, from Proposition 7.3.1, we know the form of the $p_j \vee h_j$ pdf-s, hence, in view of (7.4.1), we must find the expressions of the $\tilde{\phi}_j$s. The following results examine the later for the two particular forms of the kernel functions presented in Section 7.2.2.

Proposition 7.4.5. *Let h be the $Erlang\,(k,\beta)$ pdf and let $\phi\,(x) = e^{-\delta x} - \mathcal{L}_N\,(\delta)$, hence $f\,(x) = e^{-\delta x}$. It holds that*
i) If $N \sim Po\,(\mu)$ then $((fp) \vee h)\,(0) = e^{-\mu}$, while for $x > 0$,

$$((fp) \vee h)\,(x) = e^{-\delta-\mu-\beta x}\frac{\mu\,(\beta x)^k}{x\,(k-1)!}\,{}_0F_k\left(\{\}\,,\left\{\frac{k+1}{k},\frac{k+2}{k},...,\frac{2k-1}{k},2\right\};\frac{\mu}{e^{\delta}}\left(\frac{\beta x}{k}\right)^k\right).$$

ii) If $N \sim NB\,(r,q)$ then $((fp) \vee h)\,(0) = q^r$, while for $x > 0$,

$$\begin{aligned}((fp) \vee h)\,(x) &= \frac{rq^r\,(1-q)\,e^{-\delta-\beta x}\,(\beta x)^k}{x\,(k-1)!} \\ &\quad\times\,{}_1F_k\left(\{r+1\}\,,\left\{\frac{k+1}{k},\frac{k+2}{k},...,\frac{2k-1}{k},2\right\};\frac{1-q}{e^{\delta}}\left(\frac{\beta x}{k}\right)^k\right).\end{aligned}$$

iii) If $N \sim Log\,(\theta)$ then $((fp) \vee h)\,(0) = 0$, while for $x > 0$,

$$((fp) \vee h)\,(x) = \begin{cases} -\frac{e^{-\beta x}}{x\ln(1-\theta)}\left(e^{\theta\beta x e^{-\delta}}-1\right), k=1 \\ -\frac{ke^{-\beta x}}{x\ln(1-\theta)}\left[{}_0F_k\left(\{\}\,,\left\{\frac{1}{k},\frac{2}{k},...,\frac{k-1}{k}\right\};\frac{\theta}{e^{\delta}}\left(\frac{\beta x}{k}\right)^k\right)-1\right], k\geq 2\end{cases}.$$

The proof of Proposition 7.4.5 is omitted, being very similar with the proof of Proposition 7.3.1.

Proposition 7.4.6. *Let h be the $Erlang\,(k,\beta)$ pdf and let $\phi\,(x) = p\,(x) - \sum_{n\in\mathbb{N}} p^2\,(n)$. Then* *i) When* $N \sim Po(\mu)$ *we have* $((fp) \vee h)(0) = e^{-2\mu}$, *while for* $x > 0$,

$$((fp) \vee h)(x) = e^{-2\mu-\beta x} \frac{\mu^2 (\beta x)^k}{x(k-1)!} {}_0F_{k+1}\left(\{\}, \left\{\frac{k+1}{k}, \frac{k+2}{k}, ..., \frac{2k-1}{k}, 2, 2\right\}; \mu^2 \left(\frac{\beta x}{k}\right)^k\right).$$

ii) When $N \sim NB(r, q)$ *we have* $((fp) \vee h)(0) = q^{2r}$, *while for* $x > 0$,

$$\begin{aligned} ((fp) \vee h)(x) &= \frac{r^2 q^{2r} (1-q)^2 e^{-\beta x} (\beta x)^k}{x(k-1)!} \\ &\quad \times {}_2F_{k+1}\left(\{r+1, r+1\}, \left\{\frac{k+1}{k}, \frac{k+2}{k}, ..., \frac{2k-1}{k}, 2, 2\right\}; (1-q)^2 \left(\frac{\beta x}{k}\right)^k\right). \end{aligned}$$

iii) When $N \sim Log(\theta)$ *we have* $((fp) \vee h)(0) = 0$, *while for* $x > 0$,

$$((fp) \vee h)(x) = \begin{cases} \frac{\theta^2 e^{-\beta x} \beta x}{x \ln^2(1-\theta)} \, {}_2F_2(\{1,1\}, \{2,2\}; \theta^2 \beta x), & k = 1 \\ \frac{\theta^2 e^{-\beta x} (\beta x)^k}{x(k-1)! \ln^2(1-\theta)} \, {}_2F_{k+1}\left(\{1,1\}, \{\frac{k+1}{k}, \frac{k+2}{k}, ..., \frac{2k-1}{k}, 2, 2\}; \theta^2 \left(\frac{\beta x}{k}\right)^k\right), & k \geq 2 \end{cases}.$$

Example 7.4.1. *In this example, for illustration purposes, we consider the bivariate case. Hence, for, e.g., two insurance portfolios, we define the compound model (7.1.2) as follows:* $(S_1, S_2) = \left(\sum_{l=0}^{N_1} X_{1l}, \sum_{l=0}^{N_2} X_{2l}\right)$ *with* $X_{1l} \sim Erlang(k_1 = 2, \beta_1 = 0.9)$, $X_{2l} \sim Erlang(k_2 = 3, \beta_2 = 0.95), l \geq 1$, $X_{10} = X_{20} = 0$, *while the r.v.s* $N_1 \sim Po(\mu = 2)$, $N_2 \sim NB(r = 4, q = 0.65)$ *are joined by the Sarmanov distribution with kernels of type* $\phi(n) = e^{-n} - \mathcal{L}_N(1)$ *and dependence parameter* $\omega = 3$. *Since in this case we have* $E_1 = \mathcal{L}_{Po(2)}(1) = 0.2825$, $E_2 = \mathcal{L}_{NB(4,0.65)}(1) = 0.3098$, *the pmf of* (N_1, N_2) *is*

$$\begin{aligned} p(n_1, n_2) &= p_1(n_1) p_2(n_2) \left[1 + \omega \left(e^{-n_1} - E_1\right) \left(e^{-n_2} - E_2\right)\right] \\ &= 0.00403 \frac{2^{n_1} 0.35^{n_2} \Gamma(4 + n_2)}{n_1! n_2!} \left[1 + 3\left(e^{-n_1} - 0.2825\right)\left(e^{-n_2} - 0.3098\right)\right], n_1, n_2 \in \mathbb{N}. \end{aligned}$$

Note that this Sarmanov distribution joins different types of marginals, i.e., one Poisson and one Negative Binomial. Also, its correlation coefficient results using Lemma 7.4.4 as

$$\rho(N_1, N_2) = \omega \left(e^{-1} - 1\right)^2 \frac{q^r \sqrt{\mu r (1-q)} e^{\left(e^{-1}-1\right)\mu}}{\left(1 - (1-q) e^{-1}\right)^{r+1}} \simeq 0.2015.$$

Note that the possible ranges of ω *and* ρ *are in this case* $\omega \in (-2.0192, 4.4983)$, $\rho(N_1, N_2) \in (-0.1356, 0.3021)$.

Thus, according to (7.4.1), the joint pdf of (S_1, S_2) *can be written as*

$$(p \vee \mathbf{h})(\mathbf{s}) = \prod_{l=1}^{2} (p_l \vee h_l)(s_l) \left(1 + \omega \tilde{\phi}_1(s_1) \tilde{\phi}_2(s_2)\right), \mathbf{s} \geq \mathbf{0},$$

where $\tilde{\phi}_j = \frac{(f_j p_j) \vee h_j}{p_j \vee h_j} - E_j$ *with* $f_j(x) = e^{-x}, j = 1, 2$. *From Proposition 7.3.1, we obtain the marginal pdf-s of* S_1 *and* S_2 *as, respectively,*

$$\begin{aligned} (p_1 \vee h_1)(s) &= \begin{cases} 0.1353, s = 0 \\ 0.2192 s e^{-0.9s} \ {}_0F_2\left(\{\}, \{\frac{3}{2}, 2\}; 0.405 s^2\right), s > 0 \end{cases}, \\ (p_2 \vee h_2)(s) &= \begin{cases} 0.1785, s = 0 \\ 0.1071 s^2 e^{-0.95s} \ {}_1F_3\left(\{5\}, \{\frac{4}{3}, \frac{5}{3}, 2\}; 0.0111 s^3\right), s > 0 \end{cases}. \end{aligned}$$

Also, from Proposition 7.4.5, the kernel functions $\tilde{\phi}_1$ *and* $\tilde{\phi}_2$ *are expressed, respectively, by*

$$\begin{aligned} \tilde{\phi}_1(0) &= 1 - E_1 = 0.7175, \ \tilde{\phi}_1(s) = \frac{{}_0F_2\left(\{\}, \{\frac{3}{2}, 2\}; 0.405 e^{-1} s^2\right)}{{}_0F_2\left(\{\}, \{\frac{3}{2}, 2\}; 0.405 s^2\right) e} - 0.2825, s > 0, \\ \tilde{\phi}_2(0) &= 1 - E_2 = 0.6902, \ \tilde{\phi}_2(s) = \frac{{}_1F_3\left(\{5\}, \{\frac{4}{3}, \frac{5}{3}, 2\}; 0.0111 e^{-1} s^3\right)}{{}_1F_3\left(\{5\}, \{\frac{4}{3}, \frac{5}{3}, 2\}; 0.0111 s^3\right) e} - 0.3098, s > 0. \end{aligned}$$

Then $(p \vee \mathbf{h})(\mathbf{0}) = p(\mathbf{0}) = 0.06005$, *while for* $s > 0$,

$$\begin{aligned} (p \vee \mathbf{h})(s, 0) &= (p_1 \vee h_1)(s) p_2(0) \left(1 + \omega \tilde{\phi}_1(s) \tilde{\phi}_2(0)\right) \\ &= 0.0391 s e^{-0.9s} \ {}_0F_2\left(\{\}, \left\{\frac{3}{2}, 2\right\}; 0.405 s^2\right) \\ &\times \left[1 + 2.0706 \left(\frac{{}_0F_2\left(\{\}, \{\frac{3}{2}, 2\}; 0.405 e^{-1} s^2\right)}{{}_0F_2\left(\{\}, \{\frac{3}{2}, 2\}; 0.405 s^2\right) e} - 0.2825\right)\right], \end{aligned} \tag{7.4.3}$$

$$\begin{aligned} (p \vee \mathbf{h})(0, s) &= p_1(0) (p_2 \vee h_2)(s) \left(1 + \omega \tilde{\phi}_1(0) \tilde{\phi}_2(s)\right) \\ &= 0.0145 s^2 e^{-0.95s} \ {}_1F_3\left(\{5\}, \left\{\frac{4}{3}, \frac{5}{3}, 2\right\}; 0.0111 s^3\right) \\ &\times \left[1 + 2.1526 \left(\frac{{}_1F_3\left(\{5\}, \{\frac{4}{3}, \frac{5}{3}, 2\}; 0.0111 e^{-1} s^3\right)}{{}_1F_3\left(\{5\}, \{\frac{4}{3}, \frac{5}{3}, 2\}; 0.0111 s^3\right) e} - 0.3098\right)\right]. \end{aligned} \tag{7.4.4}$$

Finally, for $s_1 > 0, s_2 > 0$,

$$(p \vee \mathbf{h})(\mathbf{s}) = 0.0235 s_1 s_2^2 e^{-0.9 s_1 - 0.95 s_2}$$

$$\times {}_0F_2\left(\{\},\left\{\frac{3}{2},2\right\};0.405s_1^2\right)\ {}_1F_3\left(\{5\},\left\{\frac{4}{3},\frac{5}{3},2\right\};0.0111s_2^3\right)$$
$$\times\left[1+3\left(\frac{{}_0F_2\left(\{\},\left\{\frac{3}{2},2\right\};0.405e^{-1}s_1^2\right)}{{}_0F_2\left(\{\},\left\{\frac{3}{2},2\right\};0.405s_1^2\right)e}-0.2825\right)\right.$$
$$\left.\times\left(\frac{{}_1F_3\left(\{5\},\left\{\frac{4}{3},\frac{5}{3},2\right\};0.0111e^{-1}s_2^3\right)}{{}_1F_3\left(\{5\},\left\{\frac{4}{3},\frac{5}{3},2\right\};0.0111s_2^3\right)e}-0.3098\right)\right]. \qquad (7.4.5)$$

In what concerns the correlation of this compound distribution, from Proposition 7.4.2 item (i) we obtain

$$\rho\left(S_1,S_2\right)=\rho\left(N_1,N_2\right)\left(\left(\frac{1}{k_1}+1\right)\left(\frac{q}{k_2}+1\right)\right)^{-1/2}\simeq 0.1492.$$

Moreover, for the actual values of the parameters, the possible range of this correlation coefficient when ω varies in its interval is $\rho\left(S_1,S_2\right)\in(-0.1004,0.2236)$.

7.4.2 Inside-type dependency case

We shall now relax the inside-type independency condition while keeping the between-types independency one, i.e., the different types of claims are independent of each other. Hence, we assume that the claim sizes of each type are dependent and follows the multivariate Pareto distribution as defined in Section 7.2.2. As before, we rely on formula (7.4.1) to find the compound distribution $p\vee\mathbf{h}$; therefore, we must evaluate all the marginal compound distributions $p_j\vee h_j$, as well as the quantities $(f_jp_j)\vee h_j$ involved in the $\tilde{\phi}_j$s. Considering the previous three particular counting distributions, we already know the form of each $p_j\vee h_j$ from Lemma 7.3.2, while the expressions of $(f_jp_j)\vee h_j$ are given in the following two properties for two particular kernels cases.

Proposition 7.4.7. *Let the claim sizes be multivariate type II Pareto distributed with parameters (α,σ), and let $\phi\left(x\right)=e^{-\delta x}-\mathcal{L}_N\left(\delta\right)$, hence $f\left(x\right)=e^{-\delta x}$. Then*
i) If $N\sim Po\left(\mu\right)$ then $\left(\left(fp\right)\vee h\right)\left(0\right)=e^{-\mu}$, while for $x>0$,

$$\left(\left(fp\right)\vee h\right)\left(x\right)=\frac{\alpha\mu e^{-(\mu+\delta)}}{\sigma\left(1+x/\sigma\right)^{1+\alpha}}\ {}_1F_1\left(\{1+\alpha\},\{2\};\frac{\mu x e^{-\delta}}{\sigma+x}\right).$$

ii) If $N\sim NB\left(r,q\right)$ then $\left(\left(fp\right)\vee h\right)\left(0\right)=q^r$, while for $x>0$,

$$\left(\left(fp\right)\vee h\right)\left(x\right)=\frac{\alpha r q^r(1-q)e^{-\delta}}{\sigma(1+x/\sigma)^{1+\alpha}}\ {}_2F_1\left(\{1+\alpha,1+r\},\{2\};\frac{(1-q)xe^{-\delta}}{\sigma+x}\right).$$

iii) If $N \sim Log\,(\theta)$ *then* $((fp) \vee h)\,(0) = 0$, *while for* $x > 0$,

$$((fp) \vee h)\,(x) = -\frac{1}{x \ln(1-\theta)} \left[\frac{1}{(1 + (1-\theta e^{-\delta})x/\sigma)^{\alpha}} - \frac{1}{(1 + x/\sigma)^{\alpha}} \right].$$

The proof of Proposition 7.4.7 is similar to the proof of Lemma 7.3.2 given in Sarabia et al. [106], and therefore is omitted.

Proposition 7.4.8. *Let the claim sizes be multivariate type II Pareto distributed with parameters* (α, σ), *and let* $\phi(x) = p(x) - \sum_{n \in \mathbb{N}} p^2(n)$, *hence* $f = p$ *and* $E = \sum_{n \in \mathbb{N}} p^2(n)$. *Then*
i) When $N \sim Po\,(\mu)$ *we have* $((fp) \vee h)\,(0) = e^{-2\mu}$, *while for* $x > 0$,

$$((fp) \vee h)\,(x) = \frac{\alpha e^{-2\mu} \mu^2}{\sigma(1 + x/\sigma)^{\alpha+1}}\, {}_1F_2 \left(\{1+\alpha\}, \{2, 2\}; \frac{\mu^2 x}{\sigma + x} \right).$$

ii) When $N \sim NB\,(r, q)$ *we have* $((fp) \vee h)\,(0) = q^{2r}$, *while for* $x > 0$,

$$((fp) \vee h)\,(x) = \frac{\alpha r^2 q^{2r} (1-q)^2}{\sigma(1 + x/\sigma)^{\alpha+1}}\, {}_3F_2 \left(\{1+\alpha, 1+r, 1+r\}, \{2, 2\}; \frac{(1-q)^2 x}{\sigma + x} \right).$$

iii) When $N \sim Log\,(\theta)$ *we have* $((fp) \vee h)\,(0) = 0$, *while for* $x > 0$,

$$((fp) \vee h)\,(x) = 1 \frac{\alpha \theta^2}{\sigma(1 + x/\sigma)^{\alpha+1} \ln^2(1-\theta)}\, {}_3F_2 \left(\{1, 1, 1+\alpha\}, \{2, 2\}; \frac{\theta^2 x}{\sigma + x} \right).$$

Remarks 7.4.9. *Similarly to Example 7.4.1, we consider the compound model (7.1.2) in the bivariate case, i.e.,* $(S_1, S_2) = \left(\sum_{l=0}^{N_1} X_{1l}, \sum_{l=0}^{N_2} X_{2l} \right)$ *with* $N_1 \sim Po(\mu = 2)$, $N_2 \sim Log(\theta = 0.6)$, *while, for* $N_i = n_i, i = 1, 2$, *we let* $(X_{11}, \ldots, X_{1n_1}) \sim PaII_{n_1}(\alpha_1 = 4, \sigma_1 = 3)$ *and* $(X_{21}, \ldots, X_{2n_2}) \sim PaII_{n_2}(\alpha_2 = 3, \sigma_2 = 4)$. *We assume that the r.v.s* N_1 *and* N_2 *are dependent and joined by the Sarmanov distribution with kernels of type* $\phi(n) = e^{-n} - \mathcal{L}_N(1)$ *and dependence parameter* $\omega = 4.5$ *(in this case, the limiting interval is* $\omega \in (-1.9148, 4.8644)$*). Hence, with* $E_1 = \mathcal{L}_{Po(2)}(1) = 0.2825$ *and* $E_2 = \mathcal{L}_{Log(0.6)}(1) = 0.2722$, *the pmf of* (N_1, N_2) *is given by*

$$p(n_1, n_2) = 0.1477 \frac{2^{n_1} 0.6^{n_2}}{n_1! n_2} \left[1 + 4.5 \left(e^{-n_1} - 0.2825 \right) \left(e^{-n_2} 0.2722 \right) \right], n_1, n_2 \in \mathbb{N}, n_2 > 0.$$

Based on Lemma 7.4.4, its correlation coefficient is

$$\rho\,(N_1, N_2) = \omega \left(e^{-1} - 1 \right) e^{\left(e^{-1}-1\right)\mu} \sqrt{\frac{-\mu\theta}{\theta + \ln(1-\theta)}} \left(\frac{1-\theta}{e-\theta} - \frac{\ln(1-\theta e^{-1})}{\ln(1-\theta)} \right) \simeq 0.1304.$$

In fact, depending on ω, *the possible range of this correlation is* $\rho\,(N_1, N_2) \in (-0.0555, 0.1410)$.

The joint pdf of (S_1, S_2) *is given by*

$$(p \vee \mathbf{h})(\mathbf{s}) = \prod_{l=1}^{2} (p_l \vee h_l)(s_l) \left(1 + \omega \tilde{\phi}_1 (s_1) \tilde{\phi}_2 (s_2)\right), \mathbf{s} \geq \mathbf{0},$$

with $\tilde{\phi}_j = \frac{(f_j p_j) \vee h_j}{p_j \vee h_j} - E_j$ *and* $f_j(x) = e^{-x}, j = 1, 2$. *From Lemma 7.3.2, we obtain the marginal pdf-s of* S_1 *and* S_2*, respectively, as*

$$\begin{aligned}(p_1 \vee h_1)(s) &= \begin{cases} 0.1353, s = 0 \\ 0.3609(1 + s/3)^{-5}\ {}_1F_1(\{5\}, \{2\}; 2s/(s+3)), s > 0 \end{cases}, \\ (p_2 \vee h_2)(s) &= \begin{cases} 0, s = 0 \\ 1.0913 s^{-1} \left[(1 + 0.1s)^{-3} - (1 + 0.25s)^{-3}\right], s > 0 \end{cases}.\end{aligned}$$

Also, from Proposition 7.4.7, the kernel functions $\tilde{\phi}_1$ *and* $\tilde{\phi}_2$ *are expressed, respectively, by*

$$\begin{aligned}\tilde{\phi}_1(0) &= 1 - E_1 = 0.7175,\ \tilde{\phi}_1(s) = \frac{{}_1F_1\left(\{5\}, \{2\}; \frac{2e^{-1}s}{3+s}\right)}{{}_1F_1\left(\{5\}, \{2\}; \frac{2s}{3+s}\right) e} - 0.2825, s > 0, \\ \tilde{\phi}_2(0) &= 1 - E_2 = 0.7278,\ \tilde{\phi}_2(s) = \frac{(1 + 0.1948s)^{-3} - (1 + 0.25s)^{-3}}{(1 + 0.1s)^{-3} - (1 + 0.25s)^{-3}} - 0.2722, s > 0.\end{aligned}$$

Then $(p \vee \mathbf{h})(\mathbf{0}) = p(\mathbf{0}) = 0$*, while for* $s > 0$,

$$(p \vee \mathbf{h})(s, 0) = (p_1 \vee h_1)(s)\, p_2(0) \left(1 + \omega \tilde{\phi}_1(s) \tilde{\phi}_2(0)\right) = 0 \text{ as } (p_2 \vee h_2)(0) = 0,$$

$$\begin{aligned}(p \vee \mathbf{h})(0, s) &= p_1(0)\,(p_2 \vee h_2)(s) \left(1 + \omega \tilde{\phi}_1(0) \tilde{\phi}_2(s)\right) \\ &= 0.1477 s^{-1} \left[(1 + 0.1s)^{-3} - (1 + 0.25s)^{-3}\right] \\ &\quad \times \left[1 + 3.2289 \left(\frac{(1 + 0.1948s)^{-3} - (1 + 0.25s)^{-3}}{(1 + 0.1s)^{-3} - (1 + 0.25s)^{-3}} - 0.2722\right)\right].\end{aligned}$$

Finally, for $s_1 > 0, s_2 > 0$,

$$\begin{aligned}(p \vee \mathbf{h})(\mathbf{s}) &= \frac{0.3939}{(1 + s_1/3)^5 s_2}\ {}_1F_1\left(\{5\}, \{2\}; \frac{2s_1}{s_1 + 3}\right) \left[\frac{1}{(1 + 0.1s_2)^3} - \frac{1}{(1 + 0.25s_2)^3}\right] \\ &\quad \times \left[1 + 4.5 \left(\frac{{}_1F_1\left(\{5\}, \{2\}; \frac{2e^{-1}s_1}{3+s_1}\right)}{{}_1F_1\left(\{5\}, \{2\}; \frac{2s_1}{3+s_1}\right) e} - 0.2825\right)\right. \\ &\quad \times \left.\left(\frac{(1 + 0.1948s_2)^{-3} - (1 + 0.25s_2)^{-3}}{(1 + 0.1s_2)^{-3} - (1 + 0.25s_2)^{-3}} - 0.2722\right)\right].\end{aligned}$$

From Proposition 7.4.2 item (ii), using that $\mathbb{E}[N(N-1)] = Var[N] + \mathbb{E}^2[N] - \mathbb{E}[N]$, *we also find the correlation of the compound distribution as*

$$\rho(S_1, S_2) = \rho(N_1, N_2)\left(\left(\frac{\alpha_1+\mu}{\alpha_1-2}+1\right)\left(\frac{\ln(1-\theta)(\theta+(1-\theta)\alpha_2)}{(\alpha_2-2)(\theta+\ln(1-\theta))}+1\right)\right)^{-\frac{1}{2}} \simeq 0.0262$$

Considering the actual values of the parameters, when ω *varies in its limiting interval, the possible correlation range is* $\rho(S_1, S_2) \in (-0.0111, 0.0283)$. *Moreover, we numerically maximized (in Mathematica software) the corresponding formula of* $\rho(N_1, N_2)$ *with respect to the parameters* $\mu, \theta, \delta_1, \delta_2$ *and obtained that the maximum possible of the correlation is* $\rho_{max}(N_1, N_2) = 0.4702$ *for* $\mu = 0.2297, \theta = 0.9929, \delta_1 = 1.8367, \delta_2 = 0.0099$, *while* ω *resulted as 6.9073 (note that this time we let the* δ*s vary); hence, when* $\alpha_1 = 4$ *and* $\alpha_2 = 3, \rho_{max}(S_1, S_2) = 0.1753$.

7.5 Numerical example

As mentioned in the Introduction, when dealing with claim size distributions of continuous type, the usual approach consists in discretizing these distributions, and then in evaluating the corresponding discrete compound distribution by applying a specific technique such as, e.g., the recursive method, the Fast Fourier Transform (FFT) algorithm or simulation. Such an approach generates errors starting with the discretization step (by span choice), errors that are usually magnified by applying the specific technique (for some details see, e.g.,Robe-Voinea and Vernic [103] and the references therein).

Therefore, in this example, we compare the cdf-s obtained for the bivariate compound distribution presented in Example 7.4.1 by using the just described approach based on recursions, and by direct calculation.

In what concerns direct calculation, because the formulas involve the hypergeometric function, we used the facilities provided by the software R and Mathematica to numerically integrate the pdf-s (7.4.3)-(7.4.5). We note that the results were obtained immediately, taking less than a second for each integral.

Regarding the discretization approach, due to the nature of this example, we were able to apply to the resulting discrete compound distribution the recursive method presented in Vernic [119]. The main problem here was the choice of a proper discretization span h; we proceeded by trials (i.e., by successively reducing its value) and, finally, we stopped at $h = 0.001$ (we took the same span for both Erlang marginal distributions).

For comparison, we evaluated the cdf $F(s_1, s_2)$ of (S_1, S_2) by both methods, for $s_1, s_2 \in \{0, 1, 2, ..., 20\}$. Some values are presented in Table 1, while in Table 2 we display the maximum absolute error between the exact cdf values and the recursive ones; note how important are the differences between different spans.
On the other hand, the smaller the span is, the longer is the running time and the needed memory space (we wrote both R and Matlab programs). For example, it took more than one hour to evaluate the entire discretized cdf for $0 \leq s_1, s_2 \leq 20$ when $h = 0.001$, and we had to optimize the code in order to avoid "out of memory" messages. As another example, to find only the value $F(20, 20)$, the Matlab recursion-discretization code with the span $h = 0.001$ took about 25.14 seconds , while the exact integral value was obtained in about 1 second.

Therefore, we can see from this example that even if it involves some numerical integrals, direct calculation is more efficient than the classical recursion-discretization method in what concerns the accuracy of the values and the computing time.

Table 1. Comparison of exact and recursive CDF values (different spans)

CDF	(s_1, s_2)				
	$(0,0)$	$(0,5)$	$(0,10)$	$(0,15)$	$(0,20)$
Exact	0.006005	0.099282	0.121663	0.13011	0.133381
Rec. $h = 0.001$	0.006005	0.099286	0.121665	0.13011	0.133382
Rec. $h = 0.01$	0.006005	0.099317	0.121676	0.13011	0.133383
Rec. $h = 0.1$	0.006005	0.099634	0.121792	0.13016	0.133399
	$(5,0)$	$(10,0)$	$(15,0)$	$(20,0)$	$(5,5)$
Exact	0.141177	0.170763	0.177207	0.178323	0.326836
Rec. $h = 0.001$	0.141183	0.170764	0.177207	0.178323	0.326876
Rec. $h = 0.01$	0.141229	0.170774	0.177207	0.178321	0.327223
Rec. $h = 0.1$	0.141554	0.170684	0.177013	0.178106	0.330215
	$(10,10)$	$(10,15)$	$(15,10)$	$(15,15)$	$(20,20)$
Exact	0.683211	0.812865	0.735079	0.877797	0.955568
Rec. $h = 0.001$	0.683239	0.812886	0.735102	0.877809	0.955573
Rec. $h = 0.01$	0.683479	0.813060	0.735285	0.877902	0.955595
Rec. $h = 0.1$	0.684718	0.813377	0.735876	0.877289	0.954133

Table 2. Maximum absolute error between the exact and recursive CDF values for different spans

	$h = 1$	$h = 0.1$	$h = 0.01$	$h = 0.001$
$\max_{s_1, s_2 \in \{0,1,\ldots,20\}} \lvert F - F_{disc.} \rvert (s_1, s_2)$	0.143	0.005	0.00056	0.000045

7.6 Conclusions and future work

To conclude, in this paper we obtained some closed-type formulas for the multivariate pdf of some compound distributions with Sarmanov counting distribution and Erlang severity distributions; we also included some dependency between the claim sizes of a certain type by means of a multivariate Pareto distribution. Based on the hypergeometric function which is already implemented in existing software, these formulas seems to be numerically more efficient than the classical recursion-discretization approach, avoiding thus the typical discretization errors generated by the span choice, and the long running time characteristic to the multivariate case. Therefore, we think that it would be interesting to continue the search for such formulas in the case of compound distributions with continuous severity distributions, formulas expressed by means of special functions already implemented in mathematical software. Moreover, we also plan to pay special attention to the statistical inference of this type of compound distributions.

7.7 Appendix

Proof of Lemma 7.2.1. We write

$$\begin{aligned}((n+1)k-1)!(n+1) &= k^{nk}[1\times 2\times .... \times (k-1)] \\ &\times k^{-n}[k(2k)(3k)\times .... \times (nk)](n+1) \\ &\times k^{-n}[(k+1)(2k+1)\times .... \times (nk+1)] \\ &\times k^{-n}[(k+2)(2k+2)\times .... \times (nk+2)]\times .... \times \\ &\times k^{-n}[(k+(k-1))(2k+(k-1))\times .... \times (nk+(k-1))]\end{aligned}$$

$$\begin{aligned}&= (k-1)!k^{nk}(2\times(2+1)\times .... \times(2+n-1)) \\ &\times\left[\frac{k+1}{k}\left(\frac{k+1}{k}+1\right)\times .... \times\left(\frac{k+1}{k}+n-1\right)\right] \\ &\times ... \times\left[\frac{k+(k-1)}{k}\left(\frac{k+(k-1)}{k}+1\right)\times .... \times\left(\frac{k+(k-1)}{k}+n-1\right)\right],\end{aligned}$$

from where formula (7.2.6) is immediate. To obtain formula (7.2.7), we similarly write

$$\begin{aligned}(nk)! &= [(k+1)(2k+1)\times .... \times((n-1)k+1)] \\ &\times[2(k+2)(2k+2)\times .... \times((n-1)k+2)]\times .... \times \\ &\times[(k-1)(k+(k-1))(2k+(k-1))\times .... \times((n-1)k+(k-1))] \\ &\times[k(2k)(3k)\times .... \times(nk)] \\ &= k^{nk}\left[\frac{1}{k}\left(\frac{1}{k}+1\right)\times .... \times\left(\frac{1}{k}+n-1\right)\right] \\ &\times\left[\frac{2}{k}\left(\frac{2}{k}+1\right)\times .... \times\left(\frac{2}{k}+n-1\right)\right]\times .... \times \\ &\times\left[\frac{k-1}{k}\left(\frac{k-1}{k}+1\right)\times .... \times\left(\frac{k-1}{k}+n-1\right)\right]n!,\end{aligned}$$

hence the result. □

Proof of Proposition 7.3.1. *i)* In the Poisson case, based on (7.3.1) and (7.2.6) we have for $x>0$,

$$(p\vee h)(x) = \sum_{n=1}^{\infty} e^{-\mu}\frac{\mu^n}{n!}\frac{\beta^{nk}}{(nk-1)!}x^{nk-1}e^{-\beta x} = e^{-\mu-\beta x}\sum_{n=1}^{\infty}\frac{\left(\mu(\beta x)^k\right)^n}{n!(nk-1)!}x^{-1}$$

$$= e^{-\mu-\beta x}\frac{\mu(\beta x)^k}{x}\sum_{n=0}^{\infty}\frac{\left(\mu(\beta x)^k\right)^n}{(n+1)!\,((n+1)k-1)!}$$

$$= e^{-\mu-\beta x}\frac{\mu(\beta x)^k}{x(k-1)!}\sum_{n=0}^{\infty}\frac{\left(\mu\left(\frac{\beta x}{k}\right)^k\right)^n}{n!}\frac{1}{\left(\frac{k+1}{k}\right)_{(n)}\left(\frac{k+2}{k}\right)_{(n)}\times\ldots\times\left(\frac{k+(k-1)}{k}\right)_{(n)}(2)_{(n)}},$$

and using the definition (7.2.5), the formula is immediate. In the negative binomial case (ii), (7.3.1) and (7.2.6) yields

$$\begin{aligned}
(p\vee h)(x) &= \frac{q^r e^{-\beta x}}{\Gamma(r)}\sum_{n=1}^{\infty}\frac{\Gamma(r+n)}{n!}(1-q)^n\frac{\beta^{nk}}{(nk-1)!}x^{nk-1}\\
&= \frac{q^r e^{-\beta x}}{x\Gamma(r)}\sum_{n=1}^{\infty}\frac{\left((1-q)(\beta x)^k\right)^n}{n!}\frac{\Gamma(r+n)}{(nk-1)!}\\
&= \frac{q^r e^{-\beta x}}{x}\sum_{n=0}^{\infty}\frac{\left((1-q)(\beta x)^k\right)^{n+1}}{n!}\frac{\Gamma(r+n+1)}{(n+1)((n+1)k-1)!\Gamma(r)}\\
&= \frac{q^r(1-q)(\beta x)^k e^{-\beta x}}{x}\sum_{n=0}^{\infty}\frac{\left((1-q)(\beta x)^k\right)^n}{n!\,(k-1)!k^{nk}}\\
&\quad\times\frac{r(r+1)\times\ldots\times(r+1+n-1)}{\left(\frac{k+1}{k}\right)_{(n)}\left(\frac{k+2}{k}\right)_{(n)}\times\ldots\times\left(\frac{k+(k-1)}{k}\right)_{(n)}(2)_{(n)}},
\end{aligned}$$

hence the result. Finally, for the logarithmic distribution, (7.3.1) gives

$$\begin{aligned}
(p\vee h)(x) &= -\frac{e^{-\beta x}}{\ln(1-\theta)}\sum_{n=1}^{\infty}\frac{\theta^n}{n}\frac{\beta^{nk}x^{nk-1}}{(nk-1)!} = -\frac{e^{-\beta x}k}{x\ln(1-\theta)}\sum_{n=1}^{\infty}\frac{\left(\theta(\beta x)^k\right)^n}{(nk)!}\\
&= -\frac{ke^{-\beta x}}{x\ln(1-\theta)}\sum_{n=0}^{\infty}\left[\frac{\left(\theta(\beta x)^k\right)^n}{(nk)!}-1\right].
\end{aligned}$$

When $k=1$ (i.e., exponentially distributed claims) we obtain

$$(p\vee h)(x) = -\frac{e^{-\beta x}}{x\ln(1-\theta)}\left(e^{\theta\beta x}-1\right),$$

otherwise, using (7.2.7), we have

$$(p\vee h)(x) = -\frac{ke^{-\beta x}}{x\ln(1-\theta)}\sum_{n=0}^{\infty}\left[\frac{\left(\theta(\beta x)^k\right)^n}{n!k^{nk}}\frac{1}{\left(\frac{1}{k}\right)_{(n)}\left(\frac{2}{k}\right)_{(n)}\cdots\left(\frac{k-1}{k}\right)_{(n)}}-1\right],$$

i.e., the second formula of (iii), which completes the proof. $\square$

Proof of Proposition 7.4.1. This result was proved by Vernic [119] in the case when both inside-type and between-types independency assumptions hold and the claim sizes are of discrete type. Considering now the case with only between-types independence and no inside-type independence, along with continuous claim sizes, the proof is similar: inserting formula (7.2.8) and $\phi_l(x) = f_l(x) - E_l$ into (7.2.4), we have

$$\begin{aligned}
(p \vee \mathbf{h})(\mathbf{s}) &= \sum_{n_1=0}^{\infty} \dots \sum_{n_m=0}^{\infty} \left(\prod_{l=1}^{m} p_l(n_l)\, h_l^{(n_l)}(s_l) \right) \\
&\quad \times \left(1 + \sum_{k=2}^{m} \sum_{1 \le j_1 < \dots < j_k \le m} \omega_{j_1 \dots j_k}\, \phi_{j_1}(n_{j_1}) \times \dots \times \phi_{j_k}(n_{j_k}) \right) \\
&= \prod_{l=1}^{m} \left(\sum_{n_l=0}^{\infty} p_l(n_l)\, h_l^{(n_l)}(s_l) \right) + \sum_{k=2}^{m} \sum_{1 \le j_1 < \dots < j_k \le m} \omega_{j_1 \dots j_k} \\
&\quad \times \prod_{l=1}^{k} \left(\sum_{n_{j_l}=0}^{\infty} \left(f_{j_l}(n_{j_l}) - E_{j_l} \right) p_{j_l}(n_{j_l})\, h_{j_l}^{(n_{j_l})}(s_{j_l}) \right) \prod_{l=k+1}^{m} \left(\sum_{n_{j_l}=0}^{\infty} p_{j_l}(n_{j_l})\, h_{j_l}^{(n_{j_l})}(s_{j_l}) \right),
\end{aligned}$$

from where, using (7.2.1), we obtain formula (7.4.2). This formula easily yields the Sarmanov form (7.4.1); to verify that $\tilde{\phi}_j$ is indeed in the form (7.2.11), we proceed as follows: we denote by S_j the marginal r.v. having the compound distribution $p_j \vee h_j$, hence

$$\begin{aligned}
\mathbb{E}\left[\left(\frac{(f_j p_j) \vee h_j}{p_j \vee h_j} \right)(S_j) \right] &= \int_0^{\infty} \left(\frac{(f_j p_j) \vee h_j}{p_j \vee h_j} \right)(s)\, (p_j \vee h_j)(s)\, ds = \int_0^{\infty} ((f_j p_j) \vee h_j)(s)\, ds \\
&= \int_0^{\infty} \sum_{k=0}^{\infty} (f_j p_j)(k)\, h_j^{(k)}(s)\, ds = \sum_{k=0}^{\infty} (f_j p_j)(k) \int_0^{\infty} h_j^{(k)}(s)\, ds \\
&= \sum_{k=0}^{\infty} (f_j p_j)(k) = \mathbb{E}\left[f_j(N_j) \right] = E_j,
\end{aligned}$$

which completes the proof. $\square$

Proof of Proposition 7.4.2. To find $\rho(S_1, S_2)$, according to formula (7.2.12) we must evaluate

$$\begin{aligned}
\mathbb{E}\left[S_j \tilde{\phi}_j(S_j) \right] &= \int_0^{\infty} x \left(\frac{(f_j p_j) \vee h_j}{p_j \vee h_j} \right)(x)\, (p_j \vee h_j)(x)\, dx - E_j \mathbb{E}\left[S_j \right] \\
&= \sum_{n=0}^{\infty} (f_j p_j)(n) \int_0^{\infty} x h_j^{(n)}(x)\, dx - E_j \mathbb{E}\left[S_j \right].
\end{aligned}$$

In both cases with independent and dependent claim sizes, it holds that $\mathbb{E}[S_j] =$

$\mathbb{E}[N_j]\,\mathbb{E}[X_j]$ (for the dependent case see Sarabia et al. [106]); also, since $h_j^{(n)}$ is the pdf of $X_{j1}+...+X_{jn}$, we have $\int_0^\infty x h_j^{(n)}(x)\,dx = n\mathbb{E}[X_j]$, hence

$$\begin{aligned}\mathbb{E}\left[S_j\tilde{\phi}_j(S_j)\right] &= \mathbb{E}[X_j]\sum_{n=0}^{\infty} n(f_j p_j)(n) - E_j\mathbb{E}[N_j]\,\mathbb{E}[X_j] \\ &= \mathbb{E}[X_j]\left(\mathbb{E}[N_j f_j(N_j)] - E_j\mathbb{E}[N_j]\right) = \mathbb{E}[X_j]\,\mathbb{E}[N_j\phi_j(N_j)].\end{aligned}$$

Therefore,

$$\rho(S_1,S_2) = \omega\prod_{j=1}^{2}\frac{\mathbb{E}\left[S_j\tilde{\phi}_j(S_j)\right]}{\sqrt{Var[S_j]}} = \rho(N_1,N_2)\prod_{j=1}^{2}\mathbb{E}[X_j]\sqrt{\frac{Var[N_j]}{Var[S_j]}}.$$

$i)$ Under the independence assumption, we have $Var[S_j] = \mathbb{E}[N_j]\,Var[X_j] + \mathbb{E}^2[X_j]\,Var[N_j]$, which inserted into (7.4.2) easily yields the corresponding formula.
$ii)$ Under this dependence assumption, Sarabia et al. [106] proved that

$$Var[S_j] = \mathbb{E}[N_j]\,Var[X_j] + \mathbb{E}^2[X_j]\,Var[N_j] + c_j\mathbb{E}[N_j(N_j-1)],$$

and inserting it into (7.4.2) we obtain the last formula. This completes the proof.□

Proof of Lemma 7.4.4. $i)$ When $\phi(x) = e^{-\delta x} - \mathcal{L}_N(\delta)$, we have

$$\mathbb{E}[N\phi(N)] = \mathbb{E}\left[Ne^{-\delta N}\right] - \mathcal{L}_N(\delta)\,\mathbb{E}[N].$$

For $N \sim Po(\mu)$, $\mathcal{L}_N(\delta) = e^{\mu\left(e^{-\delta}-1\right)}$, while

$$\begin{aligned}\mathbb{E}[N\phi(N)] &= \sum_{n=1}^{\infty} ne^{-\delta n}e^{-\mu}\frac{\mu^n}{n!} - \mu e^{\mu\left(e^{-\delta}-1\right)} = \mu e^{-\mu-\delta}\sum_{n=0}^{\infty}\frac{\left(\mu e^{-\delta}\right)^n}{n!} - \mu e^{\mu\left(e^{-\delta}-1\right)} \\ &= \mu\left(e^{\mu\left(e^{-\delta}-1\right)-\delta} - e^{\mu\left(e^{-\delta}-1\right)}\right),\end{aligned}$$

from where we immediately obtain the stated formula. For $N \sim NB(r,q)$, $\mathcal{L}_N(\delta) = \frac{q^r}{\left(1-(1-q)e^{-\delta}\right)^r}$ and

$$\begin{aligned}\mathbb{E}\left[Ne^{-\delta N}\right] &= \sum_{n=0}^{\infty} ne^{-\delta n}\frac{\Gamma(r+n)}{n!\Gamma(r)}q^r(1-q)^n = q^r\sum_{n=0}^{\infty} n\frac{\Gamma(r+n)}{n!\Gamma(r)}\left((1-q)e^{-\delta}\right)^n \\ &= \frac{q^r}{\left(1-(1-q)e^{-\delta}\right)^r}\frac{r(1-q)e^{-\delta}}{1-(1-q)e^{-\delta}} = \frac{r(1-q)q^r e^{-\delta}}{\left(1-(1-q)e^{-\delta}\right)^{r+1}},\end{aligned}$$

hence

$$\begin{aligned}\mathbb{E}\left[N\phi\left(N\right)\right] &= \frac{r\left(1-q\right)q^{r}e^{-\delta}}{\left(1-\left(1-q\right)e^{-\delta}\right)^{r+1}} - \frac{q^{r-1}r\left(1-q\right)}{\left(1-\left(1-q\right)e^{-\delta}\right)^{r}} \\ &= \frac{r\left(1-q\right)q^{r-1}}{\left(1-\left(1-q\right)e^{-\delta}\right)^{r}}\left(\frac{qe^{-\delta}}{1-\left(1-q\right)e^{-\delta}}-1\right) = \frac{\left(e^{-\delta}-1\right)r\left(1-q\right)q^{r-1}}{\left(1-\left(1-q\right)e^{-\delta}\right)^{r+1}},\end{aligned}$$

which easily yields the stated formula. When $N \sim Log\left(\theta\right), \mathcal{L}_N\left(\delta\right) = \frac{\ln\left(1-\theta e^{-\delta}\right)}{\ln(1-\theta)}$ and using

$$\begin{aligned}\mathbb{E}\left[Ne^{-\delta N}\right] &= -\frac{1}{\ln(1-\theta)}\sum_{n=1}^{\infty} ne^{-\delta n}\frac{\theta^{n}}{n} = -\frac{\theta e^{-\delta}}{\ln(1-\theta)}\sum_{n=0}^{\infty}\left(e^{-\delta}\theta\right)^{n} \\ &= -\frac{\theta e^{-\delta}}{\ln(1-\theta)}\frac{1}{1-e^{-\delta}\theta},\end{aligned}$$

we have

$$\begin{aligned}\frac{\mathbb{E}\left[N\phi\left(N\right)\right]}{\sqrt{Var\left[N\right]}} &= \left(\frac{-\theta e^{-\delta}}{\left(1-\theta e^{-\delta}\right)\ln\left(1-\theta\right)} + \frac{\theta\ln\left(1-\theta e^{-\delta}\right)}{\left(1-\theta\right)\ln^{2}\left(1-\theta\right)}\right)\sqrt{\frac{\left(1-\theta\right)^{2}\ln^{2}\left(1-\theta\right)}{-\theta\left(\theta+\ln\left(1-\theta\right)\right)}} \\ &= \frac{-\left(1-\theta\right)\sqrt{\theta}}{\sqrt{-\left(\theta+\ln\left(1-\theta\right)\right)}}\left(\frac{-e^{-\delta}}{1-\theta e^{-\delta}} + \frac{\ln\left(1-\theta e^{-\delta}\right)}{\left(1-\theta\right)\ln\left(1-\theta\right)}\right),\end{aligned}$$

from where results the last formula of case (i).

ii) Let now $\phi\left(x\right) = p\left(x\right) - \sum_{n} p^{2}\left(n\right)$. If $N \sim Po\left(\mu\right)$ we have

$$E = \sum_{n\in\mathbb{N}} p^{2}\left(n\right) = \sum_{n\in\mathbb{N}} e^{-2\mu}\frac{\mu^{2n}}{\left(n!\right)^{2}} = e^{-2\mu}I_{0}\left(2\mu\right),$$

while

$$\begin{aligned}\mathbb{E}\left[N\phi(N)\right] &= \sum_{n=1}^{\infty} n\left(e^{-\mu}\frac{\mu^{n}}{n!}\right)^{2} - \mu E = \mu e^{-2\mu}\sum_{n=0}^{\infty}\frac{\mu^{2n+1}}{n!\left(n+1\right)!} - \mu e^{-2\mu}I_{0}\left(2\mu\right) \\ &= \mu e^{-2\mu}\left(I_{1}\left(2\mu\right) - I_{0}\left(2\mu\right)\right),\end{aligned}$$

which immediately yields the corresponding result. When $N \sim NB\left(r,q\right)$, we calculate

$$\begin{aligned}E &= \sum_{n\in\mathbb{N}}\left[\frac{\Gamma\left(r+n\right)}{\Gamma\left(r\right)}\right]^{2}\frac{q^{2r}\left(1-q\right)^{2n}}{\left(n!\right)^{2}} = q^{2r}\sum_{n=0}^{\infty}\frac{\left(1-q\right)^{2n}}{n!}\frac{\left[r\left(r+1\right)\times\ldots\times\left(r+n-1\right)\right]^{2}}{1\times2\times\ldots\times\left(1+n-1\right)} \\ &= q^{2r}\,{}_{2}F_{1}\left(\{r,r\},\{1\};\left(1-q\right)^{2}\right),\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}\left[Np(N)\right] &= q^{2r}\sum_{n=1}^{\infty} n(1-q)^{2n}\left(\frac{\Gamma(r+n)}{\Gamma(r)n!}\right)^2 = q^{2r}(1-q)^2\sum_{n=0}^{\infty}\frac{(1-q)^{2n}}{n!(n+1)!}\left(\frac{\Gamma(r+n+1)}{\Gamma(r)}\right)^2 \\
&= r^2q^{2r}(1-q)^2\sum_{n=0}^{\infty}\frac{(1-q)^{2n}}{n!}\frac{\left[(r+1)\times\ldots\times(r+1+n-1)\right]^2}{2\times\ldots\times(n+2-1)} \\
&= r^2q^{2r}(1-q)^2\ {}_2F_1\left(\{r+1,r+1\},\{2\};(1-q)^2\right).
\end{aligned}$$

Thus, based on the above, a straightforward calculation yields the stated formula. For $N\sim Log(\theta)$, we have

$$E=\sum_{n=1}^{\infty}p^2(n)=\frac{1}{\ln^2(1-\theta)}\sum_{n=1}^{\infty}\frac{\theta^{2n}}{n^2}=\frac{1}{\ln^2(1-\theta)}Li_2\left(\theta^2\right),$$

and

$$\mathbb{E}\left[Np(N)\right]=\sum_{n=1}^{\infty}\frac{n}{\ln^2(1-\theta)}\frac{\theta^{2n}}{n^2}=\frac{1}{\ln^2(1-\theta)}\sum_{n=1}^{\infty}\frac{\theta^{2n}}{n}=-\frac{\ln(1-\theta^2)}{\ln^2(1-\theta)}.$$

It follows that

$$\frac{\mathbb{E}\left[N\phi(N)\right]}{\sqrt{Var\left[N\right]}}=\left(\frac{\theta Li_2\left(\theta^2\right)}{(1-\theta)\ln^3(1-\theta)}-\frac{\ln(1-\theta^2)}{\ln^2(1-\theta)}\right)\sqrt{\frac{(1-\theta)^2\ln^2(1-\theta)}{-\theta\left(\theta+\ln(1-\theta)\right)}},$$

which easily completes the proof. □

Proof of Proposition 7.4.6. i) For $N\sim Po(\mu)$, we have$((fp)\vee h)(0)=(fp)(0)=p^2(0)=e^{-2\mu}$, while for $x>0$, using (7.2.6) and $(n+1)!=(2)_{(n)}$,

$$\begin{aligned}
((fp)\vee h)(x) &= \sum_{n=1}^{\infty}e^{-2\mu}\frac{\mu^{2n}}{(n!)^2}\frac{\beta^{nk}}{(nk-1)!}x^{nk-1}e^{-\beta x} \\
&= e^{-2\mu-\beta x}\frac{\mu^2(\beta x)^k}{x}\sum_{n=0}^{\infty}\frac{\left(\mu^2(\beta x)^k\right)^n}{\left[(n+1)!\right]^2((n+1)k-1)!} \\
&= e^{-2\mu-\beta x}\frac{\mu^2(\beta x)^k}{x(k-1)!}\sum_{n=0}^{\infty}\frac{\left(\mu^2(\beta x/k)^k\right)^n}{n!} \\
&\quad\times\frac{1}{\left(\frac{k+1}{k}\right)_{(n)}\left(\frac{k+2}{k}\right)_{(n)}\times\ldots\times\left(\frac{k+(k-1)}{k}\right)_{(n)}(2)_{(n)}(2)_{(n)}},
\end{aligned}$$

hence the stated formula. In the case (ii) when $N\sim NB(r,q)$, we obtain $((fp)\vee h)(0)=$

$p^2(0) = q^{2r}$, and for $x > 0$, using (7.2.6),

$$\begin{aligned}((fp) \vee h)(x) &= q^{2r} e^{-\beta x} \sum_{n=1}^{\infty} \left[\frac{\Gamma(r+n)}{\Gamma(r)}\right]^2 \frac{(1-q)^{2n}}{(n!)^2} \frac{\beta^{nk} x^{nk-1}}{(nk-1)!} \\ &= \frac{q^{2r} e^{-\beta x} (1-q)^2 (\beta x)^k}{x} \sum_{n=0}^{\infty} \frac{\left((1-q)^2 (\beta x)^k\right)^n}{n!} \\ &\quad \times \frac{[r(r+1) \times \ldots \times (r+1+n-1)]^2}{(n+1)(n+1)!((n+1)k-1)!} \\ &= \frac{r^2 q^{2r} e^{-\beta x} (1-q)^2 (\beta x)^k}{x(k-1)!} \sum_{n=0}^{\infty} \frac{\left((1-q)^2 (\beta x)^k\right)^n}{n! k^{nk}} \\ &\quad \times \frac{(r+1)_{(n)} (r+1)_{(n)}}{\left(\frac{k+1}{k}\right)_{(n)} \left(\frac{k+2}{k}\right)_{(n)} \times \ldots \times \left(\frac{k+(k-1)}{k}\right)_{(n)} (2)_{(n)} (2)_{(n)}},\end{aligned}$$

which immediately yields the result. To prove the formulas in case (iii) when $N \sim Log(\theta)$ and clearly $((fp) \vee h)(0) = 0$, for $x > 0$, we use

$$\begin{aligned}((fp) \vee h)(x) &= \frac{e^{-\beta x}}{\ln^2(1-\theta)} \sum_{n=1}^{\infty} \frac{\theta^{2n}}{n^2} \frac{\beta^{nk} x^{nk-1}}{(nk-1)!} = \frac{e^{-\beta x} \theta^2 (\beta x)^k}{x \ln^2(1-\theta)} \sum_{n=0}^{\infty} \frac{\left(\theta^2 (\beta x)^k\right)^n}{(n+1)^2 ((n+1)k-1)!} \\ &= \frac{e^{-\beta x} \theta^2 (\beta x)^k}{x \ln^2(1-\theta)} \sum_{n=0}^{\infty} \frac{\left(\theta^2 (\beta x)^k\right)^n}{n!} \frac{n!}{(n+1)((n+1)k-1)!} \frac{n!}{(n+1)!}.\end{aligned}$$

When $k = 1$, this gives

$$\begin{aligned}((fp) \vee h)(x) &= \frac{e^{-\beta x} \theta^2 \beta x}{x \ln^2(1-\theta)} \sum_{n=0}^{\infty} \frac{(\theta^2 \beta x)^n}{n!} \frac{n!}{(n+1)!} \frac{n!}{(n+1)!} \\ &= \frac{\theta^2 e^{-\beta x} \beta x}{x \ln^2(1-\theta)} \, {}_2F_2\left(\{1,1\},\{2,2\};\theta^2 \beta x\right),\end{aligned}$$

while for $k \geq 2$, we apply formula (7.2.6) and obtain the result. □

Proof of Proposition 7.4.8. $i)$ For $N \sim Po(\mu)$, we get

$$\begin{aligned}((fp)) \vee h)(x) &= \sum_{n=1}^{\infty} e^{-2\mu} \frac{\mu^{2n}}{(n!)^2} \frac{\Gamma(\alpha+n)}{\Gamma(\alpha)\Gamma(n)} \frac{x^{n-1}}{\sigma^n (1+x/\sigma)^{n+\alpha}} \\ &= \frac{e^{-2\mu}}{x(1+x/\sigma)^{\alpha}} \sum_{n=0}^{\infty} \frac{\Gamma(\alpha+1+n)}{\Gamma(\alpha)\Gamma(n+1)} \frac{\mu^{2(n+1)}}{[(n+1)!]^2} \left(\frac{x}{\sigma(1+x/\sigma)}\right)^{n+1} \\ &= \frac{\alpha \mu^2 e^{-2\mu}}{\sigma(1+x/\sigma)^{1+\alpha}} \sum_{n=0}^{\infty} \left(\frac{\mu^2 x}{\sigma+x}\right)^n \frac{1}{n!} \frac{(1+\alpha)_{(n)}}{(2)_{(n)} (2)_n},\end{aligned}$$

hence the result.
ii) For $N \sim NB(r,q)$, we have

$$\begin{aligned}
((fp)) \vee h)(x) &= \sum_{n=1}^{\infty}\left[\frac{\Gamma(n+r)}{\Gamma(r)n!}q^r(1-q)^n\right]^2 \frac{\Gamma(\alpha+n)}{\Gamma(\alpha)\Gamma(n)}\frac{x^{n-1}}{\sigma^n(1+x/\sigma)^{n+\alpha}} \\
&= \frac{q^{2r}}{x\left(1+\frac{x}{\sigma}\right)^{\alpha}}\sum_{n=0}^{\infty}\frac{(1-q)^{2n+2}}{n!}\left(\frac{x}{\sigma\left(1+\frac{x}{\sigma}\right)}\right)^{n+1}\left[\frac{\Gamma(r+1+n)}{\Gamma(r)(n+1)!}\right]^2\frac{\Gamma(\alpha+1+n)}{\Gamma(\alpha)} \\
&= \frac{\alpha r^2 q^{2r}(1-q)^2}{\sigma(1+x/\sigma)^{\alpha+1}}\sum_{n=0}^{\infty}\left(\frac{(1-q)^2 x}{\sigma+x}\right)^n \frac{1}{n!}\frac{(1+\alpha)_{(n)}(1+r)_{(n)}(1+r)_{(n)}}{(2)_{(n)}(2)_{(n)}},
\end{aligned}$$

yielding the result. In the case (*iii*) where $N \sim Log(\theta)$, we write

$$\begin{aligned}
((fp)) \vee h)(x) &= \left[\frac{1}{\ln(1-\theta)}\right]^2\sum_{n=1}^{\infty}\left(\frac{\theta^n}{n}\right)^2\frac{\Gamma(\alpha+n)}{\Gamma(\alpha)\Gamma(n)}\frac{x^{n-1}}{\sigma^n(1+x/\sigma)^{n+\alpha}} \\
&= \frac{1}{\ln^2(1-\theta)}\frac{1}{x(1+x/\sigma)^{\alpha}}\sum_{n=0}^{\infty}\frac{\theta^{2n+2}}{(n+1)^2}\frac{\Gamma(\alpha+n+1)}{\Gamma(\alpha)\Gamma(n+1)}\left(\frac{x}{\sigma(1+x/\sigma)}\right)^{n+1} \\
&= \frac{1}{\ln^2(1-\theta)}\frac{\alpha\theta^2}{\sigma(1+x/\sigma)^{\alpha+1}}\sum_{n=0}^{\infty}\left(\frac{\theta^2 x}{\sigma+x}\right)^n\frac{1}{n!}\frac{(\alpha+1)_{(n)}}{(n+1)^2},
\end{aligned}$$

which, using $\frac{1}{n+1} = \frac{n!}{(n+1)!}$, leads to the stated formula. $\square$

Appendix A

Copula Models

A.1 Gumbel Copula

The Gumbel copula is an Archimedean copula with generator $\psi_\theta(t) = (-\ln t)^\theta$ and distribution function defined as follows

$$Q_\theta(v_1, v_2) = e^{-\left[(-\ln v_1)^\theta + (-\ln v_2)^\theta\right]^{\frac{1}{\theta}}},$$

where $\theta \geq 1$ is the dependence parameter.
The partial derivative of Q_θ with respect to v_1 is given by

$$\frac{\partial Q_\theta(v_1, v_2)}{\partial v_1} = \frac{1}{v_1}(-\ln v_1)^{\theta-1}\left[(-\ln v_1)^\theta + (-\ln v_2)^\theta\right]^{\frac{1}{\theta-1}} e^{-\left[(-\ln v_1)^\theta + (-\ln v_2)^\theta\right]^{\frac{1}{\theta}}}. \quad \text{(A.1.1)}$$

By differentiating (A.1.1) with respect to v_2, we get the joint density copula $q_\theta(v_1, v_2)$ defined below

$$\begin{aligned} q_\theta(v_1, v_2) &= \frac{(-\ln v_1)^{\theta-1}(-\ln v_2)^{\theta-1}}{v_1 v_2}\Big([(-\ln v_1)^\theta + (-\ln v_2)^\theta]^{\frac{2}{\theta}-2} \\ &\quad +(\theta-1)[(-\ln v_1)^\theta + (-\ln v_2)^\theta]^{\frac{1}{\theta}-2}\Big) e^{-\left[(-\ln v_1)^\theta + (-\ln v_2)^\theta\right]^{\frac{1}{\theta}}}. \end{aligned}$$

A.2 Frank Copula

The Frank copula is an Archimedean copula with generator $\psi_\theta(t) = -\ln\left(\frac{e^{-\theta t}-1}{e^{-\theta}-1}\right)$ and distribution function given by

$$Q_\theta(v_1, v_2) = \frac{-1}{\theta}\ln\Big(1 + \frac{(e^{-\theta v_1}-1)(e^{-\theta v_2}-1)}{e^{-\theta}-1}\Big),$$

where $\theta \neq 1$ is the dependence parameter.
The partial derivative of Q_θ with respect to v_1 is defined as follows

$$\frac{\partial Q_\theta(v_1, v_2)}{\partial v_1} = \frac{e^{-\theta v_1}(e^{-\theta v_2}-1)}{(e^{-\theta}-1)+(e^{-\theta v_1}-1)(e^{-\theta v_2}-1)}. \tag{A.2.1}$$

By differentiating (A.2.1) with respect to v_2, the joint density copula $q_\theta(v_1, v_2)$ can be expressed as follows

$$q_\theta(v_1, v_2) = \frac{\theta(1-e^{-\theta})e^{-\theta(v_1+v_2)}}{\Big[(1-e^{-\theta})-(1-e^{-\theta v_1})(1-e^{-\theta v_2})\Big]^2}.$$

A.3 Joe copula

The Joe copula is an Archimedean copula with generator $\psi_\theta(t) = -\ln(1-(1-t)^\theta)$ and distribution function defined as follows

$$Q_\theta(v_1, v_2) = 1 - \Big[(1-v_1)^\theta + (1-v_2)^\theta - (1-v_1)^\theta(1-v_2)^\theta\Big]^{\frac{1}{\theta}},$$

where $\theta \geq 1$ is the dependence parameter.
The partial derivative of Q_θ with respect to v_1 is given by

$$\frac{\partial Q_\theta(v_1, v_2)}{\partial v_1} = (1-v_1)^{\theta-1}(1-(1-v_2)^\theta)\Big((1-v_1)^\theta + (1-v_2)^\theta - (1-v_1)^\theta(1-v_2)^\theta\Big)^{\frac{1}{\theta}-1} \tag{A.3.1}$$

By differentiating (A.3.1) with respect to v_2, the joint density copula can be written as

$$\begin{aligned} q_\theta(v_1, v_2) &= (1-v_1)^{\theta-1}(1-v_2)^{\theta-1}\Big(\theta - 1 + (1-v_1)^\theta + (1-v_2)^\theta - (1-v_1)^\theta(1-v_2)^\theta\Big) \times \\ &\quad \Big((1-v_1)^\theta + (1-v_2)^\theta - (1-v_1)^\theta(1-v_2)^\theta\Big)^{\frac{1}{\theta}-2}. \end{aligned}$$

A.4 Clayton copula

The Clayton copula is an Archimedean copula with generator $\psi_\theta(t) = \frac{1}{\theta}\left(t^{-\theta} - 1\right)$ and distribution function defined as follows

$$\begin{aligned} Q_\theta(v_1, v_2) &= \left[\max\left(u^{-\theta} + v^{-\theta} - 1; 0\right)\right]^{-\frac{1}{\theta}} \\ &:= \left(u^{-\theta} + v^{-\theta} - 1; 0\right)^{-\frac{1}{\theta}}, \end{aligned}$$

where $\theta \in [-1, \infty) \setminus \{0\}$ is the dependence parameter.
The partial derivative of Q_θ with respect to v_1 is given by

$$\frac{\partial Q_\theta(v_1, v_2)}{\partial v_1} = v_1^{-(\theta+1)}\left(v_1^{-\theta} + v_2^{-\theta} - 1\right)^{-\frac{1}{\theta}-1}. \tag{A.4.1}$$

By differentiating (A.4.1) with respect to v_2, the joint density copula can be expressed as follows

$$q_\theta(v_1, v_2) = (\theta+1)(v_1 v_2)^{-(\theta+1)}\left(v_1^{-\theta} + v_2^{-\theta} - 1\right)^{-\frac{1}{\theta}-2}.$$

A.5 Student Copula

The distribution function of the Student copula with dependence parameter $\theta \in (-1, 1)$ and m degrees of freedom is defined as follows

$$\begin{aligned} Q_{\theta,m}(v_1, v_2) &= t_{\theta,m}(t_m^{-1}(v_1), t_m^{-1}(v_2)) \\ &= \int_{-\infty}^{t_m^{-1}(v_1)} \int_{-\infty}^{t_m^{-1}(v_2)} \frac{1}{2\pi(1-\theta^2)^{1/2}}\left[1 + \frac{s^2 - 2\theta st + t^2}{m(1-\theta^2)}\right]^{-(m+2)/2} ds\, dt. \end{aligned}$$

The partial derivative with respect to v_1 is given by

$$\frac{\partial Q_{\theta,m}(v_1, v_2)}{\partial v_1} = t_{m+1}\left[\left(t_m^{-1}(v_2) - \theta t_m^{-1}(v_1)\right) \Big/ \left(\sqrt{\frac{(m + (t_m^{-1}(v_1))^2)(1-\theta^2)}{m+1}}\right)\right] \tag{A.5.1}$$

By differentiating (A.5.1) with respect to v_2, we get the joint density copula $q_\theta(v_1, v_2)$ defined below

$$q_{\theta,m}(v_1, v_2) = \frac{1}{2\pi\sqrt{1-\theta^2}} \frac{1}{dt(t_m^{-1}(v_1))\, dt(t_m^{-1}(v_2))}\left(1 + \frac{t_m^{-1}(v_1)^2 + t_m^{-1}(v_2)^2 - 2\theta t_m^{-1}(v_1) t_m^{-1}(v_2)}{m(1-\theta^2)}\right)^{-\frac{m+2}{2}},$$

where $dt(t_m^{-1}(v_i)) = \frac{\Gamma(\frac{m+1}{2})}{\Gamma(\frac{m}{2})\sqrt{\pi m}}\left(1 + \frac{t_m^{-1}(v_i)^2}{m}\right)^{-\frac{m+1}{2}}$ for $i = 1, 2$.

Bibliography

[1] A. Abdallah, J. P. Boucher, H. Cossette, and J. Trufin. Sarmanov family of bivariate distributions for multivariate loss reserving analysis. *North American Actuarial Journal*, 20(2):184–200, 2016.

[2] H. Ammeter. Note concerning the distribution function of the total loss excluding the largest individual claims. *Astin Bulletin*, 3(02):132–143, 1964.

[3] Barry C Arnold. *Pareto distribution*. Wiley Online Library, 2015.

[4] JS Arora, MW Huang, and CC Hsieh. Methods for optimisation of nonlinear problems with discrete variables: a review. *Structural optimisation*, 8(2-3): 69–85, 1994.

[5] Alexandru V Asimit, Edward Furman, and Raluca Vernic. On a multivariate pareto distribution. *Insurance: Mathematics and Economics*, 46(2):308–316, 2010.

[6] Alexandru V. Asimit, Alexandru M. Badescu, and Ka Chun Cheung. Optimal reinsurance in the presence of counterparty default risk. *Insurance Math. Econom.*, 53(3):690–697, 2013. ISSN 0167-6687. doi: 10.1016/j.insmatheco.2013.09.012. URL http://dx.doi.org/10.1016/j.insmatheco.2013.09.012.

[7] Alexandru V. Asimit, Alexandru M. Badescu, and Tim Verdonck. Optimal risk transfer under quantile-based risk measures. *Insurance Math. Econom.*, 53(1):252–265, 2013. ISSN 0167-6687. doi: 10.1016/j.insmatheco.2013.05.005. URL http://dx.doi.org/10.1016/j.insmatheco.2013.05.005.

[8] Alexandru V Asimit, Raluca Vernic, and Riċardas Zitikis. Evaluating risk measures and capital allocations based on multi-losses driven by a heavy-tailed background risk: The multivariate pareto-ii model. *Risks*, 1(1):14–33, 2013.

[9] Alexandru V. Asimit, Alexandru M. Badescu, Tak Kuen Siu, and Yuriy Zinchenko. Capital requirements and optimal investment with solvency probability constraints. *IMA J. Manag. Math.*, 26(4):345–375, 2015. ISSN 1471-678X. doi: 10.1093/imaman/dpt029. URL http://dx.doi.org/10.1093/imaman/dpt029.

[10] Alexandru V. Asimit, Yichun Chi, and Junlei Hu. Optimal non-life reinsurance under Solvency II regime. *Insurance Math. Econom.*, 65:227–237, 2015. ISSN 0167-6687.

[11] Alexandru V. Asimit, Alexandru M. Badescu, Steven Haberman, and Eun-Seok Kim. Efficient risk allocation within a non-life insurance group under Solvency II regime. *Insurance Math. Econom.*, 66:69–76, 2016. ISSN 0167-6687.

[12] A.V. Asimit and Y. Chen. Asymptotic results for conditional measures of association of a random sum. *Insurance Math. Econom.*, 60:11–18, 2015. ISSN 0167-6687. doi: 10.1016/j.insmatheco.2014.10.012. URL http://dx.doi.org/10.1016/j.insmatheco.2014.10.012.

[13] Søren Asmussen. *Ruin probabilities*, volume 2 of *Advanced Series on Statistical Science & Applied Probability*. World Scientific Publishing Co., Inc., River Edge, NJ, 2000. ISBN 981-02-2293-9. doi: 10.1142/9789812779311. URL http://dx.doi.org/10.1142/9789812779311.

[14] S. Aulbach, M. Falk, M. Hofmann, and M. Zott. Max-stable processes and the functional D-norm revisited. *Extremes*, 18(2):191–212, 2015. ISSN 1386-1999. doi: 10.1007/s10687-014-0210-0. URL http://dx.doi.org/10.1007/s10687-014-0210-0.

[15] S. Aulbach, M. Falk, and M. Zott. The space of D-norms revisited. *Extremes*, 18(1):85–97, 2015. ISSN 1386-1999. doi: 10.1007/s10687-014-0204-y. URL http://dx.doi.org/10.1007/s10687-014-0204-y.

[16] Michael Bartholomew-Biggs. *Nonlinear optimization with engineering applications*, volume 19. Springer Science & Business Media, 2008.

[17] Paul T. Boggs and Jon W. Tolle. Sequential quadratic programming. In *Acta numerica*, Acta Numer., pages 1–51. Cambridge Univ. Press, Cambridge, 1995.

[18] Paul T. Boggs and Jon W. Tolle. Sequential quadratic programming for large-scale nonlinear optimisation. *J. Comput. Appl. Math.*, 124(1-2):123–137, 2000.

ISSN 0377-0427. doi: 10.1016/S0377-0427(00)00429-5. URL http://dx.doi.org/10.1016/S0377-0427(00)00429-5. Numerical analysis 2000, Vol. IV, optimisation and nonlinear equations.

[19] A. Bücher and J. Segers. Extreme value copula estimation based on block maxima of a multivariate stationary time series. *Extremes*, 17(3):495–528, 2014. ISSN 1386-1999. doi: 10.1007/s10687-014-0195-8. URL http://dx.doi.org/10.1007/s10687-014-0195-8.

[20] A.C. Cebrian, M. Denuit, P. Lambert, et al. Analysis of bivariate tail dependence using extreme value copulas: An application to the soa medical large claims database. *Belgian Actuarial Journal*, 3(1):33–41, 2003.

[21] Yichun Chi and Hui Meng. Optimal reinsurance arrangements in the presence of two reinsurers. *Scand. Actuar. J.*, 2014(5):424–438, 2014. ISSN 0346-1238. doi: 10.1080/03461238.2012.723638. URL http://dx.doi.org/10.1080/03461238.2012.723638.

[22] H. Cossette, M. Mailhot, and E. Marceau. TVaR-based capital allocation for multivariate compound distributions with positive continuous claim amounts. *Insurance: Mathematics & Economics*, 50(2):247–256, 2012. ISSN 0167-6687. doi: 10.1016/j.insmatheco.2011.11.006. URL http://dx.doi.org/10.1016/j.insmatheco.2011.11.006.

[23] H. Cossette, M.-P. Côté, E. Marceau, and K. Moutanabbir. Multivariate distribution defined with Farlie-Gumbel-Morgenstern copula and mixed Erlang marginals: Aggregation and capital allocation. *Insurance: Mathematics & Economics*, 52:560–572, 2013.

[24] Robert J Dakin. A tree-search algorithm for mixed integer programming problems. *The computer journal*, 8(3):250–255, 1965.

[25] K. Debicki, E. Hashorva, and L. Ji. Parisian ruin of self-similar gaussian risk processes. *J. Appl. Probab*, 52:688–702, 2015.

[26] P Deheuvels. La fonction de dépendance empirique et ses propriétés. académie royale de belgique. *Bulletin de la Classe des Sciences*, 65(5):274–292, 1979.

[27] M. Denuit, J. Dhaene, M. Goovaerts, and R. Kass. *Actuarial Theory for Dependent Risks: Measures, Orders and Models*. Wiley, 2005.

[28] M. Denuit, O. Purcaruff, and I. Van Keilegorni. Bivariate archimedean copula models for censored data in non-life insurance. Technical report, Université catholique de Louvain, 2006.

[29] Michel Denuit, Jan Dhaene, Marc Goovaerts, and Rob Kaas. *Actuarial theory for dependent risks: measures, orders and models*. John Wiley & Sons, 2006.

[30] L. Devroye. A triptych of discrete distributions related to the stable law. *Statistics & Probability Letters*, 18(5):349–351, 1993.

[31] J. Dhaene and M.J. Goovaerts. Dependency of risks and stop-loss order. *Astin Bulletin*, 26(02):201–212, 1996.

[32] J. Dhaene, A. Tsanakas, E.A. Valdez, and S. Vanduffel. Optimal capital allocation principles. *Journal of Risk and Insurance*, 79(1):1–28, 2012.

[33] Peter K Dunn and Gordon K Smyth. Series evaluation of tweedie exponential dispersion model densities. *Statistics and Computing*, 15(4):267–280, 2005.

[34] P. Embrechts. Copulas: a personal view. *Journal of Risk and Insurance*, 76 (3):639–650, 2009.

[35] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling extremal events*, volume 33 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1997. For insurance and finance.

[36] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling extremal events for insurance and finance*. Springer-Verlag, Berlin, 1997.

[37] Paul Embrechts, Alexander McNeil, and Daniel Straumann. Correlation and dependence in risk management: properties and pitfalls. *Risk management: value at risk and beyond*, 176223, 2002.

[38] M. Falk, J. Hüsler, and R.-D. Reiss. Laws of small numbers: extremes and rare events. In *DMV Seminar*, volume 23, page 3rd edn. Birkhäuser, Basel, 2010.

[39] Kris Johnson Ferreira, Bin Hong Alex Lee, and David Simchi-Levi. Analytics for an online retailer: Demand forecasting and price optimization. *Manufacturing & Service Operations Management*, 18(1):69–88, 2015.

[40] Roger Fletcher and Michael JD Powell. A rapidly convergent descent method for minimization. *The computer journal*, 6(2):163–168, 1963.

[41] Marguerite Frank and Philip Wolfe. An algorithm for quadratic programming. *Naval research logistics quarterly*, 3(1-2):95–110, 1956.

[42] G. A. Fredricks and R. B. Nelsen. On the relationship between spearman's rho and kendall's tau for pairs of continuous random variables. *Journal of statistical planning and inference*, 137(7):2143–2150, 2007.

[43] E.W. Frees, V.R. Young, and Y. Luo. Case studies using panel data models. *North American Actuarial Journal*, 5(4):24–42, 2001.

[44] Harish Garg. A hybrid pso-ga algorithm for constrained optimization problems. *Applied Mathematics and Computation*, 274:292–305, 2016.

[45] C. Genest, K. Ghoudi, and L.P. Rivest. A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82(3):543–552, 1995.

[46] C. Genest, M. Gendron, and M. Bourdeau-Brien. The advent of copulas in finance. *The European Journal of Finance*, 15(7-8):609–618, 2009.

[47] Christian Genest and Anne-Catherine Favre. Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of hydrologic engineering*, 12(4):347–368, 2007.

[48] Christian Genest and Jock MacKay. The joy of copulas: Bivariate distributions with uniform marginals. *The American Statistician*, 40(4):280–283, 1986.

[49] Christian Genest and Louis-Paul Rivest. Statistical inference procedures for bivariate archimedean copulas. *Journal of the American statistical Association*, 88(423):1034–1043, 1993.

[50] Christian Genest, Étienne Marceau, and Mhamed Mesfioui. Compound poisson approximations for individual models with dependent risks. *Insurance: Mathematics and Economics*, 32(1):73–91, 2003.

[51] Christian Genest, Bruno Rémillard, and David Beaudoin. Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and economics*, 44(2):199–213, 2009.

[52] A. Y. Golubin. Optimal insurance and reinsurance policies chosen jointly in the individual risk model. *Scand. Actuar. J.*, 2016(3):181–197, 2016. ISSN 0346-1238.

[53] Marc J Goovaerts and Jan Dhaene. The compound poisson approximation for a portfolio of dependent risks. *Insurance: Mathematics and Economics*, 18(1): 81–85, 1996.

[54] K.L. Grazier, K. G'Sell Associates, and L.P. Rivest. Group medical insurance large claims database and collection. *SOA Monograph M-HB97-1*, 1997.

[55] Montserrat Guillén, Ana Maria Perez, and Manuela Alcañiz. A logistic regression approach to estimating customer profit loss due to lapses in insurance. *Document de Treball No. XREAP*, 13, 2011.

[56] Montserrat Guillén, José María Sarabia, and Faustino Prieto. Simple risk measure calculations for sums of positive random variables. *Insurance: Mathematics and Economics*, 53(1):273–280, 2013.

[57] B.E. Hansen. Bandwidth selection for nonparametric distribution estimation. *manuscript, University of Wisconsin*, 2004.

[58] E Hashorva. Extremes of conditioned elliptical random vectors. *J. Multivariate Anal.*, 98(8):1583–1591, 2007. ISSN 0047-259X.

[59] E. Hashorva and G. Ratovomirija. On Sarmanov mixed Erlang risks in insurance applications. *ASTIN Bulletin*, 45:175–205, 2015. ISSN 1783-1350. doi: 10.1017/asb.2014.24. URL http://journals.cambridge.org/article_S0515036114000245.

[60] E. Hashorva, G. Ratovomirija, and M. Tamraz. On some new dependence models derived from multivariate collective models in insurance applications. *Scandinavian Actuarial Journal*, 0(0):1–21, 0. doi: 10.1080/03461238.2016.1243574. URL http://dx.doi.org/10.1080/03461238.2016.1243574.

[61] Enkelejd Hashorva, Gildas Ratovomirija, Maissa Tamraz, and Yizhou Bai. Some mathematical aspects of price optimisation. *Scandinavian Actuarial Journal*, 0(0):1–25, 2017. doi: 10.1080/03461238.2017.1357654. URL http://dx.doi.org/10.1080/03461238.2017.1357654.

[62] S. Haug, C. Klüppelberg, and L. Peng. Statistical models and methods for dependence in insurance data. *Journal of the Korean Statistical Society*, 40 (2):125–139, 2011.

[63] A. Hernández-Bastida and M. P. Fernández-Sánchez. A Sarmanov family with beta and gamma marginal distributions: an application to the Bayes premium

in a collective risk model. *Statistical Methods & Applications*, 21(4):391–409, 2012.

[64] Hung-Hsi Huang, Yung-Ming Shiu, and Ching-Ping Wang. Optimal insurance contract with stochastic background wealth. *Scand. Actuar. J.*, 2013(2):119–139, 2013. ISSN 0346-1238.

[65] Tao Jin and Jiandong Ren. Recursions and fast fourier transforms for a new bivariate aggregate claims model. *Scandinavian Actuarial Journal*, 2014(8): 729–752, 2014.

[66] H. Joe. *Dependence modeling with copulas*, volume 134 of *Monographs on Statistics and Applied Probability.* CRC Press, Boca Raton, FL, 2015. ISBN 978-1-4665-8322-1.

[67] Harry Joe. *Multivariate models and multivariate dependence concepts.* CRC Press, 1997.

[68] B Jorgensen. The theory of dispersion models. chapman and hall. *London, UK*, 1997.

[69] M. Kanter. Stable densities under change of scale and total variation inequalities. *The Annals of Probability*, 3(4):697–707, 1975.

[70] G. Kim, M.J. Silvapulle, and P. Silvapulle. Comparison of semiparametric and parametric methods for estimating copulas. *Computational Statistics and Data Analysis*, 51(6):2836–2850, 2007.

[71] Christian Kleiber and Samuel Kotz. *Statistical size distributions in economics and actuarial sciences*, volume 470. John Wiley & Sons, 2003.

[72] Stuart A Klugman and Rahul Parsa. Fitting bivariate loss distributions with copulas. *Insurance: mathematics and economics*, 24(1):139–148, 1999.

[73] Stuart A Klugman, Harry H Panjer, and Gordon E Willmot. *Loss models: from data to decisions*, volume 715. John Wiley & Sons, 2012.

[74] Erik Kole, Kees Koedijk, and Marno Verbeek. Selecting copulas for risk management. *Journal of Banking & Finance*, 31(8):2405–2423, 2007.

[75] Samuel Kotz, N Balakrishnan, and Norman L Johnson. Bivariate and trivariate normal distributions. *Continuous Multivariate Distributions: Models and Applications, Volume 1, Second Edition*, pages 251–348, 2000.

[76] R. Kulik and P. Soulier. Heavy tailed time series with extremal independence. *Extremes*, 18(2):273–299, 2015. ISSN 1386-1999. doi: 10.1007/s10687-014-0213-x. URL http://dx.doi.org/10.1007/s10687-014-0213-x.

[77] M. L. T. Lee. Properties and applications of the Sarmanov family of bivariate distributions. *Communications in Statistics-Theory and Methods*, 25(6):1207–1222, 1996.

[78] S. C. K. Lee and X. S. Lin. Modeling and evaluating insurance losses via mixtures of Erlang distributions. *North American Actuarial Journal*, 14(1): 107–130, 2010. ISSN 1092-0277. doi: 10.1080/10920277.2010.10597580. URL http://dx.doi.org/10.1080/10920277.2010.10597580.

[79] S. C. K. Lee and X. S. Lin. Modeling dependent risks with multivariate Erlang mixtures. *Astin Bulletin*, 42(1):153–180, 2012. ISSN 0515-0361.

[80] Han Tong Loh and Panos Y Papalambros. A sequential linearization approach for solving mixed-discrete nonlinear design optimisation problems. *Journal of Mechanical Design*, 113(3):325–334, 1991.

[81] David G. Luenberger and Yinyu Ye. *Linear and nonlinear programming*. International Series in Operations Research & Management Science, 228. Springer, Cham, fourth edition, 2016. ISBN 978-3-319-18841-6; 978-3-319-18842-3. doi: 10.1007/978-3-319-18842-3. URL http://dx.doi.org/10.1007/978-3-319-18842-3.

[82] Cheng Ma and Liansheng Zhang. On an exact penalty function method for nonlinear mixed discrete programming problems and its applications in search engine advertising problems. *Applied Mathematics and Computation*, 271: 642–656, 2015.

[83] D.D. Mari and S. Kotz. *Correlation and dependence*, volume 518. World Scientific, 2001.

[84] Angel Marin and Thomas Bayley. Price optimisation for new business profit and growth. *Society of Actuaries in Ireland*, 2010.

[85] H. M. Markowitz. *Portfolio Selection: Efficient Diversification of Investments*. John Wiley & Sons, New-York, 1959.

[86] H. M. Markowitz, G Peter Todd, and William F Sharpe. *Mean-variance analysis in portfolio choice and capital markets*, volume 66. John Wiley & Sons, 2000.

[87] Harry Markowitz. Portfolio selection. *The journal of finance*, 7(1):77–91, 1952.

[88] Harry Markowitz. The optimization of a quadratic function subject to linear constraints. *Naval research logistics Quarterly*, 3(1-2):111–133, 1956.

[89] Joseph W Meador, Gerald P Madden, and David J Johnston. On the probability of acquisition of non-life insurers. *Journal of Risk and Insurance*, pages 621–643, 1986.

[90] Melanie Mitchell. *An introduction to genetic algorithms*. MIT press, 1998.

[91] I. Molchanov. Convex geometry of max-stable distributions. *Extremes*, 11 (3):235–259, 2008. ISSN 1386-1999. doi: 10.1007/s10687-008-0055-5. URL http://dx.doi.org/10.1007/s10687-008-0055-5.

[92] NAIC. Price optimization white paper. *Casualty Actuarial and Statistical (C) Task Force, Casualty Actuarial Society Committee on Ratemaking Price Optimization Working Party*, 2015.

[93] R.B. Nelsen. *An introduction to copulas*, volume 139 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1999. ISBN 0-387-98623-5.

[94] Roger B Nelsen. Introduction. In *An Introduction to Copulas*, pages 1–4. Springer, 1999.

[95] Tristan Nguyen and Robert Danilo Molinari. Solvency ii–considering risk dependencies. *Journal 2005, Vol*, 8(3):428–454, 2010.

[96] Ronald Harlan Nickel and JW Tolle. A sequential quadratic programming algorithm for solving large, sparse nonlinear programs. Technical report, DTIC Document, 1984.

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

[98] L. Peng. Joint tail of ECOMOR and LCR reinsurance treaties. *Insurance Math. Econom.*, 58:116–120, 2014. ISSN 0167-6687. doi: 10.1016/j.insmatheco.2014.06.013. URL http://dx.doi.org/10.1016/j.insmatheco.2014.06.013.

[99] Robert Lewis Phillips. *Pricing and revenue optimization*. Stanford University Press, 2005.

[100] G. Ratovomirija. On mixed Erlang reinsurance risk: aggregation, capital allocation and default risk. *European Actuarial Journal*, pages 1–27, 2016.

[101] Darryn J Reid. Genetic algorithms in constrained optimization. *Mathematical and computer modelling*, 23(5):87–111, 1996.

[102] S.I. Resnick and D. Zeber. Asymptotics of Markov kernels and the tail chain. *Adv. in Appl. Probab.*, 45(1):186–213, 2013. ISSN 0001-8678. URL http://projecteuclid.org/euclid.aap/1363354108.

[103] Elena-Gratiela Robe-Voinea and Raluca Vernic. On the recursive evaluation of a certain multivariate compound distribution. *Acta Mathematicae Applicatae Sinica, English Series*, 32(4):913–920, 2016.

[104] T. Rolski, H. Schmidli, V. Schmidt, and J. Teugels. *Stochastic processes for insurance and finance*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester, 1999. ISBN 0-471-95925-1. doi: 10.1002/9780470317044. URL http://dx.doi.org/10.1002/9780470317044.

[105] Vladimir I. Rotar. *Actuarial models*. CRC Press, Boca Raton, FL, second edition, 2015. ISBN 978-1-4822-2706-2. The mathematics of insurance.

[106] José María Sarabia, Emilio Gómez-Déniz, Faustino Prieto, and Vanesa Jordá. Risk aggregation in multivariate dependent pareto distributions. *Insurance: Mathematics and Economics*, 71:154–163, 2016.

[107] O. V. Sarmanov. Generalized normal correlation and two-dimensional Fréchet classes. *Doklady Akademii Nauk SSSR*, 168:32–35, 1966. ISSN 0002-3264.

[108] Arthur J Schwartz and Joseph G Harrington. Price optimization: The controversy, the future, and what it means for cpcus. *CPCU Society INSIGHTS*, 2015.

[109] N Ravi Shankar, G Ananda Rao, J Madhu Latha, and V Sireesha. Solving a fuzzy nonlinear optimization problem by genetic algorithm. *Int. J. Contemp. Math. Sciences*, 5(16):791–803, 2010.

[110] Joanna H Shih and Thomas A Louis. Inferences on the association parameter in copula models for bivariate survival data. *Biometrics*, pages 1384–1399, 1995.

[111] Abe Sklar. Random variables, joint distribution functions, and copulas. *Kybernetika*, 9(6):449–460, 1973.

[112] FW. Steutel and K. Van Harn. Discrete analogues of self-decomposability and stability. *The Annals of Probability*, pages 893–899, 1979.

[113] Bjørn Sundt and Raluca Vernic. *Recursions for convolutions and compound distributions with insurance applications.* Springer Science & Business Media, 2009.

[114] D. Tasche. Allocating portfolio economic capital to sub-portfolios. *Economic Capital: A Practitioner's Guide, Risk Books*, pages 275–302, 2004.

[115] R Guy Thomas. Non-risk price discrimination in insurance: market outcomes and public policy. *The Geneva Papers on Risk and Insurance-Issues and Practice*, 37(1):27–46, 2012.

[116] Sander Vandenberghe, NEC Verhoest, and Bernard De Baets. Fitting bivariate copulas to the dependence structure between storm characteristics: A detailed analysis based on 105 year 10 min rainfall. *Water resources research*, 46(1), 2010.

[117] R. Vernic. Capital allocation for Sarmanov's class of distributions. *Methodology and Computing in Applied Probability. To appear*, 2015.

[118] R. Vernic. On the distribution of a sum of Sarmanov distributed random variables. *Journal of Theoretical Probability*, 29(1):118–142, 2016.

[119] Raluca Vernic. On the evaluation of some multivariate compound distributions with sarmanov's counting distribution. *SSRN, submitted*, 2016.

[120] Costas Voglis and Isaac E Lagaris. Boxcqp: An algorithm for bound constrained convex quadratic problems. In *Proceedings of the 1st International Conference: From Scientific Computing to Computational Engineering, IC-SCCE, Athens, Greece*, 2004.

[121] G. E. Willmot and X. S. Lin. Risk modeling with the mixed Erlang distribution. *Applied Stochastic Models in Business and Industry*, 27(1):8–22, 2011. ISSN 1524-1904. doi: 10.1002/asmb.838. URL http://dx.doi.org/10.1002/asmb.838.

[122] G. E. Willmot and J.K. Woo. On some properties of a class of multivariate Erlang mixtures with insurance applications. *ASTIN Bulletin*, 45:151–173, 1

2015. ISSN 1783-1350. doi: 10.1017/asb.2014.23. URL http://journals.cambridge.org/article_S0515036114000233.

[123] Gordon E Willmot and X Sheldon Lin. Risk modelling with the mixed erlang distribution. *Applied Stochastic Models in Business and Industry*, 27(1):2–16, 2011.

[124] Gordon E Willmot and Jae-Kyung Woo. On some properties of a class of multivariate erlang mixtures with insurance applications. *ASTIN Bulletin: The Journal of the IAA*, 45(1):151–173, 2015.

[125] Y. Yang and E. Hashorva. Extremes and products of multivariate AC-product risks. *Insurance: Mathematics & Economics*, 52(2):312–319, 2013. ISSN 0167-6687. doi: 10.1016/j.insmatheco.2013.01.005. URL http://dx.doi.org/10.1016/j.insmatheco.2013.01.005.

[126] Özgür Yeniay. Penalty function methods for constrained optimization with genetic algorithms. *Mathematical and Computational Applications*, 10(1):45–56, 2005.

[127] K. Zhang and J. Lin. A new class of copulas involved geometric distribution: Estimation and applications. *Insurance Math. Econom.*, 66:1–10, 2016.

[128] Lianzeng Zhang, Xiang Hu, and Baige Duan. Optimal reinsurance under adjustment coefficient measure in a discrete risk model based on Poisson MA(1) process. *Scand. Actuar. J.*, 2015(5):455–467, 2015. ISSN 0346-1238. doi: 10.1080/03461238.2013.849615. URL http://dx.doi.org/10.1080/03461238.2013.849615.