PRICING FINANCIAL AND INSURANCE PRODUCTS IN THE MULTIVARIATE SETTING

Alice Pignatelli di Cerchiara

Department of Statistics
The London School of Economics and Political Science

A thesis submitted for the degree of Doctor of Philosophy

August 2021

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it). The copyright of this thesis rests with the author. Quotation from it is permitted, provided that full acknowledgement is made.

In accordance with the Regulations, I have deposited an electronic copy of it in LSE Theses Online held by the British Library of Political and Economic Science and have granted permission for my thesis to be made available for public reference. Otherwise, this thesis may not be reproduced without my prior written consent. I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

Statement of co-authored work

I confirm that parts of Chapter 3 were adapted into a paper entitled "The multivariate mixed Negative Binomial regression model with an application to insurance a posteriori ratemaking" jointly co-authored with Dr George Tzougas, which is currently under review.

I confirm that parts of Chapter 4 were adapted into a paper entitled "EM estimation for a new class of bivariate mixed Poisson regression models with varying dispersion: an application to a posteriori ratemaking" jointly co-authored with Dr George Tzougas, which is currently under review.

Acknowledgements

I have been blessed with three incredible supervisors, to them goes all my gratitude. To my main supervisor Dr Erik Baurdoux and to my two advisors Professor Angelos Dassios and Dr George Tzougas: I could have never done it without your amazing guidance, patience, effort and unparalleled kindness. Your enthusiasm and energy has always been of great inspiration to me, especially during those times in which I felt hopelessly stuck with my research.

I am deeply grateful to my examiners Professor Pauline Barrieu and Professor Stéphane Loisel for taking the time to read my thesis, writing detailed reports and giving me very useful feedback, which resulted in a much better thesis than the one I had originally drafted: your support, encouragement and kindness has had a crucial impact.

Dr Tobias Kley, I have learnt more math from you than in all my Bachelor and Master's years: I do not know how you do it, but I wish that one day I can become half as knowledgeable as you are.

Professor Andrea Primo and Professor Diego Zappa, thank you for making me the person I am now, I could have never achieved any of this if it were not for your invaluable lessons and encouragement, you have always trusted me more than I trust myself.

I am immensely thankful to the Economic and Social Research Council which provided me with the financial support to undertake this life-changing journey.

To Penny Montague and Ian Marshall, thank you for your continuous support and care throughout the years.

Ragvir Sabharwal and Bandna Rekhi, you are two of the most amazing people I could have ever hoped to meet. Ragvir, thank you for all the laughs and the silliness, but also for the amazing advice and help in times of need. Neha, thank you for your incredible sweetness and wisdom, for the best food of my life, the lovely knitted clothes and all the medications when I was sick.

Dr Tayfun Terzi, the first year of my PhD was the absolute best, it was never half as enjoyable after you left.
Dr President Viet Dang, thank you for all the lovely moments together and all the events.
Eduardo Ferioli Gomes, Lucia Guastadisegni, JingHan Tee, Camilo Cárdenas Hurtado and Shakeel Gavioli-Akilagun, I wish we had started and ended our PhDs journeys all at the same time, all the moments together were absolutely indescribable and I wish we could have had more of those.
To Filippo Pellegrino, Dr Xiaolin Zhu and all of former Second Floor members, my PhD would not have been half as fun without you.
To Dr José Manuel Pedraza Ramírez, I am so grateful we got to share this experience together from the very first day.
To my sister Dr Arianna Ponzini, thank you for all the light you bring into my life: our unbreakable bond has only been strengthened by this further journey we got to share. Also, thank you and my brother Jacopo Bertolone for all the magical times together in London and Oxford, they are some of my fondest memories.
Dr Nicholas Cron, I cannot imagine of better teas and chats than all the ones we had together. I am really grateful I had the privilege to be your coffee buddy all these years, you have been of huge comfort and unparalleled company on those days in which I felt nothing was going right.
Dr Julia Biggane, thank you for being such an amazing friend, doing a PhD is such a damaging experience for one's self-esteem and you have always managed to compensate for that with your incredible kindness and trust.
To my friends from Milan, I would have never done it without your constant love and help behind the scenes, a special thank you to Paola Liberatore, Andrea Lovati, Stefano Duse, Yuarvi Rómulo Díaz and Giovanni Tritto.
To Luca Fraone, you have done too much for me to be able to express it here, therefore

just thank you.

To my parents, thank you for your unconditional love and for always rooting for me, it made everything possible.

Abstract

In finance and insurance there is often the need to construct multivariate distributions to take into account more than one source of risk, where such risks cannot be assumed to be independent. In the course of this thesis we are going to explore three models, namely the copula models, the trivariate reduction scheme and mixtures as candidate models for capturing the dependence between multiple sources of risk. This thesis contains results of three different projects. The first one is in financial mathematics, more precisely on the pricing of financial derivatives (multi-asset options) which depend on multiple underlying assets, where we construct the dependence between such assets using copula models and the trivariate reduction scheme. The second and the third projects are in actuarial mathematics, more specifically on the pricing of the premia that need to be paid by policyholders in the automobile insurance when more than one type of claim is considered. We do the pricing including all the information available about the characteristics of the policyholders and their cars (i.e. a priori ratemaking) and about the numbers of claims per type in which the policyholders have been involved (i.e. a posteriori ratemaking). In both projects we model the dependence between the multiple types of claims using mixture distributions/regression models: we consider the different types of claims to be modelled in terms of their own distribution/regression model but with a common heterogeneity factor which follows a mixing distribution/regression model that is responsible for the dependence between the multiple types of claims. In the second project we present a new model (i.e. the bivariate Negative Binomial-Inverse Gaussian regression model) and in the third one we present a new family of models (i.e. the bivariate mixed Poisson regression models with varying dispersion), both as suitable alternatives to the classically used bivariate mixed Poisson regression models.

Contents

Chapter 1

Introduction

1.1 Multidimensionality in finance and insurance

Practitioners in finance and insurance are always juggling multiple sources of risk. For example in investments there is the need to make predictions on multiple stock prices on which one is simultaneously investing. Companies need many different supplies and have to sell their own goods, while the market price of these supplies and goods is in constant variation. Insurance companies have many different policyholders in their portfolios, each of them with its own level of riskiness, not to mention that insurance companies have multiple lines of business at the same time. In mathematics we call each individual source of risk ”a marginal risk” or ”a marginal”.

This need to take into account more than one source of risk is already very challenging even if we assume such risks to be independent. How do we represent the different sources of risk? How do we construct multivariate distributions to take into account more than one source of risk? Furthermore, these sources of risks are almost surely not independent. Therefore what if they are not independent? How do we relax the independence assumption and capture the existing correlation? This still remains an open question in statistics and many methods have been developed in the course of the years in the attempt to tackle this problem.

The linear correlation coefficient is the easiest method to structure the dependence and it would not be hard to find someone who still uses it as their sole measure of dependence between risky assets or exposures, despite its many significant limitations. In particular it assumes that the correlation can be well synthesised just with one statistical indicator and does not take into account that such correlation may vary depending on the percentile of the distribution of each risk factor (for instance, equities exhibit a greater tendency to crash together than to boom together) and also might not be linearly dependent on time.

Therefore the construction, study and applications of multivariate distributions - that can structure the joint distribution taking into account that the interaction between different risks may change in different points of the distribution and in different points in time - is one of the classical fields of research in statistics, and it continues to be an active one. In particular, it is important not limiting ourselves to multivariate Normal distributions, as it is now well demonstrated that normality does not work for asset returns, neither for the payouts from different options, neither for most risks in finance and insurance.

As summarised by Sarabia and Gómez Déniz (2008), in recent years several books containing theory about multivariate non-Normal distributions have been published: Hutchinson (1990), Joe (1997), Arnold et al. (1999), Kotz et al. (2004), Kotz and Nadarajah (2004), Nelsen (2007). In the discrete case specifically, the books of Kocherlakota and Kocherlakota (2004) and Johnson et al. (1997) and the review papers by Balakrishnan (2014) are to be noted. Reviews on constructions of discrete and continuous bivariate distributions are given by Lai (2004) and Lai (2006). It is though impossible producing a standard set of criteria that can always be applied to produce a unique distribution which could unequivocally be called the multivariate version (Kemp and Papageorgiou (1982)). In this sense, there is no satisfactory unified scheme of classi-

fying these methods. In the bivariate continuous case Lai (2004) and Lai (2006) have considered the following clusters of methods: marginal transformation method, methods of construction of copulae, mixing and compounding, variables in common and trivariate reduction techniques, conditionally specified distributions, marginal replacement, geometric approach, constructions of extreme-value models, limits of discrete distributions, some classical methods, distributions with a given variance-covariance matrix and transformations.
Specifically, in the course of this thesis, we are going to focus on three of the above cited methods, which allow to capture the dependence in a complete way considering the whole joint distribution function:

1. copulae (which we will be using in Chapter 2);
2. trivariate reduction scheme (which we will be using in Chapter 2);
3. mixtures (which we will be using in Chapters 3 and 4).

Such methods can be used to represent jointly as many sources of risk as needed, but in the course of this thesis, for the sake of simplicity, we will be exploring always bivariate cases: this choice was made to keep the problems as simple and clear as possible. If one wishes to describe more than two sources of risks, some re-adaptations of the models are possible in order to take into account in a more exhaustive way of the multiple dependencies that arise between different pairs of risks. Also, we only focus on positive correlation between our risks, but extensions of our models to take into account negative dependence are possible. We will be explaining more about the limitations and possible extensions of our models in the following chapters of this thesis. We will now start with an introduction to these models and then in the following chapters we will be exploring them in a variety of financial and actuarial applications.

1.2 Copulae

1.2.1 The importance of copulae

Copulae is a relatively recent method that allows flexible manipulation of risk factors and other relevant variables studied in finance. In particular, when working in multidimensional conditions (i.e. there are multiple risk factors), the assumption of a Normal multivariate distribution leads to conclusions that are distant from those found empirically. The research of new multivariate distributions that better adapt to the real behaviour of risk factors led copulae to be a useful tool for solving various problems in the financial fields.

The characteristic that best defines this instrument of financial analysis is the fact that it can separate the issues related to the univariate analysis of the margins from those that refer to the structure of dependence. To be more precise, a copula is used to separate the pure randomness of one variable (e.g. a financial asset) from the interdependencies between it and other variables. In this way, one can model each variable separately and, in addition, have a measure of the relations between those variables. Technically, this means that the univariate probability distribution, being informative on the probabilities of outcomes of one variable can be modelled by a distribution type of choice, while another variable can be modelled using another type of probability distribution. By doing so, one can choose for each and any asset in a spectrum the most appropriate type of distribution, not influencing the interdependencies between those variables. The interdependencies between those variables are represented by a multivariate probability distribution function, which is informative on the joint outcomes of the variables, and this multivariate distribution function is the copula as synthesised by Rachev et al. (2009).

1.2.2 An historical perspective

The history of copulae may be said to begin with Fréchet (1951). He investigated the following problem: given the distribution functions F_k with $k = 1, 2, ..., d$ of d random variables $X_1, X_2, ..., X_d$ defined on the same probability space $(\Omega, \mathcal{F}, P)$, what can be said about the set $\Gamma(F_1, F_2, ..., F_d)$ of the d-dimensional distribution function's whose marginals are the given F_k? It is immediate to note that the set $\Gamma(F_1, F_2, ..., F_d)$, now called the Fréchet class of F_k's, is not empty since, if $X_1, X_2, ..., X_d$ are independent, then the distribution function $(x_1, x_2, ..., x_d) \mapsto F(x_1, x_2, ..., x_d) = \prod_{k=1}^{d} F_k(x_k)$ always belongs to $\Gamma(F_1, F_2, ..., F_d)$. But it was not clear which other elements of $\Gamma(F_1, F_2, ..., F_d)$ were. Preliminary studies about this problem were conducted by Féron (1956), Gumbel (1958) and Fréchet (1956) himself, but the deepest result was obtained by Sklar (1959). He introduced the notion, and the name, of a copula, and proved the theorem that now bears his name.

Sklar's Theorem. If $F(x_1, x_2, ..., x_d)$ is a joint multivariate distribution function with univariate marginal distribution functions $F_1(x_1), F_2(x_2), ..., F_d(x_d)$, then there exists a copula C such that, for each $(x_1, x_2, ..., x_d) \in \Re$:

$$F(x_1, x_2, ..., x_d) = C(F_1(x_1), F_2(x_2), ..., F_d(x_d)).$$

The theorem also admits the following converse implication, usually very important when one wants to construct statistical models by considering, separately, the univariate behaviour of the components of a random vector and their dependence properties as captured by some copula. If C is a copula and $F_1(x_1), F_2(x_2), ..., F_d(x_d)$, are univariate distributions functions, then the function $F(x_1, x_2, ..., x_d)$ is the joint distribution function with margins $F_1(x_1), F_2(x_2), ..., F_d(x_d)$. The proof to Sklar's theorem was not given in Sklar (1959), it was instead provided later for the 2-dimension case.

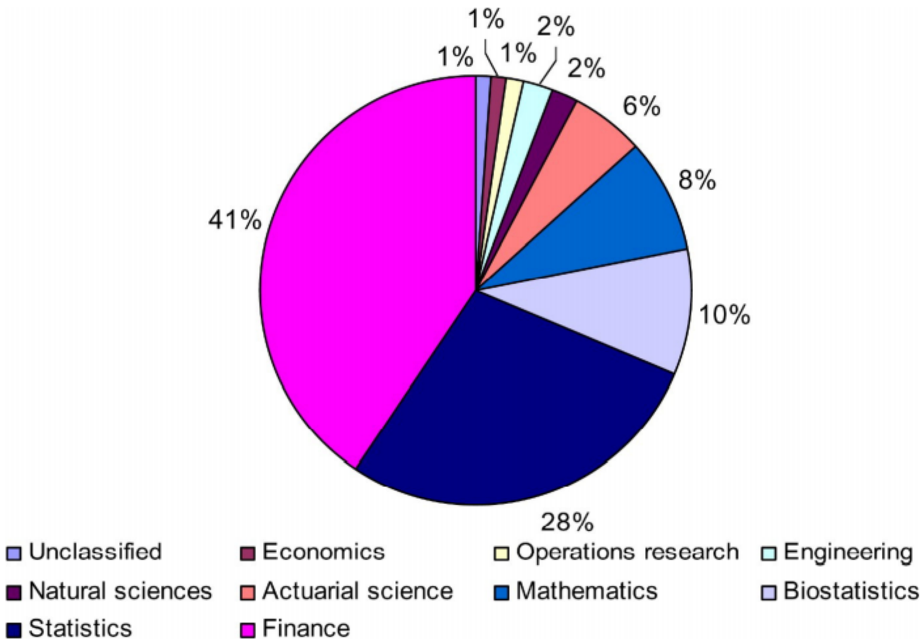
For about 15 years, all the results concerning copulae were obtained in the framework of the theory of Probabilistic Metric spaces (see Schweizer and Sklar (1983)). In the mid-seventies, Bert Schweizer realized that he could easily construct dependence measures by using copulae. However, for several other years, Chapter 6 of Schweizer and Sklar (1983), devoted to the theory of Probabilistic Metric spaces and published in 1983, was the main source of basic information on copulae. Nevertheless, since interest in questions of statistical dependence was increasing, others came to the subject from different directions. In 1990, Dall'Aglio organized the first conference devoted to copulae, called "Probability distributions with given marginals" (see Salinetti et al. (1991)). This led to a series of conferences that considerably helped the development of the field, since each of them offered the chance of presenting one's results and to learn those of other researchers.

At the end of the nineties, the notion of copulae became increasingly popular. Two books in particular became the standard references for the following decade: Joe (1997) published his work on multivariate models and in 1999 Nelsen published the first edition of his introduction to copulae (current edition Nelsen (2007)). But, the main reason of this increased interest has to be found in the discovery of the notion of copulae by researchers in several applied fields, like finance.

The advent of copulae in finance originated a wealth of investigations about them and their applications. At the same time, other fields discovered the importance of this concept for constructing more flexible multivariate models. Nowadays, it is near to impossible to give a complete account of all the applications of copulae to the main fields where they have been used. Genest et al. (2009) identified the most important areas of application of copulae. They examined 871 documents and grouped them into nine categories. They found that even though people in finance have been interested in copulae only since 2000, they produced the largest proportion of documents, i.e. 41%. Moreover, they discovered that two major phenomena account for the rise of copula modelling in finance: the lack of normality in log-returns and the dependence between extreme values of various assets. They checked 353 documents and grouped their contributions in finance into four categories:

- *Risk management*: topics included are those covering credit, market, operational risk and risk aggregation;
- *Portfolio management*: papers dealing with the dependence between international financial markets, different classes of assets and currencies;
- *Pricing of derivatives*: this category comprises work on the pricing of exotic options, collateralised debt obligations and credit default swaps;
- *Risk measurement*: papers discussing value-at-risk, expected shortfall and financial contagion.

Figure 1.1: Breakdown by discipline of 871 documents



1.2.3 Archimedean copulae

Copulae play an important role in the construction of multivariate distribution functions and, as a consequence, having at one's disposal a variety of copulae can be very useful for building stochastic models having different properties that are sometimes indispensable in practice (e.g., heavy tails, asymmetries, etc.). Therefore, several investigations have been carried out concerning the construction of different families of copulae and their properties (see Durante and Sempi (2010)). Generally, two greatest families of copulae can be distinguished: Elliptical and Archimedean family.

Elliptical copulae are the copulae with elliptical distributions, which have an elliptical form and therefore symmetry in the tails. Important copulae in this family are the Gaussian and the t-copula. The class of elliptical distributions provides useful examples of multivariate distributions because they share many of the tractable properties of the multivariate normal distribution. Furthermore, they allow to model multivariate extreme events and forms of non-normal dependencies. Simulation from elliptical distributions is easy to perform. Therefore, as a consequence of Sklar's Theorem, the simulation of elliptical copulae is also easy. However, they suffer from some drawbacks: elliptical copulae do not have closed form expressions and are restricted to have radial symmetry. In many finance and insurance applications it seems reasonable that there is a stronger dependence between big losses (e.g. a stock market crash) than between big gains. Such asymmetries cannot be modelled with elliptical copulae.

In our work, we will explore the family of Archimedean copulae. This class of copulae is worth studying since they find a wide range of applications for a number of reasons. Many interesting parametric families of copulae are Archimedean and the class of Archimedean copulae allows for a great variety of different dependence structures. Furthermore, in contrast to elliptical copulae, all commonly encountered Archimedean copulae have closed form expressions.

These copulae are not derived from multivariate distribution functions using Sklar's Theorem and a consequence of this is that we need somewhat technical conditions to assert that multivariate extensions of Archimedean 2-copulae are proper d-copulae.
A further disadvantage is that multivariate extensions of Archimedean copulae in general suffer from lack of free parameter choice in the sense that some of the entries in the resulting rank correlation matrix are forced to be equal.
As Mikosch (2006) argues, the copula C of a random vector X is a transformation of the distribution of X. As for the distribution of X we have infinitely many choices for C. If one chooses a copula it should be related to the problem at hand. For example, if we are interested in multivariate extremes the copula should be related to multivariate extreme value theory. In the literature various families of copula families with a name have been introduced. Their choice is not always based on reasoning but on mathematical convenience.
Each copula family is different for shape, behaviour and tail characteristics. These differences would allow us to fit empirical data to the optimal copula, meaning the copula that best reflects data behaviour, especially behaviour in the tails. In the literature, several methods to choose the "optimal" copula have been implemented. Generally, Maximum Likelihood or Inference Functions for Margins methods are used for the purpose. However, Maximum Likelihood Estimation is assumed to be a good method, since:

- it is sufficient: it gives complete information about parameters of interest;
- it is consistent: the true value of the parameters is recovered asymptotically for sufficiently large samples;
- it is efficient: it achieves asymptotically the lowest possible variance in parameter estimation.

Following McNeil et al. (2009) a copula C is called Archimedean if it can be written in the form

$$C(F_1(x_1), F_2(x_2), ..., F_d(x_d)) = \phi(\phi^{-1}(F_1(x_1)) + \phi^{-1}(F_2(x_2)) + ... + \phi^{-1}(F_d(x_d))),$$

for $(F_1(x_1), F_2(x_2), ..., F_d(x_d)) \in \mathcal{I}^d$ and for some generator function ϕ and its generalized inverse ϕ^{-1}.

A function $\phi : \Re^+ \to \mathcal{I}$ is said to be an (outer additive) generator if:

- it is continuous;
- it is decreasing and $\phi(0) = 1$, $\lim_{t\to+\infty} \phi(t) = 0$;
- it is strictly decreasing on $[0, t_0]$, where $t_0 := \inf\{t > 0 : \phi(t) = 0\}$.

If the function ϕ is invertible, or, equivalently, strictly decreasing on $\Re^+$, then the generator is said to be strict. If ϕ is strict, then $\phi(t) > 0$ for every $t > 0$ (and $\lim_{t\to+\infty} \phi(t) = 0$). The pseudo-inverse of ϕ is defined as follows:

$$\phi^{-1}(t) := \begin{cases} \phi^{-1}(t), & 0 \le t \le \phi(0) \\ 0, & \phi(0) \le t \le \infty. \end{cases}$$

A generator ϕ generates an Archimedean copula in dimension d if and only if it is d-monotone, that is:

- ϕ has continuous derivatives on $(0, \infty)$ up to order $d - 2$;
- $(-1)^k \phi^k(x) \ge 0$ for any $k = 1, ..., d - 2$;
- $(-1)^{d-2}\phi^{d-2}$ is non-negative, non-increasing and convex on $(0, \infty)$.

One-parameter copulae, constructed using a generator $\phi_\theta(t)$ and indexed by the parameter θ, are an important group of Archimedean copulae. In this type of copulae, the parameter θ is the measure of association, i.e. it defines the strenght of the correlation existing between the marginals: the higher the parameter θ, the higher the dependence. The existing domain of the parameter θ varies depending on the type of copula: when θ reaches its lower bound it means total independence between the marginals, whereas the upper bound means complete dependence.

Another measure of association that can be used to calculate the correlation between marginals is the rank correlation measure called Kendall's τ, which is another possible way of fitting copulae to data, because of its direct correspondence with the parameter θ, i.e knowing one you can easily extract the other one.

Let $(x_1, y_1), ..., (x_n, y_n)$ be a set of observations of the joint random variables X and Y, such that all the values of (x_i) and (y_i) are unique (ties are neglected for simplicity). Any pair of observations (x_i, y_i) and (x_j, y_j), where $i < j$, are said to be concordant if the sort order of (x_i, x_j) and (y_i, y_j) agrees: that is, if either both $x_i > x_j$ and $y_i > y_j$ holds or both $x_i < x_j$ and $y_i < y_j$; otherwise they are said to be discordant. The Kendall's τ coefficient is defined as:

$$\tau = \frac{(\text{number of concordant pairs}) - (\text{number of discordant pairs})}{n(n-1)/2}. \tag{1.1}$$

The denominator is the total number of pair combinations, so the coefficient must be in the range $-1 \leq \tau \leq 1$. If the agreement between the two rankings is perfect (i.e., the two rankings are the same) the coefficient has value 1. If the disagreement between the two rankings is perfect (i.e., one ranking is the reverse of the other) the coefficient has value -1. If X and Y are independent, then we would expect the coefficient to be approximately zero.

Three of the main one-parameter copulae will now be described. Then in Chapter 2 we are going to see an application of these three Archimedean copulae to the case where we need to capture the dependence between multiple Lévy processes: in such context these one-parameter Archimedean copulae will be called one-parameter Archimedean

Lévy copulae.
One limitation of one-parameter copulae (both in the regular and in the Lévy setting) is that they can describe only one measure of association: in our case, since we focus only on bivariate applications, one measure of dependence suffices to consider the dependence between two marginals. If there is the need to take into account of more than two, then it would be useful to use copulae that have more parameters of dependence and can therefore account for multiple different correlations between different pairs of marginals. In such cases Vine copulae (or Vine Lévy copulae in the Lévy setting) can overcome this limitation of one-parameter copulae.

Clayton copula

The Clayton copula has the following expression:

$$C(F_1(x_1), F_2(x_2), ..., F_d(x_d)) = \max\left[\left(\sum_{i=1}^{d}(F_i(x_i))^{-\theta} - (d-1)\right)^{1/\theta}, 0\right], \tag{1.2}$$

where $\theta \geq -1/(d-1)$. The limiting case $\theta = 0$ corresponds to the independence copula. The generator is:

$$\phi(t) = (\max(1 + \theta t, 0))^{1/\theta}.$$

The Kendall's τ is:

$$\tau = \frac{\theta}{\theta + 2}. \tag{1.3}$$

The Clayton copula shows an extremely uprising peak at $(0, 0)$ while a less pronounced behaviour at $(1, 1)$. Therefore, we can say that there is lower tail dependence, but no upper tail dependence, i.e.:
$\lambda_U = 0$
$\lambda_L = 2^{-1/\theta}$.

Gumbel copula

The Gumbel copula has the following expression:

$$C(F_1(x_1), F_2(x_2), ..., F_d(x_d)) = \exp\left[-\left[\sum_{i=1}^{d}(-\ln(F_i(x_i)))^{\theta}\right]^{1/\theta}\right], \tag{1.4}$$

where $\theta \geq 1$. For $\theta = 1$ we obtain the independence copula as a special case, and $\lim_{\theta \to +\infty} C(F_1(x_1), F_2(x_2), ..., F_d(x_d))$ is the co-monotonicity copula. The generator is:

$$\phi(t) = \exp(-t^{1/\theta}).$$

The Kendall's τ is:

$$\tau = 1 - \theta^{-1}. \tag{1.5}$$

The Gumbel copula shows an extremely uprising peak at $(1, 1)$ while a less pronounced behaviour at $(0, 0)$. Therefore, we can say that there is upper tail dependence, but no lower tail dependence, i.e.:
$\lambda_U = 2 - 2^{1/\theta}$
$\lambda_L = 0$.

Frank copula

The Frank copula has the following expression:

$$C(F_1(x_1), F_2(x_2), ..., F_d(x_d)) = -\frac{1}{\theta}\ln\left[1 + \frac{\prod_{i=1}^{d}(\exp(-\theta F_i(x_i)) - 1)}{(\exp(-\theta) - 1)^{d-1}}\right], \tag{1.6}$$

where $\theta > 0$. The limiting case $\theta = 0$ corresponds to the independence copula. The generator is:

$$\phi(t) = -\frac{1}{\theta}\log(1 - (1 - \exp(-\theta))\exp(-t)).$$

The Frank copula shows no extremely uprising peaks. Therefore, we can say that there is no tail dependence, i.e.:

$\lambda_U = 0$

$\lambda_L = 0.$

1.3 Trivariate reduction scheme

Copulae have a high degree of precision, which is however not needed for all applications and comes at the price of reduced analytical tractability. For this reason, several authors have proposed alternative approaches to generate dependency, which may be less flexible, but lead to simpler models. One of these methods is the trivariate reduction scheme, also known as "the variables in common method", which is a popular and old technique used for building dependent variables, both in continuous and discrete cases. This method cannot capture extreme values as well as copulae, but the estimation and simulation procedures are much more straightforward than with copulae. As previously mentioned, in the following we will be focusing mainly on the bivariate case.

The method consists of building a pair of dependent random variables starting from three (or more) random variables. These initial random variables are usually independent. The functions that connect initial variables are generally elementary functions, or are given by the structure of the variables that we want to generate.

A broad definition can be

$$\begin{cases} X = \upsilon_1(e_X, c_{XY}), \\ Y = \upsilon_2(e_Y, \tilde{c}_{XY}). \end{cases} \tag{1.7}$$

where e_X, e_Y represent two sets containing the specific variables of X and Y respectively, and c_{XY}, $\tilde{c}_{XY}$ sets containing the common or latent variables. According to Marshall and Olkin (2007), many of the couples (X, Y) here presented are associated and

$$\text{Cov}\,(u(X), v(X)) \geq 0, \tag{1.8}$$

for all increasing functions u, v for which the covariance exists, and then only positive correlations are possible. This method has been used mainly in the bivariate setting, which as we mentioned is the dimension in which we are also interested, therefore we wrote the formula for the case of two marginals. Nevertheless this representation can be used for any dimension d of marginal risks.

Over the last few years, several new dependent distributions using this method have been proposed. Some relevant models are:

- the bivariate Generalized Poisson distribution (where if the variables are Poisson random variables, we obtain the classical bivariate Poisson distribution, which is often used for obtaining compound bivariate Poisson distributions, whereas if we consider a Generalized Poisson distribution for the random variables, we obtain the model considered by Vernic (1997) and Vernic (2000));

- the bivariate Beta distribution (proposed by Olkin and Liu (2003), while Sarabia and Castillo (2006) have considered a generalization of the joint probability density function under a conditional specification);

- the bivariate t distribution (where the usual bivariate spherically symmetric distribution is defined as by Fang et al. (2018), a bivariate distribution with Student t marginals with different degrees of freedom has been proposed by Jones (2002) and an alternative bivariate t distribution including the independence case has been presented by Shaw and Lee (2008));

- the bivariate Marshall-Olkin type distributions (if the components correspond to Exponential distributions, we obtain the Marshall-Olkin distribution, see Marshall and Olkin (1967). Other survival models have been considered by Sarhan and Balakrishnan (2007) with the Exponentiated Exponential distribution, as well as a mixture of the proposed bivariate distribution. Arnold and Brockett (1983) have obtained a bivariate Gompertz-Makeham distribution using a similar construction);

- the bivariate F distribution (the classical bivariate F distribution is defined as by Balakrishnan (2000), while for a bivariate F distribution with arbitrary degrees of freedom see El-Bassiouny and Jones (2009));

- the Ballotta-Bonfiglioli model (constructed for the case when the random variables follow Lévy processes, see Ballotta and Bonfiglioli (2016)).

In Chapter 2 we are going to use this last model, which will therefore be described in such chapter. All the models presented in this section, including the Ballotta-Bonfiglioli model, can be extended to higher dimensions.

1.4 Mixtures

The use of mixtures to obtain flexible families of densities has a long history, especially in the univariate case. The advantages of the mixtures mechanism are diverse. The new classes of distributions obtained by mixing are more flexible than the original, overdispersed with tails larger than the original distribution and often providing better fits. The extension of a mixture to the multivariate case is usually simple, and the marginal distributions belong to the same family.

The limitation of this method is the same we had with the trivariate reduction scheme, i.e. that it cannot capture extreme events as well as copulae, but on the other hand, simulation and Bayesian estimation of mixtures are quite direct. Also mixtures solve the identifiability issues that can arise with copulae (more about identifiability will follow in Chapters 3 and 4). Since the introduction of simulation-based methods for inference (particularly the Gibbs sampler in a Bayesian framework), complicated densities such as those having mixture representation have been satisfactorily handled.

The multivariate distributions can be both discrete and continuous, but we will be focusing only on the discrete ones. A broad definition can be

$$P\left(x_1, x_2, ..., x_d\right) = \int_0^\infty \prod_{k=1}^{d} P\left(x_k|\lambda\right) f\left(\lambda; \boldsymbol{\theta}\right) d\lambda, \tag{1.9}$$

where x_k for $k = 1, ...d$ are the marginal risks and λ is the common mixing factor responsible for the correlation between the marginals, which follows a mixing distribution $f\left(\cdot\right)$ with vector of parameters $\boldsymbol{\theta}$.

In the case of multivariate discrete distributions, the study of the variability of multivariate counts arises in many practical situations. In ecology the counts may be the different species of animals in different geographical areas whilst in insurance, the number of claims of different policyholders in the portfolio. Some relevant mixtures are among the class of multivariate mixed Poisson models, such as:

- the multivariate Poisson-Lognormal distribution (see Aitchison and Ho (1989), other versions of this model can be viewed in Tonda et al. (2005));
- the multivariate Poisson-Generalized Inverse Gaussian distribution (departing from the Sichel distribution, also called Poisson-Generalized Inverse Gaussian distribu-

tion, in Sichel (1971) were investigated bivariate extensions of that distribution. In Stein et al. (1987) one of them is studied in order to obtain the estimation of the parameters via the likelihood method);

- the multivariate Poisson-Beta distribution (Sarabia and Gómez-Déniz (2011) have proposed multivariate versions of the Beta mixture of Poisson distribution considered by Gurland (1957) and Katti (1966). The new class of distributions can be used for modelling multivariate dependent count data when marginal overdispersion is also observed. By choosing the Sarmanov-Lee distribution described in Sarmanov (1966), Ting Lee (1996) and Kotz et al. (2004) and which has been used by Sarabia and Castillo (2006), Sarabia and Gómez-Déniz (2011) built a bivariate distribution that admits non-limited correlations of any sign).

Of course there is no need to restrict ourselves to the class of multivariate mixed Poisson models. Another relevant multivariate discrete distribution is in fact the multivariate Negative Binomial-Inverse Gaussian distribution, which was considered by Gómez-Déniz et al. (2008) as a mixture of a Negative Binomial distribution with an Inverse Gaussian distribution. This model is tractable and with attractive properties, which makes it suitable for application in disciplines where overdispersion is observed.

In Chapters 3 and 4 we will be focusing on the multivariate Negative Binomial-Inverse Gaussian model and on the class of multivariate mixed Poisson models (as previously mentioned, in our case we use $d = 2$). All these models only have one mixing distribution, but of course it is not necessary to limit oneself to only one: mixtures are a flexible approach that allows to include as many mixing distributions as needed.

1.5 Outline of the thesis

1.5.1 Chapter 2 - option pricing

Chapter 2 is in financial mathematics, more specifically in option pricing. Such work focuses on the pricing of American multi-asset options, which are options that can be exercised at any time between the date of purchase and the expiration date and whose payoff depends on the overall performance of more than one underlying asset. We use Lévy processes to represent such underlying assets, which can be described as processes that can consider a jump component.

In the literature, American multi-asset options under the Lévy processes have generally been priced under the assumption of linear dependence between the underlying assets, which was therefore synthesised by the use of the correlation coefficient ρ (which is in fact the measure of linear correlation between variables). Unfortunately linear dependence is unable to represent well the complexity of the markets and, even more critically, the jump components are implicitly considered to be independent. Since the jumps are the most important component in such processes, by assuming them to be independent, the entire dependence structure between the processes loses validity.

Therefore we resort to two of the aforementioned models to capture the dependence between marginal Lévy processes in more complete and realistic fashions:

1. copulae (as presented in Section 1.2, which in the context of Lévy processes are called Lévy copulae)

2. the trivariate reduction scheme in which the marginals are given by the sum of an idiosyncratic part plus a common part (as presented in Section 1.3, which in the context of Lévy processes is called the Ballotta-Bonfiglioli model).

Both models have been previously used only for the pricing of European multi-asset options and we have extended the literature to the American setting. We develop our

results particularly for the case of two marginal Variance Gamma processes to price eight different types of options, but such procedure may be implemented for any kind of Lévy processes and of multi-asset options. Also, under the Lévy copulae procedure, we have restricted ourselves to four types of Lévy copulae, but of course it is possible to extend the framework to other copulae.
In fact one of our limitations is that the suggested copulae can only account for positive dependence between the variables: it would be interesting to explore the procedure under copulae that can also account for negative dependence. Furthermore we use one parameter copulae, which can only describe one measure of dependence between variables, which is enough if we only have two underlying assets. It would be useful to readjust the procedure on Vine Lévy copulae, so that in the case of more than two underlying assets one is not limited by one overall dependence, but also pairwise dependence can be captured.

Our contributions to the literature are mainly the following two:

1. Regarding Lévy copulae, the literature has only covered European multi-asset options, therefore we extend the existing work to American options. The literature as well as our work is done via simulation, therefore the American feature is managed in practice through the employment of the Longstaff-Schwartz regression model.

2. Regarding the Ballotta-Bonfiglioli model, work has been done in the European setting, which we extend to the American setting. We achieve the pricing via generator (i.e. by solving the partial integro-differential equation), using a procedure via finite difference method which had been developed for the univariate setting and we extend it to take into account more underlying assets, whose dependence is captured through the aforementioned model. We then also develop the pricing via simulation, applying again the Longstaff-Schwartz algorithm, and we obtain

comparable results to the ones obtained by solving the partial integro-differential equation.

Such chapter is authored solely by me but under the invaluable guidance of Dr Tobias Kley.

1.5.2 Chapters 3 and 4 - pricing premia in the automobile insurance

Methodology

Chapters 3 and 4 are in actuarial mathematics, more specifically in automobile insurance ratemaking when there is the need to consider more than one type of claims (in our case, as mentioned, we focus on the bivariate case, therefore two types of claims), where - again - a complex dependence structure between such types of claims needs to be represented.

We structure such dependence by using mixtures (as per Section 1.4): we set the responses (i.e. the number of claims per type) to follow marginal Negative Binomial distributions/regression models (in Chapter 3) and marginal Poisson distributions/regression models (in Chapter 4) and we set a common heterogeneity factor as the mixing variable, and therefore the distributions/regression models followed by the two responses are mixtured by the same common mixing distribution/regression model. The mixing component in such two chapters is structured to be either an Inverse Gaussian, a Gamma or a Lognormal distribution/regression model. Therefore, all our models are ideally suited for capturing overdispersion and positive dependencies in the two-dimensional count data setting which, as all recent studies suggest, is the norm when the ratemaking consists of pricing different types of claim counts arising from the same policy.

It is to be noted that our models cannot capture underdispersion and negative dependencies, therefore if dealing with a data set which exhibits these characteristics, other

models are more advisable: in Chapter 3 for example we refer the reader to the models by Fung et al. (2019a) and Fung et al. (2019b) for such scenarios.
In the automobile insurance in order to distinguish between individuals with different risks profiles, Bonus-Malus Systems (BMSs) are applied as experience rating mechanisms. These mechanisms are also called a posteriori ratemaking as they take into account the number of years that the policyholder has been in the insurance's portfolio and of claims reported by each individual during the previous years and apply a penalty or a reward accordingly. In order to obtain the most accurate BMSs, Bayesian statistics is employed as it grants the optimal and fairer premia estimates. Before applying a posteriori ratemaking, the premia are already presorted through the a priori ratemaking, which uses all the available information for the policyholder and its automobile as covariate information to predetermine the level of riskiness to be expected by each individual.

Importance of the work

This work is relevant because, firstly, accurate pricing of automobile insurance has become crucial in such context as the premia for automobile insurance have been increasing alarmingly over the past years, with prospects for the future that do not seem any brighter and show the trend continuing to rise distressingly in the next years. In fact in the last report issued in the UK, the Association of British Insurers (ABI) states that average premia have risen by 9% between October and December 2017 and display an average of £481 which is the highest figure since the ABI started records in 2012, and the eighth successive quarter in which premia have risen. These increments are partially caused by a rise in the Insurance Premium Tax, which has reached 12% in June 2017, and by a decrease in April 2017 of the discount (or Odgen) rate that is applied on insurances for the cost of long-term injury claims.
Such worrying scenario is not for Europe only: in the US in fact, due to the additional

factor of catastrophic disasters (i.e. floods, storms and hurricanes), the problem is even more significant. According to the National Highway Traffic and Safety Administration, people's distraction has caused crashes to rise to numbers that have not been seen in nearly a decade. This statement is strengthened by the National Safety Council, which in fact observes an increase of 6% in fatal motor accidents from 2015 to 2016, for a total of 40,200 fatalities. This combination of natural disasters and accidents has caused a huge increment in the number of claims, which justifies these high premia.

There have been attempts to reduce the premia in order to relieve the customers, but in the UK, despite the average price having recorded its first quarterly fall in two years, average costs during the first quarter of 2018 were the highest that the ABI has ever seen at this time of year. The Ministry of Justice has anticipated a revision of the Odgen rate and new legislation that should relieve the current pressure on premia but it is unclear whether these measures will secure much of an impact.

Furthermore, the calculation of the premium differs between different countries as varying policies may be applied, as reported by the RAC Foundation: firsty, obviously the premium is calibrated on the living costs, therefore for example in Europe it is higher for northern countries; secondly, it is important using models which allow employing the a priori ratemaking, as different covariate information may be considered since each country may weigh differently specific explanatory variables. In fact for example France, Germany and Sweden focus primarily on the type of vehicle, while the UK has a strongly risk-based approach to underwriting that focuses on the driver. Therefore, despite total average premia being comparable in the aforementioned countries, they are lower for young drivers in other European countries than is the case in the UK, and conversely higher for older drivers.

In addition, when more than one type of claim is taken into account, the pricing policies between countries differ: being able to consider more than one type of claim is crucial as not all countries reason only in terms of comprehensive claims, but actually

distinguish between different types of claims. For example, in many European markets, such as France, Germany, Italy, the Netherlands, Spain and Sweden, the cover is split into third party, fire and theft on the one hand, and material damage on the other, with a lower price being applied when only the first one is chosen. On the contrary, in the UK, comprehensive cover prevails due to its cheaper price than third-party cover, since requesting for the latter is acknowledged as a declaration of being a risky profile, especially for less-experienced drivers.
Hence, it has become crucial for actuaries to tailor their models in concurrence with the need of absorbing covariate information and of considering more then one type of claim, leading to the consequent problem of depicting appropriately the correlation structure between the different kinds of claims and designing an optimal allocation of the expected costs. The current pressure on the premia has highlighted the urgency of conceiving more sophisticated models with the ability of capturing all factors affecting the premia, which could therefore be led to a more accurate and fairer calculation.

Contribution to the literature

In Chapter 3 we extend to the bivariate case a Negative Binomial-Inverse Gaussian regression model (as mentioned in Section 1.4) which had already been considered in the context of insurance ratemaking in its univariate version. We then compare it to the more traditionally used bivariate mixed Poisson models and show that, not only our model is a better fit (in terms of specification criteria) for the data set of a MTPL insurance, but also has the following beneficial characteristics:

1. since the marginal responses/type of claims are distributed as two Negative Binomial distributions/regression models, overdispersion can be taken into account also when considering separately the two marginals, and not only when considered jointly. This is actually an important aspect, because, as we will see in Chapters 3 and 4, our data set clearly shows that such overdispersion already exists when

analysing separately the two types of claims;

2. we develop a novel Expectation-Maximisation type algorithm for maximum likelihood estimation of the bivariate Negative Binomial-Inverse Gaussian model, which allows to achieve parameter estimation also including covariate information (i.e. a priori ratemaking), which as we said is necessary if we want a model flexible and versatile enough to be adaptable to any kind of individual (and its level of riskiness) and to any country/market;

3. our model shows much less extreme a posteriori, or Bonus-Malus, premia for policyholders with some claims experience than those produced by the mixed Poisson models, which is a significant asset, if we consider how high the premia already are;

4. also, for a given total number of claims, our model can enable the actuary to differentiate the premium rates based on the exact frequencies of the two types of claims, whereas the two mixed Poisson models do not allow to price discriminate by taking into account the difference in the numbers of the two types of claims.

In Chapter 4 we use the family of multivariate mixed Poisson models (as mentioned in Section 1.4), we extend them and introduce their adjusted version as a new class of models, the multivariate mixed Poisson models with varying dispersion, which allow also the mixing component to be modelled as a generalized linear model (GLM). The main advantages of our approach are the following:

1. as mentioned, we express all parameters of our model in terms of covariate information, including the mixing variable, whereas traditionally the mixing component cannot be represented as a regression model, but only as a distribution. This characteristic allows our model to take better into account all the available information, granting therefore the fairest premium rates possible which are accurately individually-tailored to each policyholder;

2. we develop an atypical Expectation-Maximisation algorithm that allows to include explanatory variables everywhere and that is easily re-adaptable to different parametric families from the ones suggested;

3. we implement the a posteriori ratemaking not only based on the expected principle, but we suggest the variance principle as a suitable alternative. In fact it needs to be noted that the majority of authors rely on the expected value principle, while the variance principle was suggested in the construction of BMSs with a frequency component reliant exclusively on the a posteriori criteria. Nevertheless, the latter principle, when BMSs are used, is much more robust than the expected value principle.

Chapters 3 and 4 are joint work with Dr George Tzougas, and my contribution to such chapters is as important as his.

Chapter 2

American multi-asset option pricing under Lévy copulae and the Ballotta-Bonfiglioli model

2.1 Introduction and outline

This chapter focuses on the pricing of American multi-asset options, which are options that can be exercised at any point in time (American) till the expiry date (if there is one) and that depend on more than one underlying asset (multi-asset). We consider the multiple underlying assets to be represented by Lévy processes, which can be described as processes that can consider a jump component: for further details on Lévy processes see Bertoin (1996), Sato (1999) and Kyprianou (2014) for the general setting and Schoutens (2003) and Tankov (2003b) for financial applications.

We develop our results particularly for the case of two underlying assets which follow two marginal Variance Gamma processes: full details on the Variance Gamma process can be found in Madan and Seneta (1990), Madan et al. (1998) and Seneta (2007), while its fitting is presented by Seneta (2004). As presented in Chapter 1, we need to take into account that the marginal underlying assets are in most cases not independent and therefore we need to model in the most accurate and complete way possible the dependence structure existing between them. In this chapter we focus on two particular models that have been developed to capture the dependence between the marginal processes: copulae (Section 1.2) and the trivariate reduction scheme (Section 1.3). These two models have been presented in a more general setting in Chapter 1, but have been properly adjusted in the existing literature for the case where the marginals follow Lévy processes.

In particular copulae are in the case of Lévy processes called Lévy copulae. Lévy copulae were introduced on $\Re^d_+$ by Tankov (2003a) and then extended to $\Re^d$ by Kallsen and Tankov (2006) and Tankov (2016). Other relevant literature is Barndorff-Nielsen and Lindner (2007) and Bäuerle et al. (2008). The estimation procedure has been discussed by Laeven (2009), Bücher et al. (2013), Grothe (2013) and Palmes et al. (2018) for the nonparametric cases and by Esmaeili and Klüppelberg (2011) and Esmaeili and Klüppelberg (2013) for the parametric cases. Lévy copulae, thanks to their degree of accuracy and their ability to calculate extreme events (see Bollerslev et al. (2013) and Grothe (2013)) have been used for many applications in finance and insurance. In particular Tankov (2004), Kettler (2006), Tankov (2006) and Linders and Schoutens (2014) have been using Lévy copulae for the representation of correlated financial processes/underlying assets for the purpose of option pricing. Applications to insurance can be found in Avanzi et al. (2011) and Bäuerle and Blatter (2011) for correlated claims, portfolios and business lines. See Böcker and Klüppelberg (2008), Böcker and Klüppelberg (2010) and Van Velsen (2012) for applications to operational risk.

The trivariate reduction scheme is called the Ballotta-Bonfiglioli model in the case where each marginal follows a Lévy process. This method was presented by Ballotta and Bonfiglioli (2016), and then the estimation procedure was developed by Ballotta et al. (2015b) and Ballotta et al. (2019a). It has been applied abundantly in financial appli-

cations, such as for the joint dynamics of FX rates and asset prices for the pricing of Quanto products in Ballotta et al. (2015a) and Ballotta et al. (2017) and then extended to a Markov-modulated switching regime by Deelstra and Simon (2017). Ballotta and Fusai (2015) and Ballotta et al. (2019b) have then used it for counterparty credit risk.
Of course there are other efficient methods to capture the correlation between the two Variance Gamma processes, such as the model suggested by Luciano and Schoutens (2006) and Linders and Stassen (2016), which represents the two Variance Gamma processes as time-changed geometric Brownian motions with a common Gamma subordinator.

The reason for which we focus on the aforementioned two models to capture the correlation between the two underlying assets is that Lévy copulae can capture extreme events better than any other model and have a degree of accuracy which still remains unparalleled, but this comes at the cost of reduced analytical tractability and much less convenient estimation and simulation procedure. The Ballotta-Bonfiglioli method can overcome these flaws, even though we lose on the level of precision. So these two models can be seen as complementary to one another.

Most importantly, as mentioned in Chapter 1, the linear correlation coefficient is still commonly used to capture the dependence between the underlying assets, despite its inability to represent the complexity of the markets and, even more critically, if we assume that the two marginal underlying assets follow two Lévy processes, it presents an unrealistic assumption of independence between the two marginal jump components (see, for example, Jinghui (2009)). Since the jumps are the most important components in such processes, by assuming them to be independent, the entire dependence structure between the processes loses validity. The two proposed methods on the other hand are perfectly able to overcome this issue.

An important limitation of our approach is that we only take into account positive correlation, but re-adaptations of our two models to take into account also negative dependence are possible.

Our contributions to the literature are mainly the following:

1. Under the framework given by Lévy copulae, the literature has only covered the pricing of European multi-asset options via simulation, therefore we extend the existing work to the pricing of American options. To price this type of options via simulation, we need to implement the Longstaff-Schwartz regression model (see Longstaff and Schwartz (2001)), which can create deficiencies such as a significant increase in the computational time when dealing with many underlying assets (see Hanbali and Linders (2019)), but for our bivariate case it has a good degree of precision and does not create any computational burden.

2. Regarding the Ballotta-Bonfiglioli model, work has been done to price European contracts via simulation, we implement again the Longstaff-Schwartz regression model to extend the literature to American contracts.

3. Hirsa and Madan (2004), Fiorani (1999) and Fiorani (2004) had priced American mono-asset options by solving the partial integro-differential equation (PIDE) via finite difference method for the case of one marginal underlying asset following a Variance Gamma process. We extend their work to the multidimensional case, where we have more underlying assets following marginal Variance Gamma processes, whose dependence is captured through the Ballotta-Bonfiglioli model. We then compare the achieved results to the ones we obtained via simulation and we see that they are perfectly comparable.

The chapter is structured as follows:

in Subsection 2.2.1 we introduce Variance Gamma processes as the marginal underlying assets, in Subsection 2.2.2 we show how to apply them for the pricing of options. In Section 2.3 we achieve the pricing of American multi-asset options under the Lévy copulae. In order to do so, the section is arranged as follows: in Subsection 2.3.1 we introduce Lévy copulae, followed by their simulation procedure in Subsection 2.3.2. We then introduce the employment of the Longstaff-Schwartz algorithm in Subsection 2.3.3, which needs to be applied on the simulated pahts for the evaluation of the continuation region and of the option price. Finally we present the option prices under the Clayton Lévy bidirectional copula in Subsection 2.3.4, under the Clayton Lévy (unidirectional) copula in Subsection 2.3.5, under the Gumbel Lévy copula in Subsection 2.3.6 and under the Frank Lévy copula in Subsection 2.3.7 for different values of the copulae's parameters that govern the dependence structure. In Section 2.4 we concentrate on the pricing of the same options, but by using the Ballotta-Bonfiglioli method to capture the dependence structure: therefore in Subsection 2.4.1 we present such method and in Subsection 2.4.2 we present our extension in pricing options by solving the PIDE developed by Hirsa and Madan (2004), Fiorani (1999) and by Fiorani (2004) to the multidimensional setting under the Ballotta-Bonfiglioli structure. Then in Subsection 2.4.3 we present some numerical results and in Subsection 2.4.4 we compare them with the results obtained via simulation. We end by drawing our conclusions in Section 2.5.

2.2 Variance Gamma processes for option pricing

2.2.1 The Variance Gamma process

Given the probability space $(\Omega, \mathcal{F}, P)$, let $x_k(t; \sigma_k, \alpha_k, \theta_k)$ for $k = 1, 2$ be our two marginal Variance Gamma processes with time t and marginal parameters σ_k, α_k and θ_k. Variance Gamma processes, as all processes belonging to the Lévy family, need to be representable by the triplet of Lévy characteristics (b_k, c_k, ν_k) given by the Lévy-Khintchine formula, such that:

$$\mathbb{E}\left[e^{iux_k}\right] = \exp\left[ib_k u - \frac{u^2 c_k}{2} + \int_{\mathbb{R}} \left(e^{iux_k} - 1 - iux_k \mathbb{1}_{\{|x_k|<1\}}\right)\nu_k(dx_k)\right]. \tag{2.1}$$

In the marginal Lévy triplet (b_k, c_k, ν_k) for each marginal Lévy process we have:

1. drift term $b_k \in \Re$
2. diffusion coefficient $c_k \in \Re_{\geq 0}$
3. Lévy measure ν_k with $\nu_k(\{0\}) = 0$

$$\int_{\Re} (1 \wedge |x_k|^2)\nu_k(dx_k) < \infty.$$

According to Madan and Seneta (1990), Madan et al. (1998), Seneta (2004) and Seneta (2007), the Variance Gamma processes $x_k(t; \sigma_k, \alpha_k, \theta_k)$, for $k = 1, 2$ in our two-dimensional case, are defined in terms of Brownian motions $B_k(t; \theta_k, \sigma_k)$ with drift θ_k and independent Gamma processes $\gamma_k(t; 1, \alpha_k)$ for the random time with unit mean rate, as

$$x_k(t; \sigma_k, \alpha_k, \theta_k) = B_k(\gamma_k(t; 1, \alpha_k); \theta_k, \sigma_k). \tag{2.2}$$

Variance Gamma processes are obtained by evaluating the Brownian motion at a time given by the Gamma process and are controlled by three parameters:

1. σ_k the volatility of the Brownian motion;
2. α_k the variance rate of the gamma time change;
3. θ_k the drift in the Brownian motion with drift.

The processes therefore provide two dimensions of control on the volatility, whereas over the skew control is attained through θ_k and over kurtosis through α_k. Variance Gamma processes have different representations and in one of them they can be expressed as the difference of two independent increasing Gamma processes, specifically

$$x_k(t; \sigma_k, \alpha_k, \theta_k) = \gamma_{p,k}(t; \mu_{p,k}, \alpha_{p,k}) - \gamma_{n,k}(t; \mu_{n,k}, \alpha_{n,k}),$$

where

$$\mu_{p,k} = \frac{1}{2}\sqrt{\theta_k^2 + \frac{2\sigma_k^2}{\alpha_k}} + \frac{\theta_k}{2}$$

$$\mu_{n,k} = \frac{1}{2}\sqrt{\theta_k^2 + \frac{2\sigma_k^2}{\alpha_k}} - \frac{\theta_k}{2}$$

$$\alpha_{p,k} = \left(\frac{1}{2}\sqrt{\theta_k^2 + \frac{2\sigma_k^2}{\alpha_k}} + \frac{\theta_k}{2}\right)^2 \alpha_k$$

$$\alpha_{n,k} = \left(\frac{1}{2}\sqrt{\theta_k^2 + \frac{2\sigma_k^2}{\alpha_k}} - \frac{\theta_k}{2}\right)^2 \alpha_k$$

and the Lévy measure would consequently be:

$$\nu_k(x_k)dx_k = \begin{cases} \mu_{n,k}^2/\alpha_{n,k}^2\left[\exp\left(-\frac{\mu_{n,k}}{\alpha_{n,k}}|x_k|\right)/|x_k|\right]dx_k, & \text{for } x_k < 0 \\ \mu_{p,k}^2/\alpha_{p,k}^2\left[\exp\left(-\frac{\mu_{p,k}}{\alpha_{p,k}}x_k\right)/x_k\right]dx_k, & \text{for } x_k > 0. \end{cases}$$

2.2.2 From Variance Gamma processes to multi-asset option pricing

We now need to apply a transformation to our marginal Variance Gamma processes, as in fact the dynamics of the stock prices are described by a transformed version of our marginal processes. The new specification for the statistical stock prices dynamics is obtained by replacing the role of the Brownian motion in the original Black-Scholes geometric Brownian motion model by the Variance Gamma process. In the Variance Gamma setting the statistical process for the stock prices is given by

$$S_k(t) = S_k(0)\exp((m + q + \omega_k)t + x_k(t;\sigma_k,\alpha_k,\theta_k)), \text{ for } k = 1, 2 \tag{2.3}$$

where $x_k(t;\sigma_k,\alpha_k,\theta_k)$ for $k=1,2$ are our marginal Variance Gamma processes, with

$$\omega_k = \frac{1}{\alpha_k}\ln\left(1-\theta_k\alpha_k-\frac{\sigma_k^2\alpha_k}{2}\right),$$

and m is the mean rate of return on the stocks under the statistical probability measure, q is the dividend and ω_k is necessary to ensure a martingale property.
This exponential transformation is needed for the marginal Variance Gamma processes to be able to behave like stock prices. Till this point we are using statistical/physical processes on the probability space $(\Omega,\mathcal{F},P)$, with corresponding marginal statistical/ physical parameters $\sigma_k,\alpha_k,\theta_k$. If we want to use the marginal Variance Gamma processes for the purpose of option pricing, we need to pass from the statistical/physical probability space $(\Omega,\mathcal{F},P)$ to an equivalent risk neutral one $(\Omega,\mathcal{F},Q)$ as per Girsanov's theorem. Under the risk neutral probability measure, the discounted stock prices are martingales and it follows that the mean rate of return on the stocks under this probability measure is the continuously compounded interest rate r. Let the risk neutral processes be given by

$$S_{RN,k}(t) = S_k(0)\exp((r+q+\omega_{RN,k})t + x_k(t;\sigma_{RN,k},\alpha_{RN,k},\theta_{RN,k})), \tag{2.4}$$

where the subscript RN on the Variance Gamma parameters indicates that these are the risk neutral parameters, and

$$\omega_{RN,k} = \frac{1}{\alpha_{RN,k}}\ln\left(1-\theta_{RN,k}\alpha_{RN,k}-\frac{\sigma_{RN,k}^2\alpha_{RN,k}}{2}\right).$$

The densities of the log stock prices relative over an interval of length t are, conditional on the realizations of the Gamma time changes, normal density functions. The unconditional densities are obtained by integrating out the Gamma variate and the results are in terms of the modified Bessel functions of the second kind. There are multiple methods to pass from $(\Omega, \mathcal{F}, P)$ to $(\Omega, \mathcal{F}, Q)$: the method we use to obtain an equivalent martingale measure Q is by mean correcting the exponentials of our marginal Lévy processes. In this case the risk-neutral process is given by

$$\begin{aligned} S_{RN,k}(t) &= S_k(0)\exp((m+q+\omega_k)t + x_k(t;\sigma_k,\alpha_k,\theta_k)) \\ &\cdot \frac{\exp(rt)}{\mathbb{E}[\exp((m+q+\omega_k)t + x_k(t;\sigma_k,\alpha_k,\theta_k))]}. \end{aligned} \tag{2.5}$$

Instead of using the marginal stock prices $S_{RN,k}(t)$ for $k = 1, 2$ under the risk-neutral measure themselves though, we work on the log-returns of the stock prices, rather than the prices themselves. The problem can therefore be transformed in logarithmic terms with the following change of variable:

$$y_{RN,k}(t) \triangleq \ln(S_{RN,k}(t)) \text{ for } k = 1, 2,$$

The log-returns can now be inserted in the payoff of the multi-asset option. It can be any kind of multi-asset option, but in our numerical application we will price four types of multi-asset options, both in their call and put versions, which have a life that goes for time 0 (today) to time T, therefore $t \in [0, T]$. Our considered multi-asset options are:

1. an American equally weighted basket option, with payoff

$$W(y_{RN,1}(t), y_{RN,2}(t), t) = \sup_{0\le t\le T} \exp(rt)\left(\sum_{k=1}^{2} w_k \exp(y_{RN,k}(t)) - K\right)^{+} \text{ with } w_k = \frac{1}{2}, \tag{2.6}$$

in its call version, and

$$W(y_{RN,1}(t), y_{RN,2}(t), t) = \sup_{0 \leq t \leq T} \exp(rt) \left(K - \sum_{k=1}^{2} w_k \exp(y_{RN,k}(t)) \right)^{+} \text{ with } w_k = \frac{1}{2}, \tag{2.7}$$

in its put version;

2. an American best-of option, with payoff

$$W(y_{RN,1}(t), y_{RN,2}(t), t) = \sup_{0 \leq t \leq T} \exp(rt) \left(\max[\exp(y_{RN,1}(t)), \exp(y_{RN,2}(t))] - K \right)^{+}, \tag{2.8}$$

in its call version and

$$W(y_{RN,1}(t), y_{RN,2}(t), t) = \sup_{0 \leq t \leq T} \exp(rt) \left(K - \max[\exp(y_{RN,1}(t)), \exp(y_{RN,2}(t))] \right)^{+}, \tag{2.9}$$

in its put version;

3. an American worst-of option, with payoff

$$W(y_{RN,1}(t), y_{RN,2}(t), t) = \sup_{0 \leq t \leq T} \exp(rt) \left(\min[\exp(y_{RN,1}(t)), \exp(y_{RN,2}(t))] - K \right)^{+}, \tag{2.10}$$

in its call version and

$$W(y_{RN,1}(t), y_{RN,2}(t), t) = \sup_{0 \leq t \leq T} \exp(rt) \left(K - \min[\exp(y_{RN,1}(t)), \exp(y_{RN,2}(t))] \right)^{+}, \tag{2.11}$$

in its put version;

4. an American barrier up-and-in equally weighted basket option with payoff given by Eq. (2.6) if call and payoff given by Eq. (2.7) if put, but with the further boundary condition that the option doesn't activate unless at least once within the time frame $[0, T]$, at least one of the two marginal stock prices $S_{RN,1}(t), S_{RN,2}(t)$ touches a certain upper barrier B with $S_1(0), S_2(0) < B$.

2.3 American multi-asset option pricing under Lévy copulae

We will now do the pricing of our four types of multi-asset options, both call and put, using Lévy copulae to describe the dependence structure between our two marginal Lévy processes. We will start by introducing the model and then the pricing will be achieved via simulation under four different types of Lévy copulae (the Clayton Lévy bidirectional copula, the Clayton Lévy unidirectional copula, the Gumbel Lévy copula and the Frank Lévy copula).

2.3.1 Lévy copulae

Lévy copulae were created in order to describe the dependence between a group of Lévy processes, paralleling the notion of a regular copula (Section 1.2) on the level of Lévy measures. In Lévy copulae in fact, instead of working on the cumulative distribution functions like we did in Chapter 1, our marginal processes are considered through their tail integrals, which are defined as:

$$U(x) = \begin{cases} \nu([x, \infty)), & \text{for } x \in (0, \infty) \\ -\nu((-\infty, 0]), & \text{for } x \in (-\infty, 0) \\ 0, & \text{for } x = \infty, -\infty. \end{cases}$$

As in regular copulae, the Lévy ones build their foundations on a modified version of Sklar's Theorem. Therefore Sklar's Theorem on the Lévy copulae setting is defined as follows:
let U be a d-dimensional tail integral with margins $U_1, U_2, ..., U_d$, there exists a unique Lévy copula F such that

$$U(x_1, ..., x_d) = F(U_1(x_1), ..., U_d(x_d)).$$

Therefore, while in the d-dimensional regular copulae you find the joint cumulative distribution function and then by taking partial derivative of order d you find the joint density function, in the Lévy setting you find the joint tail integral and then by taking partial derivative of order d you find the joint Lévy measure which governs the joint jumps.
In fact, let F be a Lévy d-copula, continuous on $\Re^d_\infty$, such that the density

$$\frac{\partial^d F(U_1(x_1), ..., U_d(x_d))}{\partial u_1 ... \partial u_d}$$

exists on $\Re^d$ and let $U_1, ..., U_d$ be one-dimensional tail integrals with densities $\nu_1, ..., \nu_d$. Then

$$\nu(dx_1, ..., dx_d) = \frac{\partial^d F(U_1(x_1), ..., U_d(x_d))}{\partial u_1 ... \partial u_d} \nu_1(dx_1), ..., \nu_d(dx_d)$$

is the Lévy density of a Lévy measure with marginal Lévy measures $\nu_1, ..., \nu_d$ and Lévy copula F.
The function $F : \bar{\Re}^d \to \bar{\Re}$, is called a Lévy copula if:

- $F(u_1, ..., u_d) \neq \infty$ for $(u_1, ..., u_d) \neq (\infty, ..., \infty)$,
- $F(u_1, ..., u_d) = 0$ if $u_k = 0$ for at least one $k \in \{1, ..., d\}$,

- F is d-increasing,

- $F^k(u) = u$ for any $k \in \{1, ...d\}, u \in \Re$.

2.3.2 Simulation of a multidimensional Lévy process

We will now do a brief review of the work by Tankov (2003a), Tankov (2006), Kallsen and Tankov (2006) and Tankov (2016), which provides us with the simulation procedure that we will implement for the pricing of options. Such work proposes to simulate multidimensional Lévy processes based on the conditional probability function, where you first condition on one variable and then simulate the others, and then you iterate the procedure conditioning on every single variable. In order to do this we need to start from the volume function which exists on the interval $(a, b]^d$. For $a, b \in \bar{\Re}^d$ we write $a \leq b$ if $a_k \leq b_k, k = 1, ..., d$. In this case, let $(a, b]$ denote a right-closed and left-open interval on $\bar{\Re}^d$:

$$(a, b] := (a_1, b_1] \times ... \times (a_d, b_d].$$

Then, according to Tankov (2006) *Definition 3.1*, let $F : S \rightarrow \bar{\Re}$ be a function, for some subset $S \subset \bar{\Re}^d$. For $a, b \in S$ with $a \leq b$ and $\overline{(a, b]} \subset S$, the F-volume of $(a, b]$ is defined by

$$V_F((a, b]) := \sum_{u \in \{a_1, b_1\} \times ... \times \{a_d, b_d\}} (-1)^{N(u)} F(u)$$

where $N(u) := \#\{k : u_k = a_k\}$, the cardinality of the set $\{k : u_k = a_k\}$.
F is called d-increasing if $V_F((a, b]) \geq 0$ for all $a, b \in S$.
For example the volume function for $k = 2$ is

$$V_F((a, b]) := F(a_1, a_2) - F(a_1, b_2) - F(b_1, a_2) + F(b_1, b_2)],$$

and similarly it can be extended to any k-dimensional case.

To simulate a Lévy process $(x_t)_{0\leq t\leq 1}$ on $\Re^k$ with Lévy measure ν, we first need to simulate a Poisson random measure on $[0,1]\times\Re^k$ with intensity measure $dt\times\nu$. The Lévy process can then be constructed via the Lévy-Itô decomposition.

Let F be a Lévy copula such that for every $i\in\{1,...,d\}$ nonempty,

$$\lim_{(x_i)_{i\in I}\to\infty} F(x_1,...,x_d) = F(x_1,...,x_d)|_{(x_k)_{k\in I}=\infty}. \tag{2.12}$$

This Lévy copula defines a positive measure μ on $\Re^d$ with Lebesgue margins such that for each $a,b\in\Re^d$ with $a\leq b$,

$$V_F((a,b]) = \mu((a,b]). \tag{2.13}$$

For a one-dimensional tail integral U, the (generalized) inverse tail integral U^{-1} is defined by

$$U^{-1}(u) := \begin{cases} \sup\{x>0 : U(x)\geq u\}\vee 0, u\geq 0 \\ \sup\{x<0 : U(x)\geq u\}, u<0. \end{cases} \tag{2.14}$$

According to Tankov (2006) *Lemma 4.1*, we have that for a Lévy measure ν on $\Re^d$ with marginal tail integrals U_k, $k=1,...,d$, and Lévy copula F satisfying Eq. (2.12), μ defined by Eq. (2.13) and

$$f:(u_1,...,u_d)\mapsto (U_1^{(-1)}(u_1),...,U_d^{(-1)}(u_d)), \tag{2.15}$$

then ν is the image measure of μ by f. To simulate the jumps of a multidimensional Lévy process (more precisely of the corresponding Poisson random measure), we first simulate the jumps in the first component, and then the jumps in the other components conditionally on the jumps in the first one. We therefore proceed by analysing the conditional distributions of μ. According to Ambrosio et al. (2000) *Theorem 2.28* there exists a family, indexed by $\xi \in \Re$, of positive Radon measures $K(\xi, dx_2...dx_d)$ on $\Re^{d-1}$, such that

$$\xi \mapsto K(\xi, dx_2, ..., dx_d)$$

is Borel measurable and

$$\mu(dx_1...dx_d) = dx_1 \times K(x_1, dx_2...dx_d). \tag{2.16}$$

In addition, $K(\xi, \Re^{d-1}) = 1$ for almost all ξ, that is, $K(\xi, \cdot)$ is, almost everywhere, a probability distribution. In the sequel we will call $\{K(\xi, \cdot)\}_{\xi \in \Re}$ the family of conditional probability distributions associated with Lévy copula F.
Let F_ξ be the distribution function of the measure $K(\xi, \cdot)$:

$$F_\xi(x_2, ..., x_d) := K(\xi, (-\infty, x_2] \times ... \times (-\infty, x_d]). \tag{2.17}$$

According to Tankov (2006) *Lemma 4.2*, let F be a Lévy copula satisfying Eq. (2.12), and F_ξ be the corresponding conditional distribution function, defined by Eq. (2.17). Then, there exists a set $N \subset \Re$ of zero Lebesgue measure such that for every fixed $\xi \in \Re \setminus N$, $F_\xi(\cdot)$ is a probability distribution function, satisfying

$$F_\xi(x_2, ..., x_d) = \mathrm{sgn}(\xi) \frac{\partial}{\partial \xi} V_F((\xi \wedge 0, \xi \vee 0] \times (-\infty, x_2] \times ... \times (-\infty, x_d]) \tag{2.18}$$

in every point $(x_2, ..., x_d)$, where F_ξ is continuous.

Simulation of multidimensional Lévy processes, finite variation case

To simulate Lévy processes it is important to pay attention to whether the chosen Lévy process has finite or infinite variation. In our case, since we want to simulate Variance Gamma processes, we are interested in the finite variation case.
As Carr et al. (2002) note, processes of finite variations are potentially more useful than those ones of infinite variation in explaining the measure change from the statistical to the risk neutral process as they allow greater flexibility between the local characteristic of the martingale components under the two measures. In the case of infinite activity processes like the Brownian motion, the volatility, and hence the local martingale component, is invariant under an equivalent change in measure. This equivalence of measure change for infinite variation jump processes implies that the difference between the risk neutral and the statistical Lévy densities is of finite variation. This requires that the two processes have the same exponent. On the other side, if the processes are themselves of finite variation, then the difference in the Lévy densities will automatically be of finite variation and therefore no parametric restriction on the processes is required.
According to Tankov (2006) *Theorem 4.3*, let ν be a Lévy measure on $\Re^d$, satisfying $\int(|x| \wedge 1)\nu(dx) < \infty$, with marginal tail integrals $U_k, k = 1, ..., d$ and Lévy copula $F(x_1, ..., x_d)$, such that the condition of Eq. (2.12) is satisfied, and let $K(x_1, dx_2..., dx_d)$ be the corresponding conditional probability distributions, defined by Eq. (2.17). Let $\{V_k\}$ be a sequence of independent random variables, uniformly distributed on $[0, 1]$. Introduce d random sequences $\{\Gamma_k^1\}, ..., \{\Gamma_k^d\}$, independent from $\{V_k\}$ such that

- $N = \sum_{k=1}^{\infty} \delta_{\{\Gamma_k^1\}}$ is a Poisson random measure on $\Re$ with Lebesgue intensity measure.

- Conditionally on Γ_k^1, the random vector $(\Gamma_k^2, ..., \Gamma_k^d)$ is independent from Γ_j^i with $j \neq k$ and all i and is distributed on $\Re^{d-1}$ with law $K(\Gamma_k^1, dx_2...dx_d)$.

Then

$$(x(\tau,t))_{0\leq t\leq 1}, \text{ with } x_i(t) = \sum_{k=1}^{\infty} U_k^{-1}(\Gamma_k^i)\mathbb{1}_{[0,t]}(V_k), \; i = 1, ..., d \tag{2.19}$$

is a Lévy process on the time interval $[0, 1]$ with characteristic function

$$\mathbb{E}[e^{i\langle u,x(t)\rangle}] = \exp\left(t\int_{\Re^d}(e^{i\langle u,x\rangle} - 1)\nu(dz)\right). \tag{2.20}$$

Therefore, the simulation procedure for two Variance Gamma processes can be summarised as follows:

1. Simulate ξ (i.e. Γ_k^1, where we define $\Gamma_k^1 = X_k(-1)^k$), for $k+1, ..., n$, with X_k being a Poisson process;
2. derive the Volume function;
3. derive the conditional distribution $F_\xi(x_2)$;
4. find the inverse of $F_\xi(x_2)$ or, if not invertible, apply a numerical solution, such as the acceptance rejection method;
5. simulate $(x_2) \mid \xi$.

2.3.3 Longstaff-Schwartz regression model

To price American multi-asset option we use the Longstaff-Schwartz method, which is ideal when option pricing needs to be achieved via simulation. The Longstaff-Schwartz method, developed by Longstaff and Schwartz (2001), uses a dynamic programming approach to find an optimal stopping time, and Monte Carlo to approximate the expected value. Dynamic programming is a general method for solving optimization problems by dividing it into smaller sub-problems and combining their solution to solve the problem.

In this case this means that we divide the interval $[0, T]$, where T is the expiry date of the option, into a finite set of time points $[0, t_1, t_2, ..., t_N]$ where $t_N = T$, and for each of these decide if it is better to exercise than to hold on to the option. Starting from time T and working backwards to time 0, we update the stopping time each time we find a time where it is better to exercise until we have found the smallest time where exercise is better.

To perform such regression the procedure is the following:

1. choose:
   -number of sample paths,
   -number of basis function for the regression,
   -type of basis functions,
   -number of observation dates;
2. simulate the sample paths for the two Variance Gamma processes at each point in time;
3. at expiry record the cash flow values, using the payoffs given by Eqs. (2.6, 2.7, 2.8, 2.9, 2.10 and 2.11);
4. move back to the previous observation date for each path where the payoff is greater than 0 and calculate the continuation value;
5. perform the regression to determine the functional form of the continuation value;
6. recalculate the continuation value;
7. for every two marginal paths of the two underlying assets calculate the corresponding cash flow value;
8. repeat this process for the previous time step until you have all the cash flows;
9. the option value at 0 is then the mean of all the discounted cash flows.

2.3.4 Option pricing under a Clayton Lévy bidirectional copula

The Clayton Lévy copula has two different representations, one on $\Re^d$ (in which case we call it the Clayton Lévy bidirectional copula) and one on $\Re^d_+$ (in which case we simply call it the Clayton Lévy copula), see Kettler (2006). For $k = 2$, the Clayton Lévy bidirectional copula on $\Re^2$ is defined as:

$$F(u,v) = (|u|^{-\theta} + |v|^{-\theta})^{-1/\theta}(\eta \mathbb{1}_{\{uv \geq 0\}} - (1-\eta)\mathbb{1}_{\{uv<0\}}) \tag{2.21}$$

where u and v are the two marginal tail integrals, which in Lévy copulae replace the marginal cumulative functions $F_1(x_1)$ and $F_2(x_2)$ that we had in regular copulae (Section 1.2). F is a Lévy copula for any $\theta > 0$ and $\eta \in [0,1]$. The parameter η determines the dependence of the sign of jumps: when $\eta = 1$, the two components always jump in the same direction, and when $\eta = 0$, positive jumps in one component are accompanied by negative jumps in the other and vice versa. The parameter θ is responsible for the dependence of absolute values of jumps in different components. In particular, if $\eta = 0$ and $\theta \to 0$, the two components become independent whereas the case $\eta = 1$ and $\theta \to \infty$ corresponds to complete dependence.

At this point, to simulate our two-dimensional Variance Gamma process under the Clayton Lévy bidirectional copula, we need to find the inverse of the conditional distribution $\{K(\xi, x_2)\}_{\xi \in \Re}$ from the volume function for a Clayton Lévy bidirectional copula using the procedure summarized in Subsection 2.3.2. For the volume function, we have:

$$\begin{aligned} V_F((a,b]) &= V_F\left(\left(\left(\begin{array}{c} a_1 \\ a_2 \end{array}\right)\right), \left(\begin{array}{c} b_1 \\ b_2 \end{array}\right)\right] \\ &= F(a_1,a_2) - F(a_1,b_2) - F(b_1,a_2) + F(b_1,b_2), \end{aligned}$$

which, for a Clayton Lévy bidirectional copula, can be rephrased as:

$$\begin{aligned}V_F((a,b]) &= (|a_1|^{-\theta} + |a_2|^{-\theta})^{-1/\theta}(\eta \mathbb{1}_{a_1a_2\geq 0} - (1-\eta)\mathbb{1}_{a_1a_2<0})\\ &- (|a_1|^{-\theta} + |b_2|^{-\theta})^{-1/\theta}(\eta \mathbb{1}_{a_1b_2\geq 0} - (1-\eta)\mathbb{1}_{a_1b_2<0})\\ &- (|b_1|^{-\theta} + |a_2|^{-\theta})^{-1/\theta}(\eta \mathbb{1}_{b_1a_2\geq 0} - (1-\eta)\mathbb{1}_{b_1a_2<0})\\ &+ (|b_1|^{-\theta} + |b_2|^{-\theta})^{-1/\theta}(\eta \mathbb{1}_{b_1b_2\geq 0} - (1-\eta)\mathbb{1}_{b_1b_2<0}).\end{aligned}$$

The quantity $V_F\left(\left(\begin{pmatrix}\xi\wedge 0\\ -\infty\end{pmatrix}, \begin{pmatrix}\xi\vee 0\\ x_2\end{pmatrix}\right]\right)$, which appears in Eq. (2.18), will now be analysed separately for the following four cases:

1. $\xi \geq 0$ and $x_2 \geq 0$
2. $\xi < 0$ and $x_2 \geq 0$
3. $\xi \geq 0$ and $x_2 < 0$
4. $\xi < 0$ and $x_2 < 0$.

For the first case we have:

$$a_1 = 0, a_2 = -\infty, b_1 = \xi \geq 0, b_2 = x_2 \geq 0$$

and by substituting in the Clayton Lévy bidirectional copula:

$$V_F(\cdot) = 0 - (|x_2|^{-\theta})^{-1/\theta}\eta - (|\xi|^{-\theta})^{-1/\theta}(-(1-\eta)) + (|\xi|^{-\theta} + |x_2|^{-\theta})^{-1/\theta}\eta.$$

For the second case we have:

$$a_1 = \xi < 0, a_2 = -\infty, b_1 = 0, b_2 = x_2 > 0$$

and by substituting in the Clayton Lévy bidirectional copula:

$$V_F(\cdot) = (|\xi|^{-\theta})^{-1/\theta}\eta - (|\xi|^{-\theta} + |x_2|^{-\theta})^{-1/\theta}(-(1-\eta)) - 0 + (|x_2|^{-\theta})^{-1/\theta}\eta.$$

For the third case we have:

$$a_1 = 0, a_2 = -\infty, b_1 = \xi > 0, b_2 = x_2 < 0$$

and by substituting in the Clayton Lévy bidirectional copula:

$$V_F(\cdot) = 0 - (|x_2|^{-\theta})^{-1/\theta}\eta - (|\xi|^{-\theta})^{-1/\theta}(-(1-\eta)) + (|\xi|^{-\theta} + |x_2|^{-\theta})^{-1/\theta}(-(1-\eta)).$$

Finally, for the fourth case we have:

$$a_1 = \xi < 0, a_2 = -\infty, b_1 = 0, b_2 = x_2 < 0$$

and by substituting in the Clayton Lévy bidirectional copula:

$$V_F(\cdot) = (|\xi|^{-\theta})^{-1/\theta}\eta - (|\xi|^{-\theta} + |x_2|^{-\theta})^{-1/\theta}\eta - 0 + (|x_2|^{-\theta})^{-1/\theta}\eta.$$

Putting these four cases together we obtain the volume function. Then, differentiating the volume function with respect to ξ, we find the conditional probability function:

$$\begin{aligned} F_\xi(x_2) = &((1-\eta) + (1 + |\xi/x_2|^{\theta})^{-1-1/\theta}(\eta \mathbb{1}_{x_2<0}))\mathbb{1}_{\xi \geq 0} \\ &+(\eta + (1 + |\xi/x_2|^{\theta})^{-1-1/\theta}(\mathbb{1}_{x_2 \geq 0} - \eta))\mathbb{1}_{\xi<0}. \end{aligned} \tag{2.22}$$

This conditional distribution function can be inverted analytically. More precisely, we have:

$$F_{\xi}^{-1}(u) = B(\xi, u)|\xi|(C(\xi, u)^{-\theta/(\theta+1)} - 1)^{-1/\theta},$$

with

$$B(\xi, u) = \operatorname{sgn}(u - 1 + \eta)\mathbb{1}_{\xi \geq 0} + \operatorname{sgn}(u - \eta)\mathbb{1}_{\xi < 0}$$

and

$$\begin{aligned} C(\xi, u) = & (((u - 1 + \eta)/\eta)\mathbb{1}_{u \geq 1-\eta} + ((1 - \eta - u)/(1 - \eta))\mathbb{1}_{u < 1-\eta})\mathbb{1}_{\xi \geq 0} \\ & + (((u - \eta)/(1 - \eta))\mathbb{1}_{u \geq \eta} + ((\eta - u)/\eta)\mathbb{1}_{u < \eta})\mathbb{1}_{\xi < 0}. \end{aligned}$$

Let us assume that the marginal Variance Gamma processes $x_k(t; \sigma_k, \alpha_k, \theta_k)$ for $k = 1, 2$ have marginal parameters $\sigma_k = 0.6$, $\alpha_k = 0.3$ and $\theta_k = 0.05$ for both processes, while our American multi-asset options have the following characteristics: $r = 0.05$, $S_1(0) = S_2(0) = K$, with $K = 300$, expiry date $T = 1$ (one year) and for the barrier option $B = 340$. We can now simulate the two marginal stock prices defined by Eq. (2.3) under the Clayton Lévy bidirectional copula: we simulate $H = 10,000$ sample paths for the bivariate Variance Gamma process and we apply variance-reduction on the simulations through the antithetic variables method. At this point we risk-neutralise each simulated path by applying the mean correction martingale method given by Eq. (2.5) and end up with the risk-neutralised paths defined in Eq. (2.4).

We can now apply the Longstaff-Schwartz algorithm (as described in Subsection 2.3.3) with number of observation dates $N = 1,000$ and three basis functions from the set of (weighted) Laguerre polynomials. Laguerre Polynomials are solutions $L_n(x)$ to the

Laguerre differential equation with $\nu = 0$, where the Laguerre differential equation is:

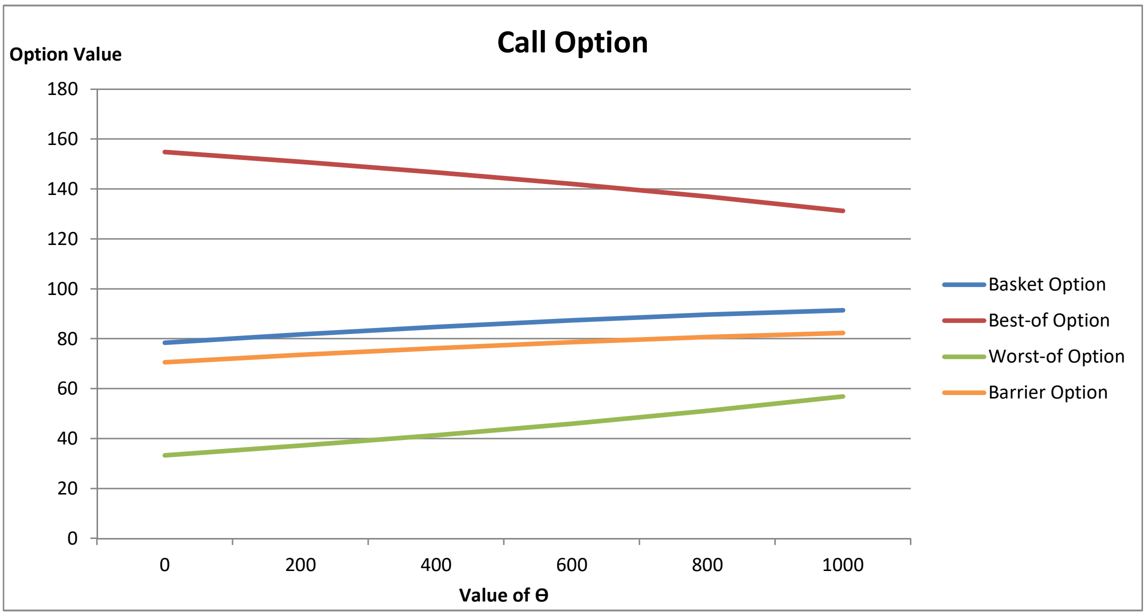
$$xy'' + (1-x)y' + \lambda y = 0.$$

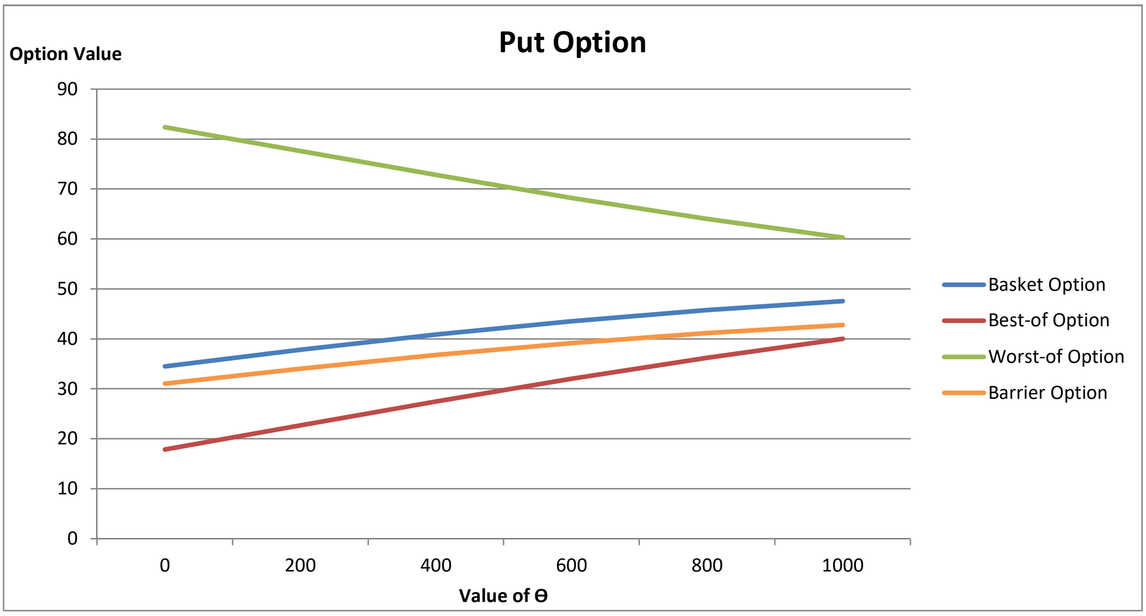
where the first three Laguerre polynomials are:

$$\begin{aligned} L_0(x) &= \exp(-x/2) \\ L_1(x) &= \exp(-x/2)(1-x) \\ L_2(x) &= \exp(-x/2)(1-2x+(x)^2/2). \end{aligned}$$

with $x = \sum_{k=1}^{2} w_k \exp(y_{RN,k}(t))$ if we are trying to price an equally-weighted basket option, $x = \max[\exp(y_{RN,1}(t)), \exp(y_{RN,2}(t))]$ in the case of a best-of option and $x = \min[\exp(y_{RN,1}(t)), \exp(y_{RN,2}(t))]$ in the case of a worst-of option. As our interest lies on how a different dependence structure may affect the price of our financial product, we now show how the prices of the options change for different values of the parameters θ and η. In Figure 2.1 we see the prices of the four types of multi-asset options, in their call and put versions, for a value of θ ranging from 0 to $1,000$ but fixing the value of η. We randomly decided to fix it at $\eta = 0.5$.

Figure 2.1: Option prices under the Clayton Lévy bidirectional copula for a varying θ

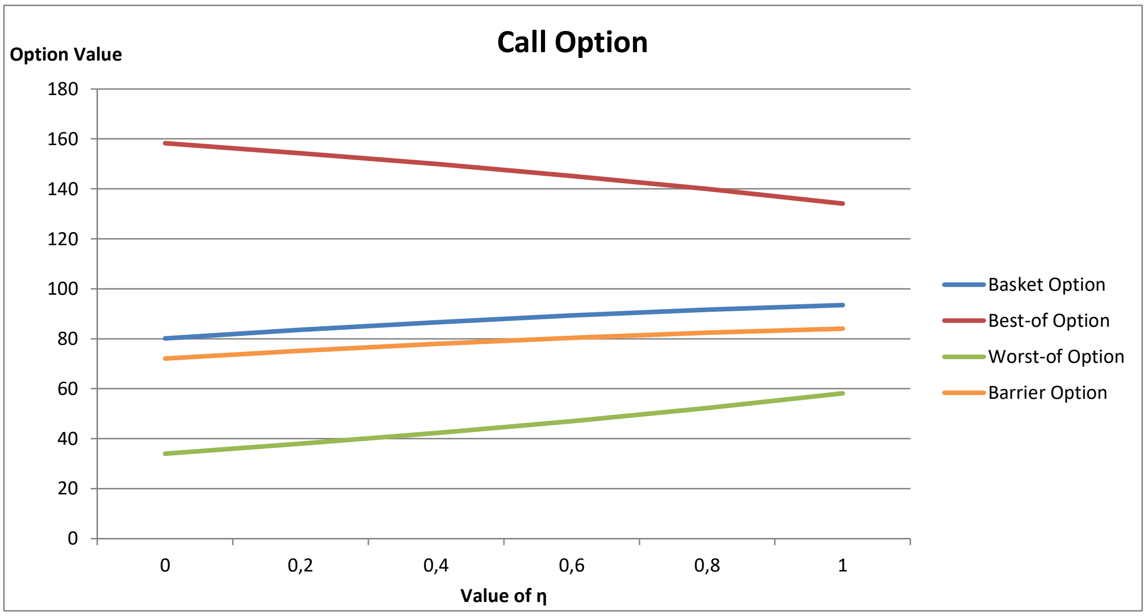


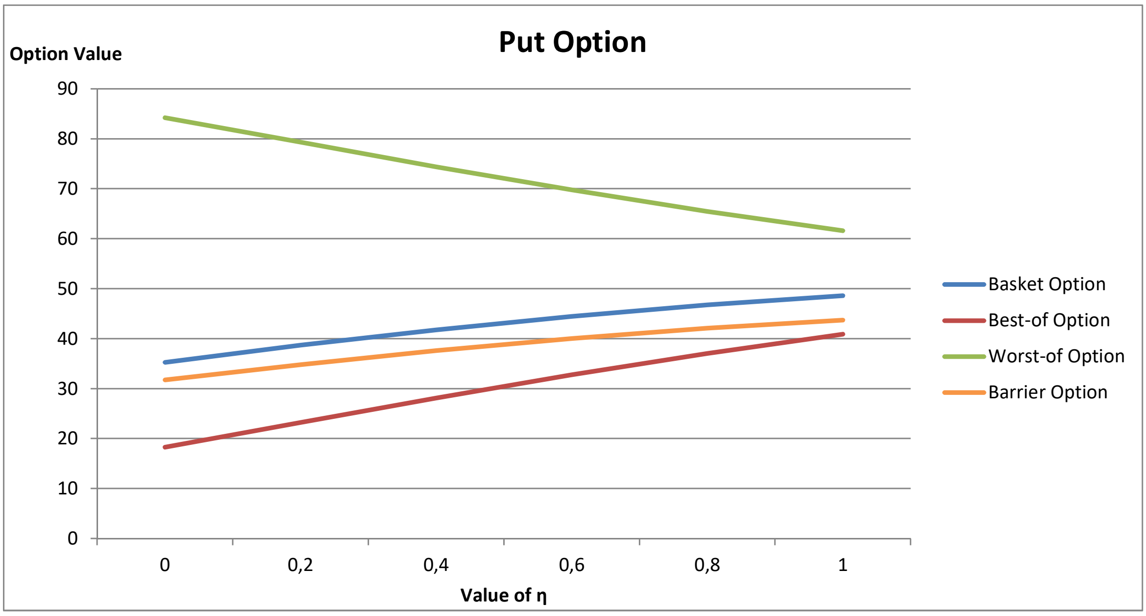


The results are perfectly aligned with what expected: the values of the options are very high and that is due to the high volatility of the Brownian motion σ_k of the marginal Variance Gamma processes and it is due to the long remaining life from 0 to T of the option.
In their call versions, the best-of option has the highest payoff, since it chooses the underlying that grants the highest payoff, the worst-of option has the lowest payoff, since it chooses the underlying that grants the lowest payoff, whereas the basket has a payoff in the middle since it takes an average of the two underlying assets. Finally, the barrier option has a payoff very similar to the one of the basket, but sightly lower, consistent with the fact that the barrier option has a very similar payoff to the one of the basket, but with an extra boundary condition that decreases its value since it reduces the number of times in which the option activates and therefore the number of times one can actually exercise the option. Since the barrier B is set quite low, while, as already mentioned, the volatility of the Brownian motion σ_k of the marginal Variance Gamma processes is quite high and the expiry date of the option is far away in the future, the value of the barrier option is only slightly lower than the basket, because the probability of at least one underlying asset hitting the barrier and therefore the option activating is very high.
Furthermore the best-of option has a payoff which decreases if the correlation increases, because of course if both assets move in the same direction the probability of having one of the two assets skyrocketing decreases. In similar fashion, the worst-of option has a payoff which increases if the correlation increases, because the probability of having one of the two assets collapsing shrinks. The basket option also has an increasing payoff at the increasing of the correlation, because the more the correlation, the more the option's volatility rises and therefore the payoff of the option increases accordingly. The same holds for the barrier option, since its payoff is very similar to the one of the basket. Since the put options have a flipped payoff, their rationale is precisely inverse to the one

which we described for the call options. In Figure 2.2 we analyse how the option prices change at the varying of the parameter η from 0 to 1, when the value of θ is fixed at a level randomly chosen as $\theta = 100$. In this case the interpretation of the results given for Figure 2.1 still holds, since we are only playing with the other measure of association.

Figure 2.2: Option prices under the Clayton Lévy bidirectional copula for a varying η





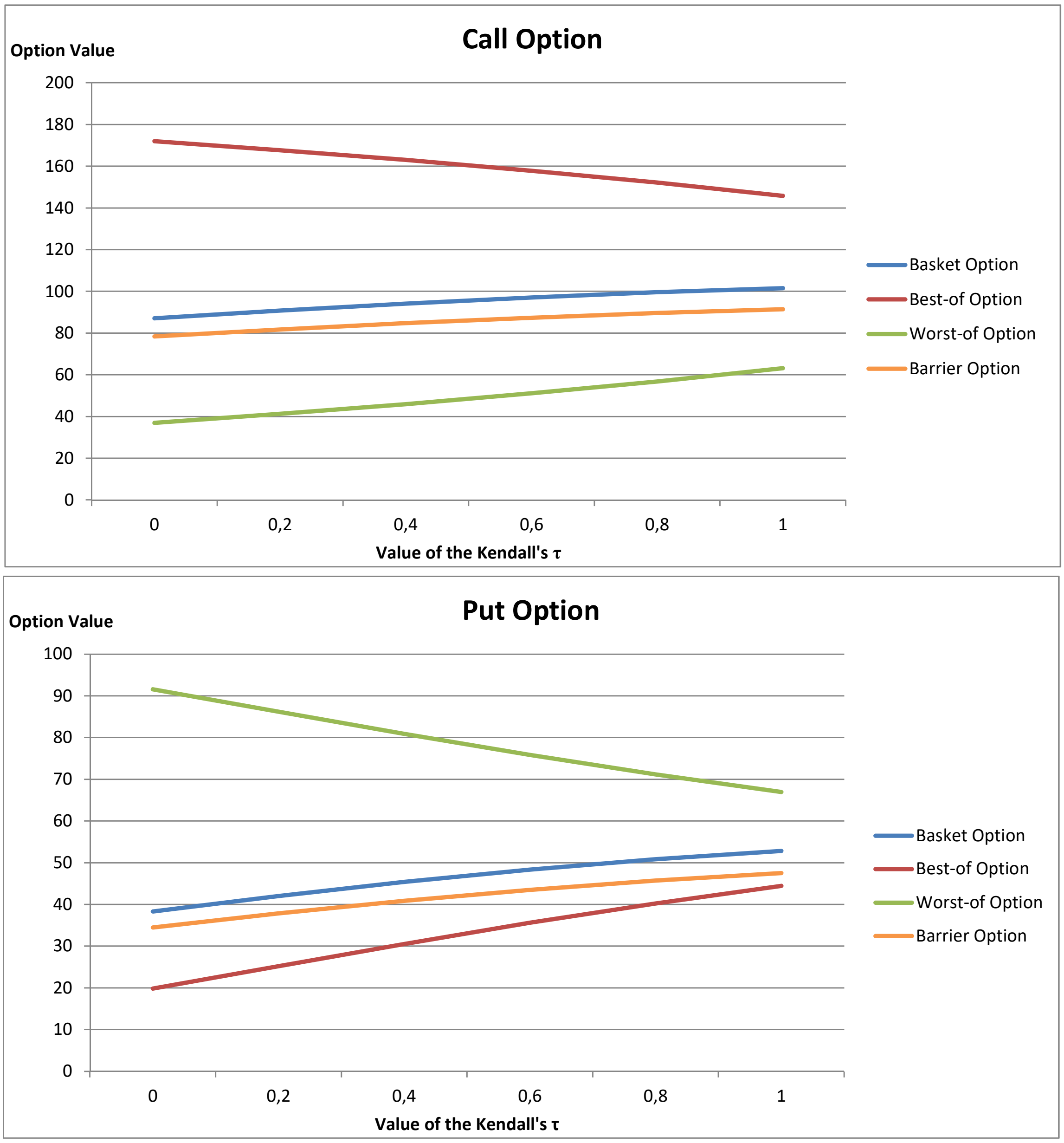
2.3.5 Option pricing under a Clayton Lévy copula

We now repeat the procedure of Subsection 2.3.2 for the case of a Clayton Lévy (unidirectional) copula. The formula for the joint cumulative function in the case of a regular Clayton (unidirectional) copula was given by Eq. (1.2): we will now readjust such formula to take into account that we are not dealing with two marginal cumulative functions $F_1(x_1)$ and $F_2(x_2)$ but we have two marginal tail integrals u and v respectively. Therefore the joint tail integral is given by:

$$F(u, v) = \max\left[\left(u^{-\theta} + v^{-\theta} - 1\right)^{1/\theta}, 0\right], \tag{2.23}$$

where $\theta > 0$. We now want to simulate paths for our two marginal Variance Gamma processes with the marginal parameters and number of simulations defined in Subsection 2.3.4 under this copula structure: we therefore follow the procedure of Subsection 2.3.2, we then apply the variance reduction technique and risk-neutralise like we did in Subsection 2.3.4, we employ the Longstaff-Schwartz algorithm, with the type and number of basis functions of Subsection 2.3.4, and obtain the option pricing results for the four types of multi-asset options, in their call and put versions, with options' characteristics defined in Subsection 2.3.4. In Figure 2.3 we present the results for our options if we consider a positive correlation between the two Variance Gamma processes described by a Kendall's τ ranging from 0 to 1. The rationale explained before is still valid also for the unidirectional version of the Clayton Lévy copula.

Figure 2.3: Option prices under the Clayton Lévy copula for a varying Kendall's τ



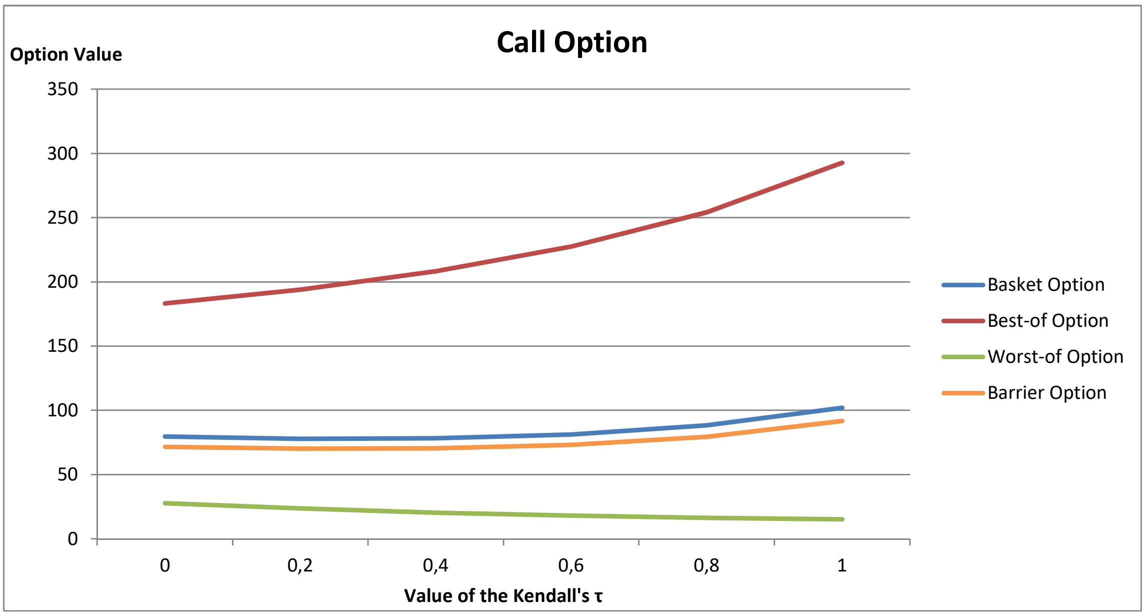
2.3.6 Option pricing under a Gumbel Lévy copula

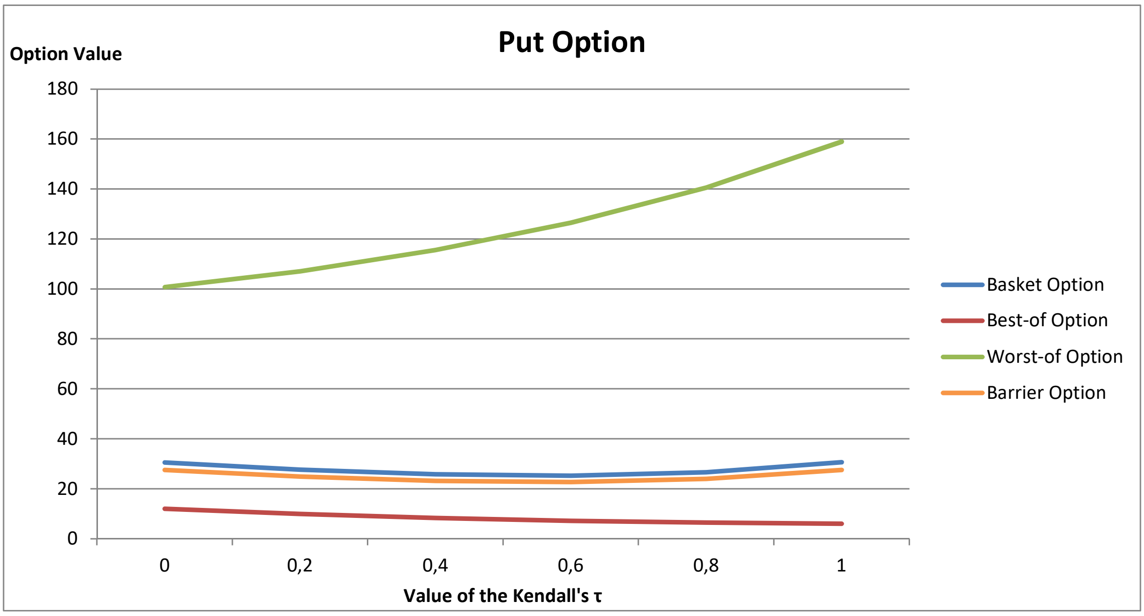
The formula for the joint cumulative function in the case of a regular Gumbel copula was given by Eq. (1.4): readjusting it to the Lévy setting, we obtain the following joint tail integral:

$$F(u,v) = \exp\left[-\left[(-\ln(u))^{\theta} + (-\ln(v))^{\theta}\right]^{1/\theta}\right], \tag{2.24}$$

where $\theta > 0$. Following always the same procedure, in Figure 2.4 we present the results for our options if we consider a positive correlation between the two Variance Gamma processes described by a Kendall's τ ranging from 0 to 1. For the Gumbel copula the trends of the best-of and worst-of options at the increasing on the correlation are symmetrical to the ones under the Clayton copula. This is due to the fact that the Clayton copula stresses the correlation on the lower tail, while the Gumbel stresses the dependence on the upper tail of the joint Lévy measure and therefore the rationale is symmetrical to the one identified under the Clayton.

Figure 2.4: Option prices under the Gumbel Lévy copula for a varying Kendall's τ





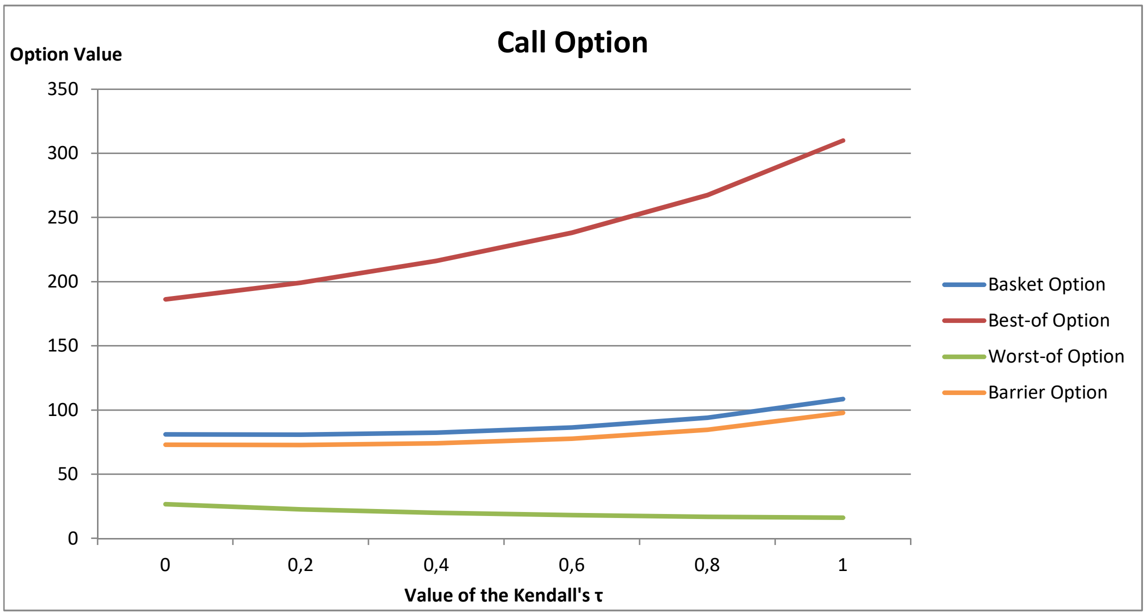
2.3.7 Option pricing under a Frank Lévy copula

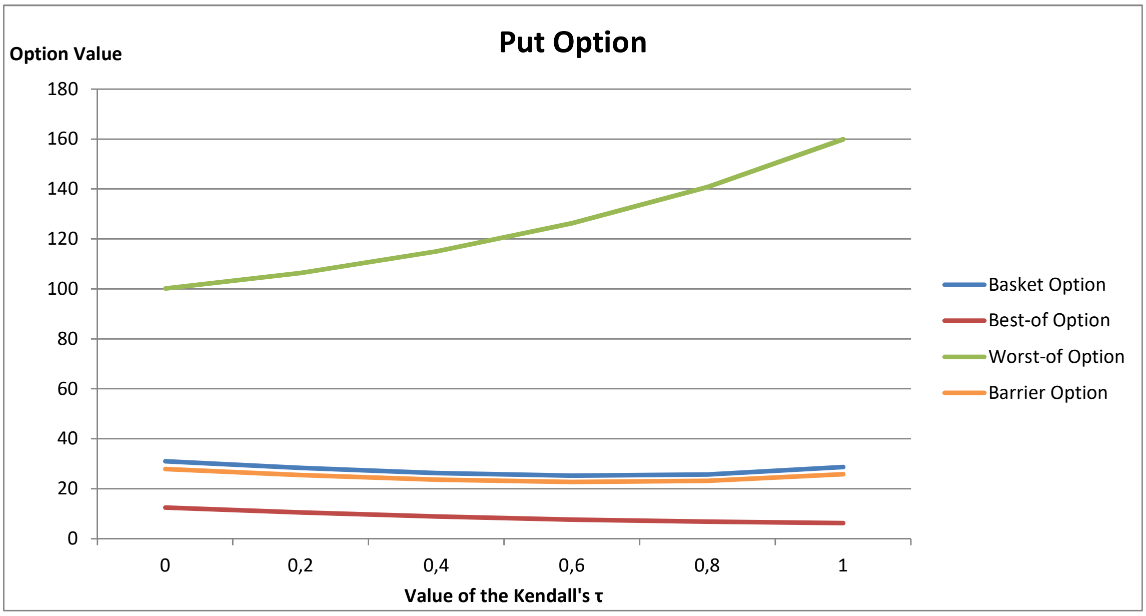
The formula for the joint cumulative function in the case of a regular Frank copula was given by Eq. (1.6): readjusting it to the Lévy setting, we obtain the following joint tail integral:

$$F(u,v) = -\frac{1}{\theta}\ln\left[1+\frac{(\exp(-\theta u)-1)\cdot(\exp(-\theta v)-1)}{(\exp(-\theta)-1)}\right], \tag{2.25}$$

where $\theta > 0$. Following always the same procedure, in Figure 2.5 we present the results for our options if we consider a positive correlation between the two Variance Gamma processes described by a Kendall's τ ranging from 0 to 1. For the Frank copula, like under the Gumbel, the trends of the best-of and worst-of options at the increasing on the correlation are symmetrical to the ones under the Clayton copula. In fact the Frank copula, like the Gumbel, stresses the dependence in higher percentiles of the joint Lévy measure and therefore the rationale is symmetrical to the one identified under the Clayton.

Figure 2.5: Option prices under the Frank Lévy copula for a varying Kendall's τ





2.4 American multi-asset option pricing under the Ballotta-Bonfiglioli model

We will now do the pricing of our four types of multi-asset options, both call and put, using the Ballotta-Bonfiglioli model to describe the dependence structure between our two marginal Lévy processes. We will start by introducing the model and then the pricing will be achieved in two ways: firstly, by solving the partial integro-differential equation (i.e. PIDE) through finite-difference method and secondly, via simulation.

2.4.1 The Ballotta-Bonfiglioli model on a two-dimensional multi-asset option

The Ballotta-Bonfiglioli model in Ballotta and Bonfiglioli (2016) structures the dependence between our two marginal Lévy processes by describing each Lévy process as the sum of two random variables, where the first one is an idiosyncratic part (which is therefore unique for each Lévy process) and the second one is a common factor (which is the same for each Lévy process and is therefore responsible for the dependence between the two Lévy processes). In each Lévy process the common factor is multiplied by a constant b_k (with $k = 1, 2$ for our two-dimensional case) which can be different for each Lévy process. The joint two-dimensional process is proved to be still Lévy. In our case we have two marginal Lévy processes $Y_1(t)$ and $Y_2(t)$, which will be represented as:

$$
\begin{aligned}
Y_1(t) &= x_1(t; \sigma_1, \alpha_1, \theta_1) + b_1 x_3(t; \sigma_3, \alpha_3, \theta_3) \\
Y_2(t) &= x_2(t; \sigma_2, \alpha_2, \theta_2) + b_2 x_3(t; \sigma_3, \alpha_3, \theta_3)
\end{aligned}
\tag{2.26}
$$

where $x_1(t; \sigma_1, \alpha_1, \theta_1)$, $x_2(t; \sigma_2, \alpha_2, \theta_2)$ and $x_3(t; \sigma_3, \alpha_3, \theta_3)$ are independent Variance Gamma processes on the probability space $(\Omega, \mathcal{F}, P)$. The two idiosyncratic processes are denoted as $x_1(t; \sigma_1, \alpha_1, \theta_1)$ and $x_2(t; \sigma_2, \alpha_2, \theta_2)$ with characteristic functions $\phi_{x_k(t;\sigma_k,\alpha_k,\theta_k)}(u, t)$, for $k = 1, 2$ respectively, and the common factor being $x_3(t; \sigma_3, \alpha_3, \theta_3)$ with characteristic function $\phi_{x_3(t;\sigma_3,\alpha_3,\theta_3)}(u, t)$ and $b_1, b_2 \in \Re$ being the constants. The joint two-dimensional process

$$\begin{aligned} Y(t) &= (Y_1(t), Y_2(t))^T \\ &= (x_1(t; \sigma_1, \alpha_1, \theta_1) + b_1 x_3(t; \sigma_3, \alpha_3, \theta_3), x_2(t; \sigma_2, \alpha_2, \theta_2) + b_2 x_3(t; \sigma_3, \alpha_3, \theta_3))^T \end{aligned}$$

is a Lévy process on $\Re^2$ with characteristic function:

$$\phi_{Y(t)}(u; t) = \phi_{x_3(t;\sigma_3,\alpha_3,\theta_3)}\left(\sum_{k=1}^{2} b_k u_k; t\right) \prod_{k=1}^{2} \phi_{x_k(t;\sigma_k,\alpha_k,\theta_k)}(u_k; t)$$

with $u \in \Re^2$.

The correlation between the two marginal Lévy processes $Y_1(t)$ and $Y_2(t)$ is equal to:

$$\text{Corr}(Y_1(t), Y_2(t)) = \frac{b_1 b_2 \text{Var}(x_3(t; \sigma_3, \alpha_3, \theta_3))}{\sqrt{\text{Var}(x_1(t; \sigma_1, \alpha_1, \theta_1))}\sqrt{\text{Var}(x_2(t; \sigma_2, \alpha_2, \theta_2))}}. \tag{2.27}$$

At this point, to price our four types of options, both call and put, as described in Subsection 2.2.2, we need to exponentially transform our two marginal Lévy processes and we need to pass to the risk neutral space $(\Omega, \mathcal{F}, Q)$ and we obtain:

$$\begin{aligned} S_{RN,k}(t) = S_k(0) \exp((r + q + \omega_{RN,k} + \omega_{RN,3})t + x_k(t; \sigma_{RN,k}, \alpha_{RN,k}, \theta_{RN,k}) \\ + b_k x_3(t; \sigma_{RN,3}, \alpha_{RN,3}, \theta_{RN,3}), \quad \text{for } k = 1, 2, \end{aligned}$$

and therefore the log-transforms are:

$$\begin{aligned} y_{RN,k}(t) = \ln(S_k(0)) + (r + q + \omega_{RN,k} + \omega_{RN,3})t + x_k(t; \sigma_{RN,k}, \alpha_{RN,k}, \theta_{RN,k}) \\ + b_k x_3(t; \sigma_{RN,3}, \alpha_{RN,3}, \theta_{RN,3}), \quad \text{for } k = 1, 2. \end{aligned}$$

The price of our options therefore depends on four sources of randomness: $t, x_1(\cdot), x_2(\cdot)$ and $x_3(\cdot)$. Our problem has therefore gained an additional dimension from the Lévy copulae setting: from a two-dimensional Variance Gamma process to a three-dimensional one. Let us assume $b_1 = b_2 = b$, $S_1(0) = S_2(0) = S(0)$, and let us isolate the common factor $x_3(t; \sigma_{RN,3}, \alpha_{RN,3}, \theta_{RN,3})$, so to maintain $x_k(t; \sigma_{RN,k}, \alpha_{RN,k}, \theta_{RN,k})$ for $k = 1, 2, 3$ as three independent sources of randomness. We can rewrite the payoff of our function $W(y_{RN,1}(t), y_{RN,2}(t), t)$ as:

$$W(y_{RN,1}(t), y_{RN,2}(t), t) = W(z_1, z_2, z_3, t) = \left(\frac{1}{2}\exp(z_3)(\exp(z_1) + \exp(z_2) - K)\right)^+,$$

in the case of an equally-weighted basket call option; as:

$$W(y_{RN,1}(t), y_{RN,2}(t), t) = W(z_1, z_2, z_3, t) = \left(K - \frac{1}{2}\exp(z_3)(\exp(z_1) + \exp(z_2))\right)^+,$$

in the case of an equally-weighted basket put option; as:

$$W(y_{RN,1}(t), y_{RN,2}(t), t) = W(z_1, z_2, z_3, t) = \Big(\max[\exp(z_1)\exp(z_3), \exp(z_2)\exp(z_3)] - K\Big)^+,$$

in the case of a best-of call option; as:

$$W(y_{RN,1}(t), y_{RN,2}(t), t) = W(z_1, z_2, z_3, t) = \Big(K - \max[\exp(z_1)\exp(z_3), \exp(z_2)\exp(z_3)]\Big)^+,$$

in the case of a best-of put option; as:

$$W(y_{RN,1}(t), y_{RN,2}(t), t) = W(z_1, z_2, z_3, t) = \Big(\min[\exp(z_1)\exp(z_3), \exp(z_2)\exp(z_3)] - K\Big)^+,$$

in the case of a worst-of call option; and as:

$$W(y_{RN,1}(t), y_{RN,2}(t), t) = W(z_1, z_2, z_3, t) = \Big(K - \min[\exp(z_1)\exp(z_3), \exp(z_2)\exp(z_3)]\Big)^+,$$

in the case of a worst-of put option; with:

$$z_1 = \frac{1}{2}\ln(S(0)) + \frac{1}{2}(r+q)t + x_1(t; \sigma_{RN,1}, \alpha_{RN,1}, \theta_{RN,1}) + \omega_{RN,1}t,$$

$$z_2 = \frac{1}{2}\ln(S(0)) + \frac{1}{2}(r+q)t + x_2(t; \sigma_{RN,2}, \alpha_{RN,2}, \theta_{RN,2}) + \omega_{RN,2}t,$$

$$z_3 = \frac{1}{2}\ln(S(0)) + \frac{1}{2}(r+q)t + bx_3(t; \sigma_{RN,3}, \alpha_{RN,3}, \theta_{RN,3}) + \omega_{RN,3}t.$$

2.4.2 Multidimensional PIDE on Variance Gamma processes

We are now ready to extend the work by Hirsa and Madan (2004), Fiorani (1999) and Fiorani (2004) to the multidimensional case. These authors, in fact, had considered an American mono-asset option (mainly they focused on the cases of a plain vanilla and of a basket option) assuming that the one underlying asset followed a Variance Gamma process and had achieved the pricing solving the PIDE via finite difference method. We will follow the same procedure but considering four sources of randomness, i.e. $t, x_1(\cdot), x_2(\cdot)$ and $x_3(\cdot)$, instead of the two under which the work by those authors was built, i.e. t and the one Variance Gamma process called $x(\cdot)$.

As mentioned, the procedure considers doing the option pricing by solving the corresponding partial integro-differential equation, which needs to be built via generator, given that f is a sufficiently smooth function. The generator in our three-dimensional

Lévy process can be defined as:

$$
\begin{aligned}
\mathcal{L}(f) \triangleq\ & \frac{\partial f(z_1, z_2, z_3, t)}{\partial t} + (r - q + \omega_{RN,1}) \frac{\partial f(z_1, z_2, z_3, t)}{\partial z_1} \\
& + (r - q + \omega_{RN,2}) \frac{\partial f(z_1, z_2, z_3, t)}{\partial z_2} + (r - q + \omega_{RN,3}) \frac{\partial f(z_1, z_2, z_3, t)}{\partial z_3} \\
& + \frac{1}{2}(r - q + \omega_{RN,1}) \frac{\partial^2 f(z_1, z_2, z_3, t)}{\partial z_1^2} + \frac{1}{2}(r - q + \omega_{RN,2}) \frac{\partial^2 f(z_1, z_2, z_3, t)}{\partial z_2^2} \\
& + \frac{1}{2}(r - q + \omega_{RN,3}) \frac{\partial^2 f(z_1, z_2, z_3, t)}{\partial z_3^2} \\
& + \int_{-\infty}^{\infty} [f(z_1 + x_1, z_2, z_3, t) - f(z_1, z_2, z_3, t)] \nu_1(x_1) dx_1 \\
& + \int_{-\infty}^{\infty} [f(z_1, z_2 + x_2, z_3, t) - f(z_1, z_2, z_3, t)] \nu_2(x_2) dx_2 \\
& + \int_{-\infty}^{\infty} [f(z_1, z_2, z_3 + x_3, t) - f(z_1, z_2, z_3, t)] \nu_3(x_3) dx_3 \\
& - r f(z_1, z_2, z_3, t),
\end{aligned}
\tag{2.28}
$$

where $\nu_k(x_k)$ for $k = 1, ..., 3$ are the Lévy measures for the three independent marginal Variance-Gamma processes. We will now only show the procedure step by step for the case of the payoff of an American call equally-weighted basket option: showing it under all four types of multi-asset options in their call and put versions would be redundant, as the procedure does not vary, only the payoff of the option needs to be replaced accordingly. We define the exercise region as the area where

$$
\begin{aligned}
& \frac{1}{2} S(0) \exp((r + q + \omega_{RN,1} + \omega_{RN,3})t + x_1(t; \sigma_{RN,1}, \alpha_{RN,1}, \theta_{RN,1}) + b x_3(t; \sigma_{RN,3}, \alpha_{RN,3}, \theta_{RN,3})) \\
& + \frac{1}{2} S(0) \exp((r + q + \omega_{RN,2} + \omega_{RN,3})t + x_2(t; \sigma_{RN,2}, \alpha_{RN,2}, \theta_{RN,2}) + b x_3(t; \sigma_{RN,3}, \alpha_{RN,3}, \theta_{RN,3})) \\
& > \frac{1}{2} S(0) \exp((r + q + \omega_{RN,1} + \omega_{RN,3})\tau + x_1(\tau; \sigma_{RN,1}, \alpha_{RN,1}, \theta_{RN,1}) + b x_3(\tau; \sigma_{RN,3}, \alpha_{RN,3}, \theta_{RN,3})) \\
& + \frac{1}{2} S(0) \exp((r + q + \omega_{RN,2} + \omega_{RN,3})\tau + x_2(\tau; \sigma_{RN,2}, \alpha_{RN,2}, \theta_{RN,2}) + b x_3(\tau; \sigma_{RN,3}, \alpha_{RN,3}, \theta_{RN,3})),
\end{aligned}
$$

which basically means the area in which the underlying at time t is more valuable than the underlying at time τ, since we are considering a call option, for a put option it would be the area in which the underlying at time t is less valuable than the underlying at time τ. In such region the following equation is true:

$$
\begin{aligned}
&W(y_{RN,1}(t), y_{RN,2}(t), t) = W(z_1, z_2, z_3, t) \\
&= \frac{1}{2}\exp(y(0) + (r + q + \omega_{RN,1} + \omega_{RN,3})t + x_1(t; \sigma_{RN,1}, \alpha_{RN,1}, \theta_{RN,1}) + bx_3(t; \sigma_{RN,3}, \alpha_{RN,3}, \theta_{RN,3})) \\
&+ \frac{1}{2}\exp(y(0) + (r + q + \omega_{RN,2} + \omega_{RN,3})t + x_2(t; \sigma_{RN,2}, \alpha_{RN,2}, \theta_{RN,2}) + bx_3(t; \sigma_{RN,3}, \alpha_{RN,3}, \theta_{RN,3})) \\
&- K \\
&= \left(\frac{1}{2}\exp(z_3)(\exp(z_1) + \exp(z_2) - K)\right),
\end{aligned}
$$

where $y(0) = \ln(S(0))$. From our generator the continuous time PIDE can be built and solved through the use of finite difference method by discretizing such PIDE and writing it down as a linear system. The most significant part of the process is of course the jump component of the three Variance Gamma processes $x_1(t; \sigma_1, \alpha_1, \theta_1)$, $x_2(t; \sigma_2, \alpha_2, \theta_2)$ and $x_3(t; \sigma_3, \alpha_3, \theta_3)$, which is represented by the three integrals in Eq. (2.28), which we have to rewrite differently to allow the numerical computation. Let us start defining the range of values we are going to consider in our computations as

$$
\begin{gathered}
[0, T] \times [z_{1,\min}, z_{1,\max}] \\
[0, T] \times [z_{2,\min}, z_{2,\max}] \\
[0, T] \times [z_{3,\min}, z_{3,\max}].
\end{gathered}
$$

We can now discretize the system using $M + 1$ mesh points in the z_ks-directions for $k = 1, 2, 3$ and $N + 1$ mesh points in the t-direction, with the size of space and time intervals given respectively by:

$$
\begin{aligned}
\Delta Y_1 &= \frac{z_{1,\max} - z_{1,\min}}{M} \\
\Delta Y_2 &= \frac{z_{2,\max} - z_{2,\min}}{M} \\
\Delta Y_3 &= \frac{z_{3,\max} - z_{3,\min}}{M}
\end{aligned}
$$

and

$$
\Delta t = \frac{T - 0}{N}.
$$

The notation $W(z_{1,i}, z_{2,i}, z_{3,i}, t_j)$ refers to the value of $W(\cdot)$ at the node (i, j) and we use the following approximation of the partial derivatives:

$$
\begin{aligned}
\frac{\partial f(z_1, z_2, z_3, t)}{\partial t} &\approx \frac{W(z_{1,i}, z_{2,i}, z_{3,i}, t_{j+1}) - W(z_{1,i}, z_{2,i}, z_{3,i}, t_j)}{\Delta t}, \\
\frac{\partial f(z_1, z_2, z_3, t)}{\partial z_1} &\approx \frac{W(z_{1,i+1}, z_{2,i}, z_{3,i}, t_j) - W(z_{1,i-1}, z_{2,i}, z_{3,i}, t_j)}{2\Delta z_1}, \\
\frac{\partial f(z_1, z_2, z_3, t)}{\partial z_2} &\approx \frac{W(z_{1,i}, z_{2,i+1}, z_{3,i}, t_j) - W(z_{1,i}, z_{2,i-1}, z_{3,i}, t_j)}{2\Delta z_2}, \\
\frac{\partial f(z_1, z_2, z_3, t)}{\partial z_3} &\approx \frac{W(z_{1,i}, z_{2,i}, z_{3,i+1}, t_j) - W(z_{1,i}, z_{2,i}, z_{3,i-1}, t_j)}{2\Delta z_3}.
\end{aligned}
$$

Finally, we define:

$$
\begin{aligned}
h_1 &\triangleq \frac{(r - q + \omega_{RN,1})\Delta t}{2\Delta z_1} \\
h_2 &\triangleq \frac{(r - q + \omega_{RN,2})\Delta t}{2\Delta z_2} \\
h_3 &\triangleq \frac{(r - q + \omega_{RN,3})\Delta t}{2\Delta z_3}.
\end{aligned}
$$

The PIDE for an American call equally-weighted basket option can be now written according to Hirsa and Madan (2004), Fiorani (1999) and Fiorani (2004) as

$$
\begin{aligned}
&(1+r\Delta t)W(z_{1,i},z_{2,i},z_{3,i},t_j)\\
&+h_1W(z_{1,i-1},z_{2,i},z_{3,i},t_j)-h_1W(z_{1,i+1},z_{2,i},z_{3,i},t_j)\\
&+h_2W(z_{1,i},z_{2,i-1},z_{3,i},t_j)-h_2W(z_{1,i},z_{2,i+1},z_{3,i},t_j)\\
&+h_3W(z_{1,i},z_{2,i},z_{3,i-1},t_j)-h_3W(z_{1,i},z_{2,i},z_{3,i+1},t_j)\\
&=W(z_{1,i},z_{2,i},z_{3,i},t_{j+1})\\
&+\Delta t\int_{-\infty}^{\infty}\Big[W(z_{1,i}+x_1,z_{2,i},z_{3,i},t_{j+1})-W(z_{1,i},z_{2,i},z_{3,i},t_{j+1})\Big]\nu_1(x_1)dx_1\\
&+\Delta t\int_{-\infty}^{\infty}\Big[W(z_{1,i},z_{2,i}+x_2,z_{3,i},t_{j+1})-W(z_{1,i},z_{2,i},z_{3,i},t_{j+1})\Big]\nu_2(x_2)dx_2\\
&+\Delta t\int_{-\infty}^{\infty}\Big[W(z_{1,i},z_{2,i},z_{3,i}+x_3,t_{j+1})-W(z_{1,i},z_{2,i},z_{3,i},t_{j+1})\Big]\nu_3(x_3)dx_3\\
&-\mathbb{1}_{z_{1,i}>z_1(\tau_j+1)}\Delta t\Bigg[rK-\frac{q}{2}e^{z_{3,i}}[e^{z_{1,i}}+e^{z_{2,i}}]\\
&+\int_{-\infty}^{z_1(\tau_j+1)-z_{1,i}}\Big(W(z_{1,i}+x_1,z_{2,i},z_{3,i},t_{j+1})+K-\frac{1}{2}e^{z_{3,i}}[e^{z_{1,i}+x_1}+e^{z_{2,i}}]\Big)\nu_1(x_1)dx_1\Bigg]\\
&-\mathbb{1}_{z_{2,i}>z_2(\tau_j+1)}\Delta t\Bigg[rK-\frac{q}{2}e^{z_{3,i}}[e^{z_{1,i}}+e^{z_{2,i}}]\\
&+\int_{-\infty}^{z_2(\tau_j+1)-z_{2,i}}\Big(W(z_{1,i},z_{2,i}+x_2,z_{3,i},t_{j+1})+K-\frac{1}{2}e^{z_{3,i}}[e^{z_{1,i}}+e^{z_{2,i}+x_2}]\Big)\nu_2(x_2)dx_2\Bigg]\\
&-\mathbb{1}_{z_{3,i}>z_3(\tau_j+1)}\Delta t\Bigg[rK-\frac{q}{2}e^{z_{3,i}}[e^{z_{1,i}}+e^{z_{2,i}}]\\
&+\int_{-\infty}^{z_3(\tau_j+1)-z_{3,i}}\Big(W(z_{1,i},z_{2,i},z_{3,i}+x_3,t_{j+1})+K-\frac{1}{2}e^{z_{3,i}+x_3}[e^{z_{1,i}}+e^{z_{2,i}}]\Big)\nu_3(x_3)dx_3\Bigg].
\end{aligned}
$$

As mentioned, the main challenge in the procedure is solving the integrals in the Heaviside term, in fact the technique requires to break the integrals in pieces depending on the size of the jumps as follows by Fiorani (1999) and Fiorani (2004) in *Appendix A*:

$$
\begin{aligned}
&\int_{-\infty}^{+\infty} \Big[W(z_{1,i}+x_1, z_{2,i}, z_{3,i}, t_{j+1}) - W(z_{1,i}, z_{2,i}, z_{3,i}, t_{j+1})\Big]\nu_1(x_1)dx_1 = \\
&\int_{-\infty}^{z_{1,0}-z_{1,i}} \Big[W(z_{1,i}+x_1, z_{2,i}, z_{3,i}, t_{j+1}) - W(z_{1,i}, z_{2,i}, z_{3,i}, t_{j+1})\Big]\nu_1(x_1)dx_1 \\
&+ \int_{z_{1,0}-z_{1,i}}^{-\Delta z_1} \Big[W(z_{1,i}+x_1, z_{2,i}, z_{3,i}, t_{j+1}) - W(z_{1,i}, z_{2,i}, z_{3,i}, t_{j+1})\Big]\nu_1(x_1)dx_1 \\
&+ \int_{-\Delta z_1}^{0} \Big[W(z_{1,i}+x_1, z_{2,i}, z_{3,i}, t_{j+1}) - W(z_{1,i}, z_{2,i}, z_{3,i}, t_{j+1})\Big]\nu_1(x_1)dx_1 \\
&+ \int_{0}^{\Delta z_1} \Big[W(z_{1,i}+x_1, z_{2,i}, z_{3,i}, t_{j+1}) - W(z_{1,i}, z_{2,i}, z_{3,i}, t_{j+1})\Big]\nu_1(x_1)dx_1 \\
&+ \int_{\Delta z_1}^{z_{1,N}-z_{1,i}} \Big[W(z_{1,i}+x_1, z_{2,i}, z_{3,i}, t_{j+1}) - W(z_{1,i}, z_{2,i}, z_{3,i}, t_{j+1})\Big]\nu_1(x_1)dx_1 \\
&+ \int_{z_{1,N}-z_{1,i}}^{+\infty} \Big[W(z_{1,i}+x_1, z_{2,i}, z_{3,i}, t_{j+1}) - W(z_{1,i}, z_{2,i}, z_{3,i}, t_{j+1})\Big]\nu_1(x_1)dx_1,
\end{aligned}
$$

with the same procedure being applied for

$$\int_{-\infty}^{+\infty} \Big[W(z_{1,i}, z_{2,i}+x_2, z_{3,i}, t_{j+1}) - W(z_{1,i}, z_{2,i}, z_{3,i}, t_{j+1})\Big]\nu_2(x_2)dx_2$$

and for

$$\int_{-\infty}^{+\infty} \Big[W(z_{1,i}, z_{2,i}, z_{3,i}+x_3, t_{j+1}) - W(z_{1,i}, z_{2,i}, z_{3,i}, t_{j+1})\Big]\nu_3(x_3)dx_3.$$

Given that we are in the discrete environment, we need

$$z_k(\tau_{j+1}) - z_{k,i} \geq \Delta z_k \text{ with } k = 1, 2, 3,$$

and we define:

$$z_k(\tau_{j+1}) \triangleq l\Delta z_k = z_{k,l} \text{ with } k = 1, 2, 3,$$

for some integer l between 0 and N whose value determines the position of the exercise boundary in the grid. The main trick that we use to discretize our intervals is using linear interpolation for which, given a jump of size $x_k \in [g\Delta z_k, (g+1)\Delta z_k]$ for $k = 1, 2, 3$ and $g = (i-l), (i-l+1), ..., (i-1)$, we can write:

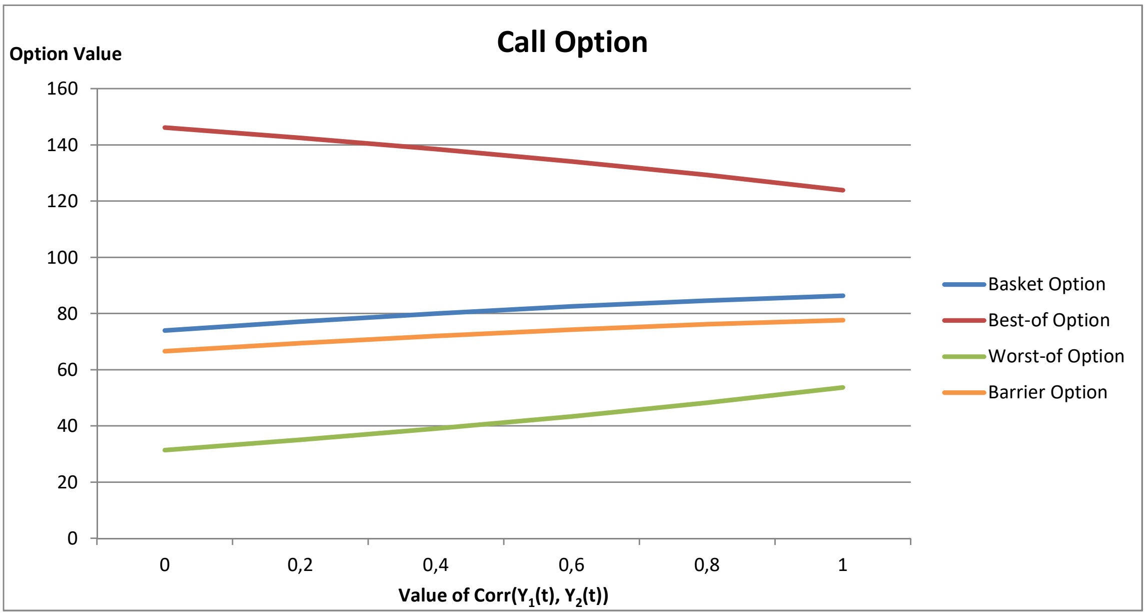
$$W(z_{1,i} + x_1, z_{2,i}, z_{3,i}, t_{j+1}) - W(z_{1,i+k}, z_{2,i}, z_{3,i}, t_{j+1})$$
$$\approx \frac{W(z_{1,i+k+1}, z_{2,i}, z_{3,i}, t_{j+1}) - W(z_{1,i+k}, z_{2,i}, z_{3,i}, t_{j+1})}{\Delta z_1}(x_1 - g\Delta z_1)$$

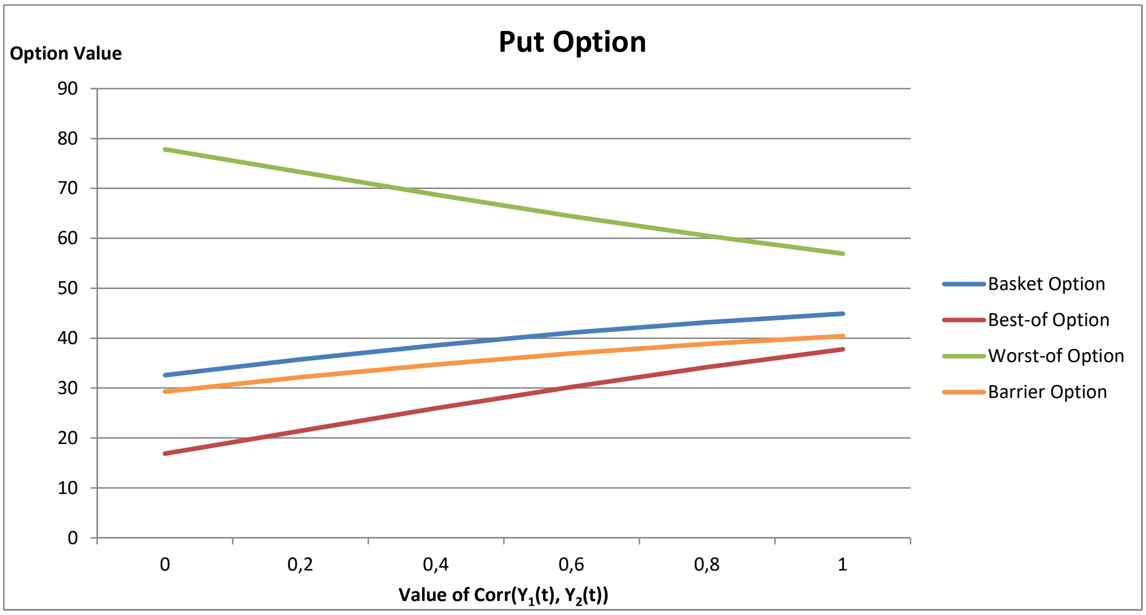
and the same interpolation can be used for z_2 and z_2. As shown in Fiorani (1999) and Fiorani (2004) for the one-dimensional case, which therefore considers the one integral related to the jump component x_1 of z_1, the jump component is integrated as the sum of six smaller integrals. Such integrals still hold for our three independent integrals for the jump components of z_1, z_2 and z_3, so we can simply plug Fiorani's results in our PIDE.

2.4.3 Results of the option pricing via finite-difference method

Now that we have solved the PIDE via finite-difference method, we present some numerical results for the prices of our four types of options, both in their call and put versions, assuming the same marginal parameters for the Variance Gamma processes and the same options' characteristics used in Section 2.3. In Figure 2.6 we present the results for our options if we consider a positive correlation $\text{Corr}(Y_1(t), Y_2(t))$ between the two Variance Gamma processes (see Eq. (2.27) for the formula) ranging from 0 to 1.

Figure 2.6: Option prices via finite-difference method under the Ballotta-Bonfiglioli model for a varying $\text{Corr}(Y_1(t), Y_2(t))$



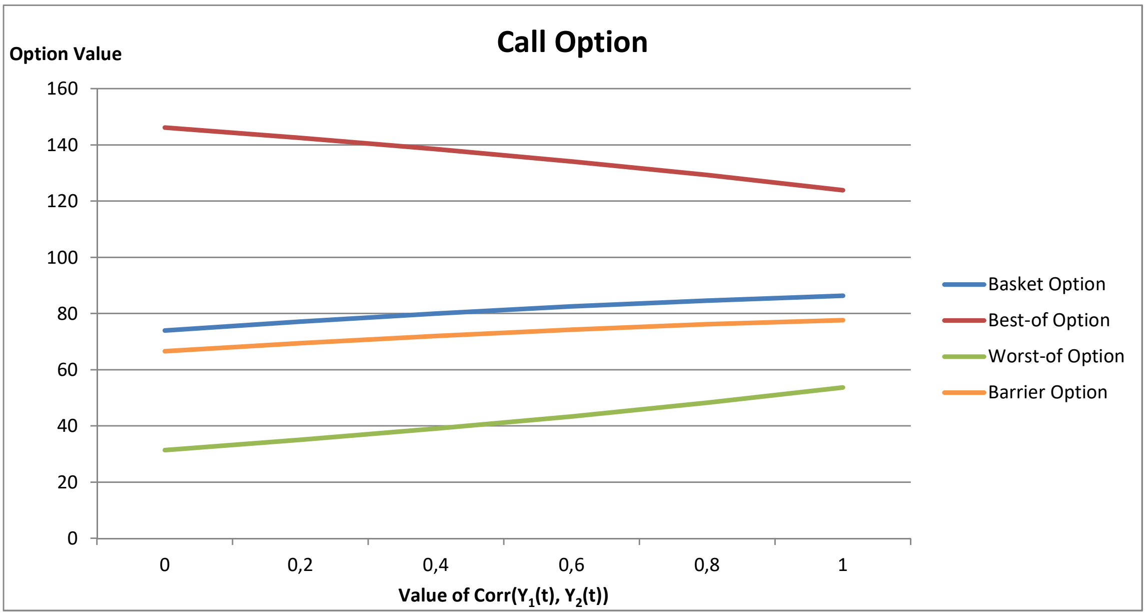


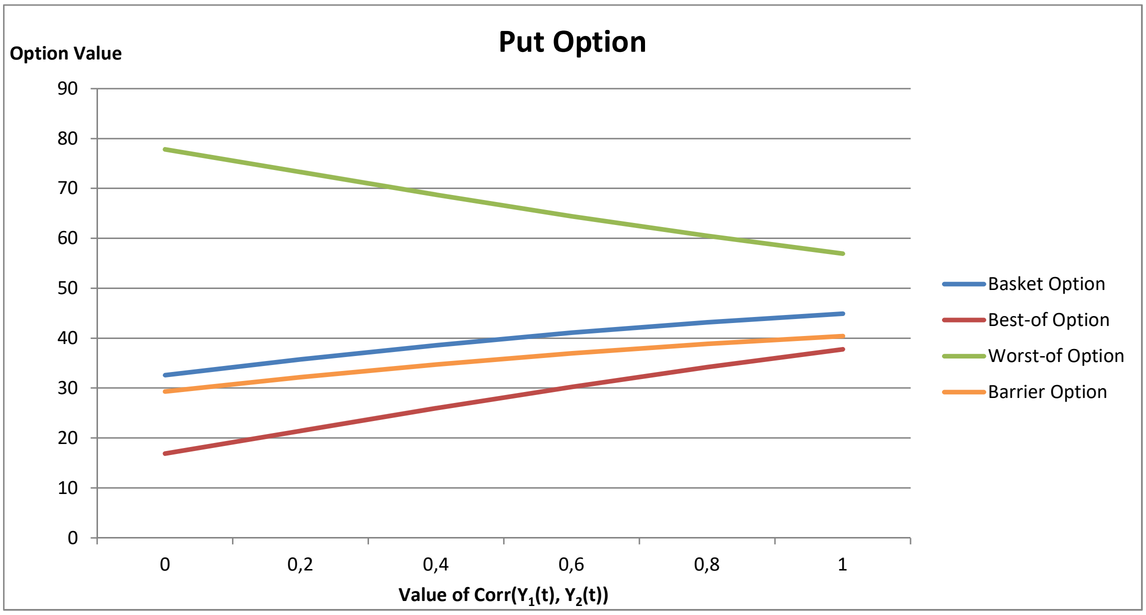
The results here are similar to the ones of Figures 2.1, 2.2 and 2.3: this is due to the fact that this model cannot capture extreme events as well as copulae. The Clayton copula (both bidirectional and unidirectional) can only capture extreme events on the left tail, which for option pricing is anyways capped at 0. Whereas the right tail is not capped but the Clayton copula cannot capture dependence on that tail. This is the reason for the similarity in the results. Whereas the Gumbel and Frank can retain all the information for the tails, because they stress dependence on higher percentiles of the joint distributions which are not capped, while they do not put any weight on the capped left tail anyways. Therefore the Clayton copula and the Ballotta-Bonfiglioli model present similar findings, while the Gumbel and Frank have symmetrical results.

2.4.4 Results of the option pricing via simulation

For comparative purposes, we also do the pricing via simulation: we simulate $Y_1(t)$ and $Y_2(t)$ with the usual marginal parameters and the same number of simulations H used previously, we apply the mean correcting martingale to risk-neutralise our simulated paths and we employ the Longstaff-Schwartz algorithm with the same number and type of basis functions as before. In Figure 2.7 we present the results for our options if we consider a positive correlation $\text{Corr}(Y_1(t), Y_2(t))$ between the two Variance Gamma processes ranging from 0 to 1.

Figure 2.7: Option prices via simulation under the Ballotta-Bonfiglioli model for a varying $\mathrm{Corr}(Y_1(t), Y_2(t))$





As we see, under the Ballotta-Bonfiglioli model, the results via simulation are consistent with the ones obtained through the finite-difference method: all differences in values arise after the third decimal point, therefore they cannot be detected visually by looking at the graphs. We carried out robustness checks to see whether with different marginal parameters for the Variance Gamma processes and with different options' characteristics the discrepancy between the results obtained via finite difference method and the ones obtained via simulation varies, but in all our tests the differences in values were stable and always after the third decimal point. In Table 2.1 we present as an example the difference in the results via finite difference method and via simulation if we keep the marginal parameters for the Variance Gamma processes and the options' characteristics used before and we only change the value of $S_1(0) = S_2(0) = K$ and the expiry date T.

Table 2.1: Difference in options' results via finite difference method and via simulation for different T and K

K/T	$T = 0.25$	$T = 0.5$	$T = 0.75$	$T = 1$
$K = 300$	0.000053	0.000035	0.000074	0.000056
$K = 305$	0.000073	0.000093	0.000048	0.000045
$K = 310$	0.000060	0.000053	0.000074	0.000035
$K = 315$	0.000046	0.000075	0.000065	0.000045
$K = 320$	0.000046	0.000095	0.000049	0.000051
$K = 325$	0.000081	0.000044	0.000093	0.000092
$K = 330$	0.000076	0.000064	0.000058	0.000087

We computed the difference in the results for each of the 8 payoffs of the 4 types of options, both in their call and put versions, and in Table 2.1 we reported the average of the 8 differences for each value of K and for each T.

In Table 2.2 we present as an example the difference in the results via finite difference method and via simulation if we keep the options' characteristics used before and we change the values for the marginal parameters for the Variance Gamma processes σ_k and α_k while we keep θ_k unvaried. As before, for simplicity, let us have the same marginal parameters for both processes, i.e. $\sigma_1 = \sigma_2$ and $\alpha_1 = \alpha_2$.

Table 2.2: Difference in options' results via finite difference method and via simulation for different σ_k and α_k

σ_k/α_k	$\alpha_k = 0.2$	$\alpha_k = 0.3$	$\alpha_k = 0.4$	$\alpha_k = 0.5$	$\alpha_k = 0.6$	$\alpha_k = 0.7$	$\alpha_k = 0.8$
$\sigma_k = 0.2$	0.000063	0.000073	0.000048	0.000073	0.000093	0.000045	0.000056
$\sigma_k = 0.3$	0.000040	0.000089	0.000045	0.000093	0.000078	0.000052	0.000072
$\sigma_k = 0.4$	0.000073	0.000094	0.000082	0.000052	0.000047	0.000073	0.000092
$\sigma_k = 0.5$	0.000071	0.000050	0.000094	0.000083	0.000092	0.000039	0.000071
$\sigma_k = 0.6$	0.000057	0.000059	0.000073	0.000091	0.000045	0.000082	0.000086
$\sigma_k = 0.7$	0.000072	0.000077	0.000083	0.000038	0.000067	0.000054	0.000059
$\sigma_k = 0.8$	0.000081	0.000056	0.000045	0.000059	0.000085	0.000071	0.000087

As before, we computed the difference in the results for each of the 8 payoffs of the 4 types of options, both in their call and put versions, and in Table 2.2 we reported the average of the 8 differences for each value of σ_k and for each value of α_k.

2.5 Concluding remarks

We therefore showed two possible ways to structure the dependence between the underlying assets in the context of American multi-asset option pricing. Lévy copulae are more precise and can describe extreme events with high accuracy, but this comes at the cost of reduced analytical tractability and much less convenient estimation and simulation procedure. The Ballotta-Bonfiglioli method on the other hand is highly flexible: the great advantage of the Ballotta-Bonfiglioli model is that, since it considers the marginal Lévy processes for the idiosyncratic components and for the common factor to be independent, one can easily rely on the work by Hirsa and Madan (2004), Fiorani (1999) and Fiorani (2004) and add as many components as needed without affecting the general structure. Fiorani has provided us with solutions for the jump component, which still hold in increasing dimensions, therefore such work can be used for any kind of option and on any dimension. Not only one can increase on the overall number of underlying assets and idiosyncratic components, it is also possible to add extra common factors to take into account of the dependence between multiple underlying assets in a more complete way, so that even pairwise dependence can be absorbed (in the case of more than two underlying assets, since with only two underlying assets one measure of dependence is enough, as mentioned in Subsection 1.2.3). It would indeed be interesting to extend the work on Lévy copulae to Vine Lévy copulae (see Grothe and Nicklas (2013)), so that in the case of more than two underlying assets one is not limited by one overall dependence, but also pairwise dependence can be captured. Another possible line of further research would be to extend the European multi-asset option pricing on a Markov-modulated switching regime by Deelstra and Simon (2017) to the American contracts. Also, as we said, in our case we only took into account positive correlation, readaptations for negative dependence would be an interesting further development.

Chapter 3

The bivariate Negative Binomial-Inverse Gaussian regression model with an application to insurance a posteriori ratemaking

3.1 Introduction

3.1.1 From univariate to multivariate regression models

Over the last few decades, univariate mixed Poisson regression models, with the Negative Binomial (NB) and Poisson-Inverse Gaussian (PIG) models being the most traditional choices, have been established by various previous studies as the appropriate statistical formalism in the a priori rakemaking process for Motor Third Party Liability (MTPL) insurance due to their efficiency for quantifying the relation between the overdispersed claim counts and the characteristics of the policyholders and their cars. Furthermore, such models can be used for deriving a posteriori ratemaking mechanisms, or Bonus-Malus Systems (BMSs), which can take into account both the a priori and a posteriori criteria, i.e. all the factors that could not be identified, measured and introduced in the a priori tariff. An excellent account of BMSs can be found in Lemaire (1995). Further references for BMSs include, among many others, Tremblay (1992), Picech (1994), Pinquet (1997), Pinquet (1998), Brouhns et al. (2003), Mert and Saykan (2005), Denuit et al. (2007), Boucher and Denuit (2008), Gómez-Déniz et al. (2008), Gómez-Déniz et al. (2014), Ni et al. (2014b), Ni et al. (2014a), Tzougas and Frangos (2014), Santi et al. (2016), Gómez-Déniz and Calderín-Ojeda (2018), Karlis et al. (2018), Tzougas et al. (2018), Tzougas et al. (2019) and Tzougas et al. (2020).

However, by adopting the univariate mixed Poisson count regression modelling approach, the actuary can only specify a separate model for different claim types. Nevertheless, it is not uncommon for an insurer to find the need in non-life insurance practice to model the positive association between claim counts of two (and/or multiple) types. In fact, various studies have reported evidence of a positive correlation between different types of claims, see, for instance, Bermúdez (2009), Bermúdez and Karlis (2011), Bermúdez and Karlis (2012), Shi and Valdez (2014) and Abdallah et al. (2016), Bermúdez and Karlis (2017) and Bermúdez et al. (2018).

As far as MTPL insurance is concerned, which refers to a person's legal liability for the bodily injury and property damage sustained by another as the result of an accident, modelling the two types of claims, which are conceivably positively correlated with each other, and their associated claim counts, is required for making the Bonus-Malus price discrimination even more fair and reasonable when the a posteriori correction is going to be calculated. Nevertheless, the Bayesian approach for calculating Bonus-Malus premia in the bivariate setting has only been addressed very recently by Bermúdez and Karlis (2017). The contribution of the latter article can be regarded as a significant improvement over prior ratemaking literature, which only focused on bivariate experience rating models that were derived based on the credibility approach.

3.1.2 The bivariate Negative Binomial-Inverse Gaussian regression model

In the present chapter, the bivariate extension of the Negative Binomial-Inverse Gaussian (NBIG) regression model, which was considered by Tzougas et al. (2019) in its univariate version, will be employed for examining the relation between the frequency of the positively correlated claims from MTPL bodily injury and property damage and the characteristics of the policyholders and their cars. Furthermore, motivated by the paper of Bermúdez and Karlis (2017), the bivariate Negative Binomial-Inverse Gaussian (BNBIG) regression model will be employed for calculating Bonus-Malus premium rates in a way which integrates a priori and a posteriori information on an individual basis. In what follows we provide a thorough discussion about how our contribution extends both the statistical and actuarial literature concerning bivariate count regression models, putting special emphasis on the probabilistic predictive modelling, computational ML estimation and practical MTPL insurance pricing perspectives.

Firstly, even if, as it was previously mentioned, a plenitude of books and scholarly articles consistent with the standard probabilistic predictive modelling and MTPL ratemaking practice have been devoted on the use of univariate mixed Poisson regression models, there is no guarantee that overdispersion and variation in claim propensity have precisely the distributional forms implied by mixed Poisson models. Moreover, due to the complexity of MTPL insurance data, their bivariate versions will not always necessarily efficiently model the relationship between MTPL bodily injury and property damage claims and a set of explanatory variables. Therefore, unless the assumption that the count data are distributed according to a particular member of the mixed Poisson family is valid, then an inappropriate imposition of the mixed Poisson model may lead to huge financial impacts for the insurance company, since, due to the economic importance of MTPL insurance[1], very accurate predictions are required by the actuary for pricing risks. Furthermore, it should be noted that alternative mixed Poisson models usually lead to Bonus-Malus premium rates which are not substantially different for policyholders with some claim experience and hence there is in principle no reason why attention should be confined to this family of models.

Thus, given the increasing interdisciplinary demand for data driven predictive models and maximum likelihood (ML) estimation methods, a very important aspect of the actuary's job is to be able to construct viable alternatives to the traditional mixed Poisson models that can capture the stylized characteristics of the data, since very accurate predictions are required for pricing, reserving, estimating future company liabilities and understanding the implications of these claims to the solvency of the company.

Mixed Negative Binomial models have thick tails and can be considered as candidate models for analysing highly overdispersed count data in numerous univariate and bivariate (and/or multivariate) domains. Nevertheless, even if the literature on mixed Poisson models is abundant, only very few mixed Negative Binomial models have been studied in depth because their log-likelihood is complicated and hence its maximisation needs a special effort. In particular, the Negative Binomial-Pareto distribution (see Shengwang et al. (1999) and Gómez-Déniz and Vázquez-Polo (2003)), the Negative Binomial-Beta regression model (see Boucher et al. (2008)), the Negative Binomial-Gamma (see Gençtürk and Yiğiter (2016)), the Negative Binomial-Lindley distribution (see Zamani and Ismail (2010) and Gómez-Déniz and Calderín-Ojeda (2017)), the Negative Binomial-Inverse Gaussian (see Gómez-Déniz et al. (2008) and Tzougas et al. (2019) who considered the cases with and without covariate information) and the Negative Binomial-Reciprocal Inverse Gaussian (see Ahmad et al. (2019)) have been considered in the univariate setting.

[1]For instance, MTPL insurance, according to the most recent report by Insurance Europe, accounted for almost one third of non-life business in the European Union, see Insurance Europe (2015).

Moreover, the literature on the bivariate (and/or multivariate) extensions of mixed Negative Binomial models is even smaller since computational complexity increases even further when considering jointly two or more count variables. In fact, the only notable exceptions so far are the articles by Gómez-Déniz et al. (2008) and Calderín-Ojeda and Gómez-Déniz (2019) who introduced the multivariate versions of the Negative Binomial-Inverse Gaussian and the Negative Binomial-Lindley distributions, considered ML estimation methods and gave a very detailed description of statistical methods connected to both models.

3.1.3 Contribution to the literature

This is the first time that the BNBIG regression model is used in a statistical or actuarial context because, due to algebraic intractability, direct maximisation of its log-likelihood is difficult and has not been addressed in the literature so far. The BNBIG regression model will be constructed based on a mixing between two marginal Negative Binomial distributions and an Inverse Gaussian distribution. At this point, we would like to call attention to the fact that the construction of bivariate (and/or multivariate) count regression models that can appropriately model overdispersed two-dimensional positively correlated count data has only focused on bivariate (and/or multivariate) extensions of the Poisson distribution and mixed Poisson models, see, for instance, Stein and Juritz (1987) and Stein et al. (1987), Kocherlakota (1988), Munkin and Trivedi (1999), Gurmu and Elder (2000) and Ghitany et al. (2012) among many others. The BNBIG model, which we consider in this study, can be regarded as a prominent candidate for modelling bivariate positively correlated count data when marginal overdispersion is observed, a situation which is quite common in the field of MTPL insurance since bodily injury and property damage claim counts often exhibit a variance that noticeably exceeds their mean.

Secondly, from a ML estimation point of view, the main contribution of the present study is that we develop an EM type algorithm that reduces the computational burden when maximising the likelihood surface of the BNBIG regression model. In particular, the EM scheme we present does not require knowledge of the joint probability mass function (jpmf) of the BNBIG model, which cannot be written in closed form, and can be implemented by taking advantage of the quintuple Poisson-Gamma-Poisson-Gamma-Inverse Gaussian mixture representation of the model. Additionally, it is worth noting that the model is suitable for application not only for modelling the positive association between the MTPL bodily injury and property damage claims but in can be immediately generalized to any vector size of positively correlated response variables. However, for large data sets with several explanatory variables and potentially higher dimensions, computational speed is a disadvantage since most of the computational costs in the case of a multidimensional response variable will come from evaluating the expectations involved at the E-Step of the algorithm, which is not tractable. In such cases, parallel computing is necessary to achieve a substantial reduction of the computing effort. In fact, due to the structure of the algorithm for the BNBIG model, the E- and M-Steps can be executed in parallel across multiple threads to take advantage of the processing power available in modern-day multicore machines.

Finally, to examine the suitability of the BNBIG model for experience rating purposes, the a posteriori, or Bonus-Malus, premium rates resulting from this model will be calculated via the net premium principle and compared to those determined by the bivariate Negative Binomial (BNB) and bivariate Poisson-Inverse Gaussian (BPIG) models, which can be regarded as natural extensions of the NB and PIG models that have been routinely used by actuaries for pricing risks in the univariate setting. The main finding is that the BNBIG model will show much less extreme a posteriori, or Bonus-Malus, premia for policyholders with some MTPL bodily injury and property damage claims experience than those produced by the BNB and BPIG models. Therefore, the work presented herein can be viewed as complementary to the articles of Shengwang et al. (1999),

Gómez-Déniz et al. (2008) and Tzougas et al. (2019) who reported similar findings regarding the comparison of the mixed Negative Binomial models, which they developed with the traditional NB and PIG models in a univariate a posteriori ratemaking context. Also, another striking difference between the BNBIG and the BNB and BPIG models, is that for a given total number of claims, the former model can enable the actuary to differentiate the premium rates based on the exact frequencies of MTPL bodily injury and property damage claims, whereas the latter two mixed Poisson models do not allow to price discriminate by taking into account the difference in the numbers of the two types of claims.
Overall, from a practical business perspective, since MTPL remains the most widely purchased non-life product in the world's markets with policyholders shopping around for the best deals, due to the aforementioned reasons, the employment of the new model is beneficial for insurance companies, since compared to the two bivariate mixed Poisson models, it can enable them to better refine their a priori risk classification and restore fairness by designing merit rating plans in accordance with the a priori ratemaking structure of the company.

3.1.4 Outline

The rest of this chapter proceeds as follows: Section 3.2 presents the derivation of the BNBIG regression model. Section 3.3 fully describes the ML estimation through the EM algorithm. Section 3.4 briefly explains the bivariate mixed Poisson models, to which the BNBIG model is being compared and defended as a suitable alternative. Section 3.5 contains an application to a data set concerning MTPL insurance bodily injury and property damage claim counts. Finally, concluding remarks can be found in Section 3.6.

3.2 Description of the BNBIG regression model

Assume that $k_{i,j}$ is the number of claims for the policyholder j, with $j = 1, ..., n$, where $i = 1, 2$ represents the MTPL bodily injury and property damage claims respectively. Furthermore, let $k_i = \sum_{j=1}^{n} k_{i,j}$ denote the total number of claims per claim type $i = 1, 2$ that have been reported to the insurance company by all the n individuals in the portfolio. Also, suppose that $\boldsymbol{x_{i,j}}$ are the vectors of individual characteristics and/or characteristics of the car related to the j-th insured person per claim type $i = 1, 2$. The two responses $k_{1,j}, k_{2,j}|\boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}$ are assumed to follow a BNBIG regression model, which can be constructed as follows.

Consider that $k_{i,j}|\boldsymbol{x_{i,j}}, \lambda_j$, per claim type $i = 1, 2$, follows an NB distribution with probability mass function (pmf)

$$P\left(k_{i,j}|\boldsymbol{x_{i,j}}, \lambda_j\right) = \frac{\Gamma\left(k_{i,j} + \sigma_i\right)}{k_{i,j}!\Gamma\left(\sigma_i\right)} \left(\frac{\lambda_j \boldsymbol{\varepsilon_{i,j}}}{\sigma_i + \lambda_j \boldsymbol{\varepsilon_{i,j}}}\right)^{k_{i,j}} \left(\frac{\sigma_i}{\sigma_i + \lambda_j \boldsymbol{\varepsilon_{i,j}}}\right)^{\sigma_i}, \tag{3.1}$$

with $k_{i,j} = 0, 1, 2, 3, ...$, $\lambda_j > 0$, $\sigma_i > 0$, where $\boldsymbol{\varepsilon_{i,j}} = \exp\left(\boldsymbol{x_{i,j}^T \beta_i}\right)$ and where $\boldsymbol{\beta_i}$ are the two vectors of the regression coefficients for the two types of claims $i = 1, 2$. The mean and the variance of $k_{i,j}|\boldsymbol{x_{i,j}}, \lambda_j$ are given by

$$\mathbb{E}\left(k_{i,j}|\boldsymbol{x_{i,j}}, \lambda_j\right) = \boldsymbol{\varepsilon_{i,j}}\lambda_j$$

and

$$\mathrm{Var}\left(k_{i,j}|\boldsymbol{x_{i,j}}, \lambda_j\right) = \boldsymbol{\varepsilon_{i,j}}\lambda_j \left[1 + \frac{\boldsymbol{\varepsilon_{i,j}}\lambda_j}{\sigma_i}\right].$$

It is worth noting that the scale parameter σ_i controls the responsiveness of overdispersion to the mean number of claims, with the degree of overdispersion decreasing when σ_i increases per claim type $i = 1, 2$. Note also that, in the limit, when σ_i approaches infinity, $P\left(k_{i,j}|\boldsymbol{x_{i,j}}, \lambda_j\right)$ tends to the Poisson distribution with mean equal to $\boldsymbol{\varepsilon_{i,j}}\lambda_j$ per claim type $i = 1, 2$.

Let us now assume that λ_j are independent and identically distributed (i.i.d) random variables from an Inverse Gaussian (IG) distribution with probability density function (pdf)

$$f(\lambda_j; \gamma) = \frac{\gamma}{\sqrt{2\pi}} \exp\left(\gamma^2\right) \lambda_j^{-\frac{3}{2}} \exp\left[-\frac{1}{2}\left(\frac{\gamma^2}{\lambda_j} + \gamma^2 \lambda_j\right)\right], \tag{3.2}$$

with $\gamma > 0$, mean $\mathbb{E}(\lambda_j) = 1$ and variance $\text{Var}(\lambda_j) = \frac{1}{\gamma^2}$. The IG prior, or mixing, distribution given by Eq. (3.2) has to have unit mean, in order for the model to be estimable, otherwise identifiability[2] issues might arise. Because of this restriction, it follows that the overdispersion linked to the simple Exponential distribution is $\frac{1}{\gamma^2}$ and hence the IG will reduce to the Exponential if γ tends to infinity. For more information about the IG distribution, which is a special case of the Generalized Inverse Gaussian (GIG) distribution, the interested reader can refer to Jørgensen (1982). Note also that several other parametrisations of the IG can be found in Seshadri (1993).

Under the assumptions in Eqs. (3.1 and 3.2), the joint probability mass function (jpmf) of the BNBIG distribution is [3]

$$P(k_{1,j}, k_{2,j} | \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}) = \int_0^\infty \prod_{i=1}^{2} P(k_{i,j} | \boldsymbol{x_{i,j}}, \lambda_j) f(\lambda_j; \gamma) \, d\lambda_j. \tag{3.3}$$

[2]The unit mean requirement for the mixing density is essential for constructing mixture models with regression structures, otherwise identifiability issues can ensue, see, for example, Karlis (2001), Rigby et al. (2008), Barreto-Souza and Simas (2016), Ghitany et al. (2012) and Tzougas (2020) among many others. Furthermore, in order to show that no particular identifiability problem exists from a practical point of view, we checked with many initial values for all the parameters to examine whether the EM algorithm was trapped with different solutions. This did not happen, and for all cases the algorithm converged to the same solution.

[3]Note that, due to its quintuple mixture decomposition in Section 3.3, the jpmf of the BNBIG model can be written in the form of the jpmf of the bivariate Poisson-Gamma-Inverse Gaussian distribution. Thus, the model has all the desirable theoretical properties of mixed Poisson models, see for instance Barndorff-Nielsen (1965), Yakowitz and Spragins (1968), Tallis (1969), Teicher (1963), Ord (1972), Xekalaki (1981), Al-Hussaini and Ahmad (1981), Xekalaki and Panaretos (1983), Lynch (1988), Lindsay and Roeder (1993), Willmot (1990) and Sapatinas (1995). Note also that Gómez-Déniz et al. (2008) gave an excellent account of statistical methods connected to both the univariate and multivariate versions of the NBIG distribution. The BNBIG distribution as in Eq. (3.3) may be distinguished from the one by Gómez-Déniz et al. (2008) as the latter does not allow to include covariate information.

The last integral cannot be solved in closed form but can be calculated through numerical integration.

Furthermore, three important properties, in the context of MTPL insurance, associated with the model are given below.

1. The marginal distribution of $k_{i,j}|\boldsymbol{x_{i,j}}$, with $i = 1, 2$, is an NBIG distribution. Also, using the laws of total expectation and total variance and the moments of the NB distribution, one can find that the mean and the variance of $k_{i,j}|\boldsymbol{x_{i,j}}$ are given by

$$\mathbb{E}\left(k_{i,j}|\boldsymbol{x_{i,j}}\right) = \mathbb{E}_{\lambda_j}\left[\mathbb{E}\left(k_{i,j}|\boldsymbol{x_{i,j}}, \lambda_j\right)\right] = \boldsymbol{\varepsilon_{i,j}}\mathbb{E}_{\lambda_j}\left[\lambda_j\right] = \boldsymbol{\varepsilon_{i,j}} \tag{3.4}$$

   and

$$\begin{aligned} \mathrm{Var}\left(k_{i,j}|\boldsymbol{x_{i,j}}\right) &= \mathbb{E}_{\lambda_j}\left[\mathrm{Var}\left(k_{i,j}|\boldsymbol{x_{i,j}}, \lambda_j\right)\right] + \mathrm{Var}_{\lambda_j}\left[\mathbb{E}\left(k_{i,j}|\boldsymbol{x_{i,j}}, \lambda_j\right)\right] \\ &= \boldsymbol{\varepsilon_{i,j}}\left[1 + \boldsymbol{\varepsilon_{i,j}}\left(\frac{1+\sigma_i+\gamma^2}{\sigma_i\gamma^2}\right)\right]. \end{aligned} \tag{3.5}$$

2. Based on the laws of total variance and total covariance, we can see that the covariance (Cov) and correlation (Corr) between $k_{1,j}$ and $k_{2,j}$ are given by

$$\mathrm{Cov}(k_{1,j}, k_{2,j}|\boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}) = \left(\frac{1}{\gamma^2}\right)\boldsymbol{\varepsilon_{1,j}}\boldsymbol{\varepsilon_{2,j}} \tag{3.6}$$

   and

$$\begin{aligned} &\mathrm{Corr}(k_{1,j}, k_{2,j}|\boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}) \\ &= \frac{\frac{1}{\gamma^2}\boldsymbol{\varepsilon_{1,j}}\boldsymbol{\varepsilon_{2,j}}}{\sqrt{\left[\boldsymbol{\varepsilon_{1,j}} + \boldsymbol{\varepsilon_{1,j}^2}\left(\frac{1+\sigma_1+\gamma^2}{\sigma_1\gamma^2}\right)\right]\cdot\left[\boldsymbol{\varepsilon_{2,j}} + \boldsymbol{\varepsilon_{2,j}^2}\left(\frac{1+\sigma_2+\gamma^2}{\sigma_2\gamma^2}\right)\right]}}. \end{aligned} \tag{3.7}$$

3. The generalized variance ratio (GVR) between the BNBIG model, as defined in

Eq. (3.3), and a simple NB model, i.e. $y_{i,j} \sim \text{NB}\left(\sigma_i, \frac{\boldsymbol{\varepsilon_{i,j}}}{\sigma_i + \boldsymbol{\varepsilon_{i,j}}}\right)$ is given by

$$\text{GVR}\left(k_{1,j}, k_{2,j} | \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}\right) = \frac{\sum_{i=1}^{2} \text{Var}\left(k_{i,j} | \boldsymbol{x_{i,j}}\right) + 2\text{Cov}\left(k_{1,j}, k_{2,j} | \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}\right)}{\sum_{i=1}^{2} \text{Var}\left(y_{i,j} | \boldsymbol{x_{i,j}}\right)}$$

$$= 1 + \frac{1}{\gamma^2} \frac{\left[\frac{\boldsymbol{\varepsilon}_{\boldsymbol{1,j}}^2}{\sigma_1} + \frac{\boldsymbol{\varepsilon}_{\boldsymbol{2,j}}^2}{\sigma_2} + \left(\boldsymbol{\varepsilon_{1,j}} + \boldsymbol{\varepsilon_{2,j}}\right)^2\right]}{\left[\frac{\boldsymbol{\varepsilon}_{\boldsymbol{1,j}}^2}{\sigma_1} + \frac{\boldsymbol{\varepsilon}_{\boldsymbol{2,j}}^2}{\sigma_2} + \boldsymbol{\varepsilon_{1,j}} + \boldsymbol{\varepsilon_{2,j}}\right]}. \tag{3.8}$$

The BNBIG allows for the positive correlation between the bodily injury and property damage claims since $\text{Corr}\left(k_{1,j}, k_{2,j} | \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}\right) > 0$, see Eq. (3.7), and it accommodates the bivariate overdispersion in the data since GVR > 1. Also, the GVR increases as the variance of the IG distribution increases, see Eq. (3.8).

Thus, the BNBIG model is ideally suited for modelling two-dimensional MTPL insurance data since, as was previously mentioned, positive correlation and overdispersion are the two phenomena that have been most commonly reported in the pricing literature in the bivariate setting.

Finally, it should be noted that there are many factors which cannot be directly observed by the actuary but can simultaneously affect the joint dynamics of MTPL bodily injury and property damage claims, leading to extra variation occurring in their associated claim counts. Thus, the choice of the mixing density, which measures the level of unobservable risk associated with each of the policies, is crucial since a potential distribution misspecification can result in biased and unreliable parameter estimates, which, in turn, can have an impact on how insurers price the policy, leading to financial implications for the company, since, if the punishment of all policyholders is not justified on a sound risk measuring basis, then they may switch to competing companies.

In this study, motivated by the characteristics of the MTPL insurance data which we will analyse in Section 3.5, we proposed the use of the IG distribution as a suitable mixing density. In fact, as we will observe in Section 3.5, the resulting BNBIG model will provide better fitting performances compared to bivariate mixed Poisson benchmarks which can be derived in a similar way. However, as it can be clearly understood, different mixing distributions might be more appropriate for different data sets. The EM type algorithm which we will present in Section 3.3 has the sufficient flexibility to estimate alternative bivariate (and/or multivariate) Negative Binomial mixture models stemming from several other continuous and at least twice differentiable mixing distributions with a unit mean in order to avoid identifiability issues.

3.3 The EM algorithm for ML estimation of the BNBIG regression model

Let $(k_{1,j}, k_{2,j}; \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}})$ be a sample of observations, $j = 1, ..., n$, where the responses $k_{i,j}$ are the number of claims for the policyholder j and where $\boldsymbol{x_{i,j}}$ are the vectors of covariate information per claim type $i = 1, 2$. Considering that the data are produced according to the BNBIG model, its log-likelihood can be expressed as

$$l(\phi) = \sum_{j=1}^{n} \log \left(P\left(k_{1,j}, k_{2,j} | \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}\right)\right), \tag{3.9}$$

where $\phi = (\gamma, \sigma_1, \sigma_2, \boldsymbol{\beta_1}, \boldsymbol{\beta_2})$ is the vector of the parameters and where $P\left(k_{1,j}, k_{2,j} | \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}\right)$ is the jpmf of the BNBIG distribution which is given by Eq. (3.3). The log-likelihood given by Eq. (3.9) does not exist in closed form and hence ϕ cannot be estimated via traditional numerical maximisation methods. In such cases, it is necessary to resort to the EM algorithm (see Dempster et al. (1977) and McLachlan and Krishnan (2007)). In particular, if one augments the unobserved data, denoted by λ_j herein, to $(k_{1,j}, k_{2,j}; \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}})$ for $j = 1, ..., n$, then the complete data log-likelihood factorises into two parts:

$$l_c(\phi) = \sum_{j=1}^{n}\sum_{i=1}^{2}\log\left(P\left(k_{i,j}|\boldsymbol{x_{i,j}},\lambda_j\right)\right) + \sum_{j=1}^{n}\log\left(f\left(\lambda_j;\gamma\right)\right), \tag{3.10}$$

where $P\left(k_{i,j}|\boldsymbol{x_{i,j}},\lambda_j\right)$ is the pmf of each of the two NB distributions, which are given by Eq. (3.1) for $i = 1, 2$, and where $f(\lambda_j;\gamma)$ represents the pdf of the IG mixing distribution, which is given by Eq. (3.2). Direct maximisation of the complete data log-likelihood, as given by Eq. (3.10), with respect to ϕ is cumbersome. Fortunately, its ML estimation can be easily achieved via the EM algorithm, if one takes advantage of the following quintuple mixture derivation of the model. In particular, the number of claims $k_{1,j}$ is distributed as:

$$\begin{aligned} k_{1,j} &\sim \text{Poisson}\left(\boldsymbol{\vartheta_{1,j}}\right) \\ \text{with } \boldsymbol{\vartheta_{1,j}} &\sim \text{Gamma}\left(\sigma_1, \frac{\sigma_1}{\lambda_j\boldsymbol{\varepsilon_{1,j}}}\right), \end{aligned} \tag{3.11}$$

the number of claims $k_{2,j}$ is distributed as:

$$\begin{aligned} k_{2,j} &\sim \text{Poisson}\left(\boldsymbol{\vartheta_{2,j}}\right) \\ \text{with } \boldsymbol{\vartheta_{2,j}} &\sim \text{Gamma}\left(\sigma_2, \frac{\sigma_2}{\lambda_j\boldsymbol{\varepsilon_{2,j}}}\right) \end{aligned} \tag{3.12}$$

and both $k_{1,j}$ and $k_{2,j}$ share the same unobserved heterogeneity variable λ_j which is distributed as

$$\lambda_j \sim \text{Inverse Gaussian}\left(\gamma\right). \tag{3.13}$$

Also, let us denote

$$P\left(k_{i,j}|\boldsymbol{\vartheta_{i,j}}\right) = e^{-\boldsymbol{\vartheta_{i,j}}}\frac{\boldsymbol{\vartheta_{i,j}}^{k_{i,j}}}{k_{i,j}!},$$

for $i = 1, 2$, to be the two pmfs of $k_{1,j}$ and $k_{2,j}$ respectively, and

$$g_1\left(\boldsymbol{\vartheta_{1,j}}|\boldsymbol{\beta_1},\sigma_1,\lambda_j\right) = \boldsymbol{\vartheta}_{\boldsymbol{1,j}}^{\sigma_1-1}\exp\left(-\frac{\sigma_1}{\lambda_j\boldsymbol{\varepsilon_{1,j}}}\boldsymbol{\vartheta_{1,j}}\right)\left(\frac{\sigma_1}{\lambda_j\boldsymbol{\varepsilon_{1,j}}}\right)^{\sigma_1}/\Gamma\left(\sigma_1\right)$$

and

$$g_2\left(\boldsymbol{\vartheta_{2,j}}|\boldsymbol{\beta_2},\sigma_2,\lambda_j\right) = \boldsymbol{\vartheta}_{\boldsymbol{2,j}}^{\sigma_2-1}\exp\left(-\frac{\sigma_2}{\lambda_j\boldsymbol{\varepsilon_{2,j}}}\boldsymbol{\vartheta_{2,j}}\right)\left(\frac{\sigma_2}{\lambda_j\boldsymbol{\varepsilon_{2,j}}}\right)^{\sigma_2}/\Gamma\left(\sigma_2\right)$$

to be the pdfs of the two Gamma distributions.

Then, using the mixture representation in Eqs. (3.11, 3.12 and 3.13) the complete data log-likelihood is proportional to

$$\begin{aligned} l_c\left(\phi\right) \quad \propto \quad & \sum_{j=1}^{n}\log\left(g_1\left(\boldsymbol{\vartheta_{1,j}}|\boldsymbol{\beta_1},\sigma_1,\lambda_j\right)\right) + \\ & \sum_{j=1}^{n}\log\left(g_2\left(\boldsymbol{\vartheta_{2,j}}|\boldsymbol{\beta_2},\sigma_2,\lambda_j\right)\right) + \sum_{j=1}^{n}\log\left(f\left(\lambda_j;\gamma\right)\right). \end{aligned} \tag{3.14}$$

The regression coefficients $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$ and the parameters σ_1 and σ_2 are involved in the first and second terms and the parameter γ is involved in the third term of Eq. (3.14), which correspond to the log-likelihoods of the two Gamma components and the Inverse Gaussian component respectively.

Therefore, the Q-function, which is the conditional expectation of the complete data log-likelihood, is proportional to

$$\begin{aligned} Q\left(\phi;\phi_{(r)}\right) &\equiv \mathbb{E}_{\lambda_j}\left(l_c\left(\phi\right)|k_{1,j},k_{2,j},\boldsymbol{x_{1,j}},\boldsymbol{x_{2,j}},\phi_{(r)}\right) \\ &\propto \mathbb{E}_{\lambda_j}\left[\sum_{j=1}^{n}\log\left(g_1\left(\boldsymbol{\vartheta_{1,j}}|\boldsymbol{\beta_{1,(r)}},\sigma_{1,(r)},\lambda_j\right)\right)\right] \\ &\quad + \mathbb{E}_{\lambda_j}\left[\sum_{j=1}^{n}\log\left(g_2\left(\boldsymbol{\vartheta_{2,j}}|\boldsymbol{\beta_{2,(r)}},\sigma_{2,(r)},\lambda_j\right)\right)\right] \\ &\quad + \mathbb{E}_{\lambda_j}\left[\sum_{j=1}^{n}\log\left(f\left(\lambda_j;\gamma_{(r)}\right)\right)\right], \end{aligned} \tag{3.15}$$

where $\phi_{(r)} = \left(\gamma_{(r)}, \sigma_{1,(r)}, \sigma_{2,(r)}, \boldsymbol{\beta_{1,(r)}}, \boldsymbol{\beta_{2,(r)}}\right)$ is the estimate of ϕ in the E-Step of our EM algorithm.

In what follows, some functions of the unobserved data λ_j which are involved in Eq. (3.15) will be calculated for implementing the E-Step of the algorithm, while the M-Step involves maximising the Q-function with respect to ϕ. Also, the following posterior distributions will be needed in the E-Step of the EM algorithm:

$$\boldsymbol{\vartheta_{1,j}}|k_{1,j}, \boldsymbol{x_{1,j}}, \sigma_1, \boldsymbol{\beta_1} \sim \text{Gamma}\left(k_{1,j} + \sigma_1, \frac{\sigma_1}{\lambda_j \boldsymbol{\varepsilon_{1,j}}} + 1\right) \tag{3.16}$$

and

$$\boldsymbol{\vartheta_{2,j}}|k_{2,j}, \boldsymbol{x_{2,j}}, \sigma_2, \boldsymbol{\beta_2} \sim \text{Gamma}\left(k_{2,j} + \sigma_2, \frac{\sigma_2}{\lambda_j \boldsymbol{\varepsilon_{2,j}}} + 1\right). \tag{3.17}$$

The EM algorithm can now be formally described as follows.

E-Step:

- Compute the pseudo-values for $j = 1, 2, ..., n$, using the parameters' values after the r-th iteration

$$w_{1,j} = \mathbb{E}_{\lambda_j}\left(\lambda_j | k_{1,j}, k_{2,j}, \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \phi_{(r)}\right) = \frac{\int_0^\infty \lambda_j P\left(k_{1,j}, k_{2,j} | \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \lambda_j\right) f\left(\lambda_j; \gamma\right) d\lambda_j}{\int_0^\infty P\left(k_{1,j}, k_{2,j} | \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \lambda_j\right) f\left(\lambda_j; \gamma\right) d\lambda_j}, \tag{3.18}$$

$$w_{2,j} = \mathbb{E}_{\lambda_j}\left(\frac{1}{\lambda_j} | k_{1,j}, k_{2,j}, \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \phi_{(r)}\right) = \frac{\int_0^\infty \frac{1}{\lambda_j} P\left(k_{1,j}, k_{2,j} | \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \lambda_j\right) f\left(\lambda_j; \gamma\right) d\lambda_j}{\int_0^\infty P\left(k_{1,j}, k_{2,j} | \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \lambda_j\right) f\left(\lambda_j; \gamma\right) d\lambda_j}, \tag{3.19}$$

$$\begin{aligned} w_{3,j} &= \mathbb{E}_{\lambda_j}\left(\frac{1}{\left(\lambda_j \boldsymbol{\varepsilon_{1,j}} + \sigma_{1,(r)}\right)} | k_{1,j}, k_{2,j}, \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \phi_{(r)}\right) \\ &= \frac{\int_0^\infty \frac{1}{\left(\lambda_j \boldsymbol{\varepsilon_{1,j}} + \boldsymbol{\sigma_{1,(r)}}\right)} P\left(k_{1,j}, k_{2,j} | \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \lambda_j\right) f\left(\lambda_j; \gamma\right) d\lambda_j}{\int_0^\infty P\left(k_{1,j}, k_{2,j} | \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \lambda_j\right) f\left(\lambda_j; \gamma\right) d\lambda_j}, \end{aligned} \tag{3.20}$$

$$w_{4,j} = \mathbb{E}_{\lambda_j}\left(\log\left(\lambda_j \boldsymbol{\varepsilon_{1,j}} + \sigma_{1,(r)}\right)|k_{1,j}, k_{2,j}, \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \phi_{(r)}\right)$$

$$= \frac{\int\limits_0^\infty \log\left(\lambda_j \boldsymbol{\varepsilon_{1,j}} + \sigma_{1,(r)}\right) P\left(k_{1,j}, k_{2,j}|\boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \lambda_j\right) f\left(\lambda_j;\gamma\right) d\lambda_j}{\int\limits_0^\infty P\left(k_{1,j}, k_{2,j}|\boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \lambda_j\right) f\left(\lambda_j;\gamma\right) d\lambda_j}, \tag{3.21}$$

$$w_{5,j} = \mathbb{E}_{\lambda_j}\left(\frac{1}{\left(\lambda_j \boldsymbol{\varepsilon_{2,j}} + \sigma_{2,(r)}\right)}|k_{1,j}, k_{2,j}, \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \phi_{(r)}\right)$$

$$= \frac{\int\limits_0^\infty \frac{1}{\left(\lambda_j \boldsymbol{\varepsilon_{2,j}} + \sigma_{2,(r)}\right)} P\left(k_{1,j}, k_{2,j}|\boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \lambda_j\right) f\left(\lambda_j;\gamma\right) d\lambda_j}{\int\limits_0^\infty P\left(k_{1,j}, k_{2,j}|\boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \lambda_j\right) f\left(\lambda_j;\gamma\right) d\lambda_j}, \tag{3.22}$$

$$w_{6,j} = \mathbb{E}_{\lambda_j}\left(\log\left(\lambda_j \boldsymbol{\varepsilon_{2,j}} + \sigma_{2,(r)}\right)|k_{1,j}, k_{2,j}, \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \phi_{(r)}\right)$$

$$= \frac{\int\limits_0^\infty \log\left(\lambda_j \boldsymbol{\varepsilon_{2,j}} + \sigma_{2,(r)}\right) P\left(k_{1,j}, k_{2,j}|\boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \lambda_j\right) f\left(\lambda_j;\gamma\right) d\lambda_j}{\int\limits_0^\infty P\left(k_{1,j}, k_{2,j}|\boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \lambda_j\right) f\left(\lambda_j;\gamma\right) d\lambda_j}. \tag{3.23}$$

- Using Eqs. (3.16, 3.17, 3.20, 3.21, 3.22 and 3.23) we obtain that

$$s_{1,j} = \mathbb{E}_{\lambda_j}\left[\mathbb{E}_{\boldsymbol{\vartheta_{1,j}}}\left(\frac{\boldsymbol{\vartheta_{1,j}}}{\lambda_j \boldsymbol{\varepsilon_{1,j}}}|k_{1,j}, k_{2,j}, \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \phi_{(r)}\right)\right] = \left(k_{1,j} + \sigma_{1,(r)}\right) w_{3,j},$$

$$s_{2,j} = \mathbb{E}_{\lambda_j}\left[\mathbb{E}_{\boldsymbol{\vartheta_{1,j}}}\left(\log\left(\frac{\boldsymbol{\vartheta_{1,j}}}{\lambda_j \boldsymbol{\varepsilon_{1,j}}}\right)|k_{1,j}, k_{2,j}, \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \phi_{(r)}\right)\right] = \Psi\left(k_{1,j} + \sigma_{1,(r)}\right) - w_{4,j}$$

and

$$s_{3,j} = \mathbb{E}_{\lambda_j}\left[\mathbb{E}_{\boldsymbol{\vartheta_{2,j}}}\left(\frac{\boldsymbol{\vartheta_{2,j}}}{\lambda_j \boldsymbol{\varepsilon_{2,j}}}|k_{1,j}, k_{2,j}, \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \phi_{(r)}\right)\right] = \left(k_{2,j} + \sigma_{2,(r)}\right) w_{5,j},$$

$$s_{4,j} = \mathbb{E}_{\lambda_j}\left[\mathbb{E}_{\boldsymbol{\vartheta_{2,j}}}\left(\log\left(\frac{\boldsymbol{\vartheta_{2,j}}}{\lambda_j \boldsymbol{\varepsilon_{2,j}}}\right)|k_{1,j}, k_{2,j}, \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \phi_{(r)}\right)\right] = \Psi\left(k_{2,j} + \sigma_{2,(r)}\right) - w_{6,j},$$

with $\Psi(\cdot)$ representing the digamma function. Clearly the expectations given by Eqs. (3.18, 3.19, 3.20, 3.21, 3.22 and 3.23) cannot be written in closed form and thus they need to be evaluated numerically. Alternatively, a Monte Carlo approach can also be used based on a rejection algorithm. The latter case leads to variants of the EM algorithm such as the Monte Carlo EM algorithm, see, for example, Booth and Hobert (1999), Booth et al. (2001) and Karlis (2005)).

M-Step:

In the M-Step, the pseudo-values from the E-Step can be used to maximise the Q-function.

- Firstly, the Newton-Raphson algorithm is employed to obtain ML estimates of the two vectors of regression coefficients $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$. Differentiating $Q\left(\phi;\phi_{(r)}\right)$ with respect to $\boldsymbol{\beta_1}$, we find:

$$g_1\left(\boldsymbol{\beta_1}\right) = \mathbb{E}_{\lambda_j}\left(\frac{\partial l_c}{\partial \boldsymbol{\beta_1}} | k_{1,j}, k_{2,j}, \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \phi\right) = \sigma_1 \sum_{j=1}^{n} \left(s_{1,i} - 1\right) \boldsymbol{x_{1,j}}$$

  and

$$\begin{aligned} G_1\left(\boldsymbol{\beta_1}\right) &= \mathbb{E}_{\lambda_j}\left(\frac{\partial^2 l_c}{\partial \boldsymbol{\beta_1} \partial \boldsymbol{\beta_1^T}} | k_{1,j}, k_{2,j}, \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \phi\right) \\ &= -\sigma_1 \sum_{j=1}^{n} s_{1,i} \boldsymbol{x_{1,j}} \boldsymbol{x_{1,j}^T} = -\sigma_1 \mathbf{X_1^T} \mathbf{W_1} \mathbf{X_1}, \end{aligned}$$

  where $\mathbf{W_1} = \text{diag}\{s_{1,i}\}$.
  Therefore, the Newton-Raphson iterative procedure for obtaining ML estimates of the elements of β_1 goes as follows:

$$\boldsymbol{\beta_{1,(r+1)}} \equiv \boldsymbol{\beta_{1,(r)}} - \left[G_1\left(\boldsymbol{\beta_{1,(r)}}\right)\right]^{-1} g_1\left(\boldsymbol{\beta_{1,(r)}}\right).$$

Then, following the same procedure for $\boldsymbol{\beta_2}$, we obtain:

$$g_2\left(\boldsymbol{\beta_2}\right) = \mathbb{E}_{\lambda_j}\left(\frac{\partial l_c}{\partial \boldsymbol{\beta_2}}|k_{1,j}, k_{2,j}, \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \phi\right) = \sigma_2 \sum_{j=1}^{n}\left(s_{3,i} - 1\right)\boldsymbol{x_{2,j}}$$

and

$$\begin{aligned} G_2\left(\boldsymbol{\beta_2}\right) &= \mathbb{E}_{\lambda_j}\left(\frac{\partial^2 l_c}{\partial \boldsymbol{\beta_2} \partial \boldsymbol{\beta_2}^T}|k_{1,j}, k_{2,j}, \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \phi\right) \\ &= -\sigma_2 \sum_{j=1}^{n} s_{3,i} \boldsymbol{x_{2,j}}, \boldsymbol{x_{2,j}}^T = -\sigma_2 \mathbf{X_2}^T \mathbf{W_2} \mathbf{X_2}, \end{aligned}$$

where $\mathbf{W_2} = \text{diag}\{s_{3,i}\}$.

Therefore the iterated $\boldsymbol{\beta_2}$ is:

$$\boldsymbol{\beta_{2,(r+1)}} \equiv \boldsymbol{\beta_{2,(r)}} - \left[G_2\left(\boldsymbol{\beta_{2,(r)}}\right)\right]^{-1} g_2\left(\boldsymbol{\beta_{2,(r)}}\right).$$

- Secondly, the one step ahead Newton iteration is used twice for updating σ_1 and σ_2:

$$\sigma_{1,(r+1)} = \sigma_{1,(r)} - \frac{\Psi\left(\sigma_{1,(r)}\right) + \bar{s}_1 - \bar{s}_2 - \log\left(\sigma_{1,(r)}\right) - 1}{\Psi_3\left(\sigma_{1,(r)}\right) - \frac{1}{\sigma_{1,(r)}}},$$

$$\sigma_{2,(r+1)} = \sigma_{2,(r)} - \frac{\Psi\left(\sigma_{2,(r)}\right) + \bar{s}_3 - \bar{s}_4 - \log\left(\sigma_{2,(r)}\right) - 1}{\Psi_3\left(\sigma_{2,(r)}\right) - \frac{1}{\sigma_{2,(r)}}},$$

where $\Psi_3\left(\cdot\right)$ denotes the trigamma function.

- Finally, update γ with

$$\gamma_{(r+1)} = \left(\bar{w}_1 + \bar{w}_2 - 2\right)^{-\frac{1}{2}}.$$

- Note also that if the regression components of the model for the two responses $k_{1,j}$ and $k_{2,j}$ are limited to the constants $\beta_{1,0}$ and $\beta_{2,0}$, then we have that $\mathbb{E}\left(k_{1,j}|\boldsymbol{x_{1,j}}\right) = \exp\left(\beta_{1,0}\right) = \mu_1$ and $\mathbb{E}\left(k_{2,j}|\boldsymbol{x_{2,j}}\right) = \exp\left(\beta_{2,0}\right) = \mu_2$; and hence the ML estimation for the bivariate distribution, i.e. without regression components, can be computed via the EM type algorithm.

3.4 The BNB and BPIG regression models

In our numerical illustration the a posteriori, or Bonus-Malus, premium rates resulting from the BNB and BPIG models will be compared to those determined by the BNBIG model. Therefore, in this section we give some rudimentary facts concerning the BNB and the BPIG models.

Consider that $k_{i,j} = 0, 1, 2, 3, ...$ are the number of bodily injury and the number of property damage claims, when $i = 1, 2$ respectively, for the policyholder j, with $j = 1, ..., n$ and suppose that $\boldsymbol{x_{i,j}}$ are the vectors of individual characteristics and/or characteristics of the car related to the j-th insured person per claim type $i = 1, 2$. Also, let $\boldsymbol{\varepsilon_{i,j}} = \exp\left(\boldsymbol{x}_{\boldsymbol{i,j}}^T \boldsymbol{\beta_i}\right)$, where $\boldsymbol{\beta_i}$ are the two vectors of the regression coefficients for $i = 1, 2$.

- The jpmf of the BNB model is given by

$$P\left(k_{1,j}, k_{2,j} | \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}\right) = \frac{\Gamma\left(\sum_{i=1}^{2} k_{i,j} + \gamma\right)}{\Gamma(\gamma) \prod_{i=1}^{2} k_{i,j}!} \frac{\gamma^{\gamma} \prod_{i=1}^{2} \boldsymbol{\varepsilon_{i,j}}^{k_{i,j}}}{(\gamma + \sum\limits_{i=1}^{2} \boldsymbol{\varepsilon_{i,j}})^{\gamma + \sum_{i=1}^{2} k_{i,j}}}, \tag{3.24}$$

  for $\gamma > 0$.

- The jpmf of the BPIG is given by

$$P\left(k_{1,j}, k_{2,j} | \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}\right) = \frac{2\gamma e^{\gamma^2}}{\sqrt{2\pi}} K_{\sum_{i=1}^{2} k_{i,j} - \frac{1}{2}}(\gamma\omega) \left(\frac{\gamma}{\Delta}\right)^{\sum_{i=1}^{2} k_{i,j} - \frac{1}{2}} \prod_{i=1}^{2} \frac{\boldsymbol{\varepsilon}_{\boldsymbol{i,j}}^{k_{i,j}}}{k_{i,j}!}, \tag{3.25}$$

  where $\gamma > 0$, $\omega = \sqrt{\gamma^2 + 2\sum_{i=1}^{2} \boldsymbol{\varepsilon_{i,j}}}$ and where $K_r(\cdot)$ denotes the modified Bessel function of the third kind of order r.

Similarly to the proposed model, the BNB and BPIG are plausible models for overdispersed two-dimensional positively correlated MTPL claim count data. For more details about the BNB and BPIG models, the interested reader can refer to Stein and Juritz (1987), Stein et al. (1987), Kocherlakota (1988), Munkin and Trivedi (1999), Gurmu and Elder (2000) and Ghitany et al. (2012).

3.5 Numerical illustration

The data were kindly provided by a major European insurance company and concern a MTPL insurance portfolio which was observed during the year 2017.

We model two types of claims and their associated claim counts are recorded as $k_{1,j}$ and $k_{2,j}$, for the policyholder j, with $j = 1, ..., n$, which, as was previously mentioned, represent MTPL bodily injury and property damage claims respectively. The sample comprised insured parties with complete records; i.e., with the availability of all the explanatory variables which affect both $k_{1,j}$ and $k_{2,j}$. There were $n = 5186$ observations that met our criteria. Additionally, an exploratory analysis was carried out in order to adequately select the subset of explanatory variables with the highest predictive power for both $k_{1,j}$ and $k_{2,j}$. Additionally, in light of the heterogeneity that exists within the portfolio, we grouped the levels of each explanatory variable with respect to similar risk profiles with regard to the MTPL bodily injury and property damage claim frequencies. This is necessary as it will enable us to achieve ratemaking accuracy and balance homogeneity and sufficiency of the volume of data in each cell in order to provide credible patterns. We therefore started with a data set containing ten explanatory variables: the age of the driver, the brand of the vehicle, the car cubism, the type/price of policy, the horsepower of the car, the insurance duration, the payment way, the city population, the vehicle age and the sum insured. We then computed some specification criteria to test the predictive power of the explanatory variables and kept the ones which returned the best values for those criteria. These explanatory variables[4] are summarized in Table 3.1.

Table 3.1: The explanatory variables and their description

Variables	Categories		
	C1	C2	C3
City population (v1)	$\leq$ 1,000,000	1,000,001-2,000,000	$\geq$ 2,000,001
Number of years that the policyholder has been registered with the insurance company (v2)	$<$ 5 years	$>$ 5 years	-
Horsepower of the vehicle (v3)	0-1400 cc	1400-1800 cc	$\geq$ 1800 cc

Table 3.2 presents a summary of the effects of the covariates on the MTPL bodily injury and property damage claim counts $k_{1,j}$ and $k_{2,j}$ based on all 5186 observations. In the first column there is a list of all explanatory variables, all broken down in their respective categories. The second column represents how many policyholders, out of our data set of 5186 observations, fall into each subgroup/category for every covariate. The rest of the table shows, conditionally on being included in a certain category per explanatory variable, the percentage of policies with claim frequencies equal to $0, 1, \geq 2$ for $k_{1,j}$ and $k_{2,j}$ respectively. For example, from Table 3.2, we can make the following observations. Firstly, in the case of the variable city population (v1) regarding the 2203 policyholders who live in a small city (C1), 92.81% of them have not made bodily injury claims and 94.68% of them have had no property damage claims. On the other hand, the 597 individuals who live in a large city (C3) seem to make more claims per both types, since the percentage that has resulted claim-free dropped to 92.34% and 93.46% for bodily injury and property damage claims respectively. Secondly, as far as the variable number of years that the policyholder has been registered with the insurance company (v2) is concerned, we see that the longer a policyholder has been with the company (C2), the bigger is the probability of its getting involved in an accident. Thirdly, regarding the

[4]Note that it would be interesting to fit the same models to larger data sets in order to study the effect of other categorical and continuous explanatory variables such as age of driver, driving experience or driving zone, which have been traditionally used in MTPL insurance.

variable horsepower of the car (v3), we observe that a high horsepower (C3) seems to be a risky category which corresponds to a lower number of claim-free policyholders, for both $k_{1,j}$ and $k_{2,j}$.

Table 3.2: Summary statistics for claim frequencies as classified by the explanatory variables

		k_1			k_2		
Covariates	Total	Count=0 (%)	Count=1 (%)	Count≥2 (%)	Count=0 (%)	Count=1 (%)	Count≥2 (%)
v1 C1	2203	92.81	5.28	1.91	94.68	5.04	0.28
v1 C2	2386	92.98	4.65	2.37	93.76	5.94	0.30
v1 C3	597	92.34	4.86	2.80	93.46	6.35	0.19
v2 C1	4491	93.00	4.68	2.32	94.25	5.48	0.27
v2 C2	695	91.69	6.81	1.50	93.19	6.48	0.33
v3 C1	2372	92.77	5.02	2.21	94.28	5.50	0.22
v3 C2	1815	93.84	3.99	2.17	94.35	5.36	0.29
v3 C3	999	91.14	6.51	2.35	93.27	6.30	0.43

At this point, in order to motivate the BNBIG regression model and the two bivariate mixed Poisson models which were presented in Sections 3.2 and 3.4 respectively, we initially perform a marginal analysis on each claim count response variable $k_{1,j}$ and $k_{2,j}$. Table 3.3 shows some standard descriptive statistics for the bodily injury and property damage claims $k_{1,j}$ and $k_{2,j}$ respectively, along with the value of the Kendall's τ correlation coefficient.

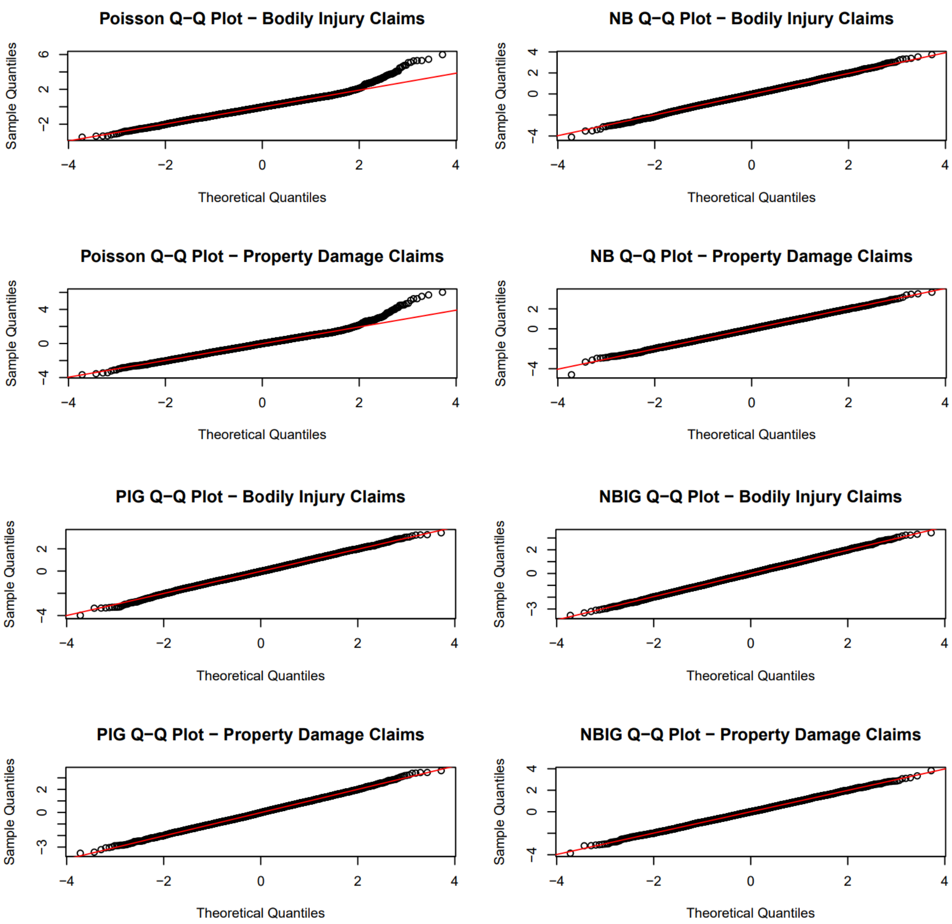
Table 3.3: Descriptive statistics for the two responses

k_1		k_2	
statistic	value	statistic	value
Minimum	0	Minimum	0
Median	0	Median	0
Mean	0.0954	Mean	0.0618
Variance	0.1375263	Variance	0.06439364
Maximum	4	Maximum	3
Kendall's τ: 0.17595			

Also, we fit the univariate Poisson, NB, PIG and NBIG regression models for the number of claims[5] and we rely on normalized quantile residuals, see Dunn and Smyth (1996), as an exploratory graphical device for investigating the adequacy of the fit of the Poisson, NB, PIG and NBIG models for both bodily injury and property damage claim counts $k_{1,j}$ and $k_{2,j}$ respectively. For these discrete response distributions, the normalized randomized quantile residuals are defined as $\hat{r}_j = \Phi^{-1}(u_j)$, where Φ^{-1} is the inverse cumulative distribution function of a standard Normal distribution and where u_j is defined as a random value from the uniform distribution on the interval $\left[F_j(x_j - 1|\hat{\boldsymbol{\phi}}), F_j(x_j|\hat{\boldsymbol{\phi}})\right]$, where F_j is the cumulative distribution function estimated for the j-th individual and where $\hat{\boldsymbol{\phi}}$ contains all estimated model parameters and x_j is the corresponding observation. Figure 3.1 depicts the normalized (random) quantiles for the Poisson, NB, PIG and NBIG models per claim type $k_{1,j}$ and $k_{2,j}$.

[5] All computing was done using the statistical computing environment language R. The parameters of the univariate NBIG regression model were estimated via the EM algorithm which was proposed by Tzougas et al. (2019). Also, ML estimation of the univariate NB and PIG regression models, for which the definition of a log-likelihood function in closed form is feasible, was straightforward by using standard statistical packages in R, such as the GAMLSS package. For more details on the GAMLSS package, see Stasinopoulos et al. (2008).

Figure 3.1: Normalized quantiles for the Poisson, NB, PIG and NBIG regression models



The findings of the marginal analysis on each claim count $k_{1,j}$ and $k_{2,j}$ indicate that the flexibility of the two mixed Poisson models and the NBIG model can be extended to the bivariate MTPL insurance claim count data setting.

In particular, as anticipated, from Table 3.3 it is evident that $k_{1,j}$ and $k_{2,j}$ have variances which are greater than their means, indicating that the two marginal variables are overdispersed already when analysed separately. Marginal overdispersion can be attributed to the differences among policyholders, which cannot be observed by the actuary and lead to extra variation in $k_{1,j}$ and $k_{2,j}$. Similarly, from Figure 3.1, we see that the NB, PIG and NBIG regression models are better assumptions than the Poisson regression model which does not capture the tails of the claim frequency distributions of $k_{1,j}$ and $k_{2,j}$. Specifically, the residuals of the former three models accounting for overdispersion are very close to the diagonals and indicate a very good fit to the distributions of both $k_{1,j}$ and $k_{2,j}$, whereas the sample quantiles of the Poisson model, due to the equidispersion constraint, near the tail end of the distributions of both $k_{1,j}$ and $k_{2,j}$ are significantly higher than the theoretical quantiles.

Furthermore, regarding the bivariate extensions of the three models, it should be noted that the BNBIG is more flexible in capturing overdispersion than its bivariate mixed Poisson counterparts since, as was previously noted, the dispersion parameters σ_1 and σ_2 control the extent of overdispersion of the individual bodily injury and property damage claim count distributions. Moreover, as σ_1 and σ_2 approach ∞, each individual claim distribution approaches a Poisson distribution and hence the BPIG model is a special case of the BNBIG model. Additionally, the common unobserved heterogeneity term, which is distributed according to the IG in the case of the proposed BNBIG model, introduces positive correlation which is the case for this data since, as we can observe from Table 3.3, Kendall's τ is positive. Thus, overall the BNBIG model can be considered as a plausible model for overdispersed and positively correlated MTPL bodily injury and property damage claim count data which we use in this study.

On the other hand, the proposed model allows only for positive correlations between claim counts, which is the case in the context of MTPL insurance, but in some other cases negative correlations may be of interest as well. Moreover, underdispersion is not covered by the model and this is definitely encountered in modeling claim counts in other insurance settings.

Also, as we can observe from Table 3.3 the value of the Kendall's τ for our data is computed at 0.17595 which is positive and small and hence the BNBIG model is a sensible choice for this data but if a negative correlation or a higher right-tailed correlation is found then copula based models can be used to more accurately model the correlation structure. Regarding the use of copulae for analysing the correlation structure of discrete variables, see, for example, Cameron et al. (2004), Nikoloulopoulos and Karlis (2009a) and Shi and Valdez (2014) among many others.

Finally, the Erlang Count Logit-weighted Reduced Mixture of Experts model (EC-LRMoE) which was recently proposed by Fung et al. (2019a) and Fung et al. (2019b) can be used in an abundance of actuarial count data settings as it can take into account both over-and-under-dispersion, positive and negative correlation between the responses and it is dense, meaning that it is guaranteed the existence of a model within the class of LRMoE that resembles well the input data, potentially avoiding the need of ad-hoc model selection procedures where multiple classes of models are fitted by trial and error in order to obtain a model that adequately represents the data.

3.5.1 Modelling results

This subsection presents the modelling results of the BNB, BPIG, and BNBIG distributions/regression models. The EM algorithm described in Section 3.3 was used to estimate the BNBIG model both for the cases without and with explanatory variables. The BNBIG model converged after a few iterations using a rather strict stopping criterion. In particular, we iterated between the E-Step and the M-Step until the relative change in log-likelihood, which is given by Eq. (3.9), between two successive iterations was smaller than 10^{-12}.

We also emphasise that for this model the choice of initial values for the vectors of the regression coefficients $\boldsymbol{\beta_i}$ and the parameters σ_i, with $i = 1, 2$, and the parameter γ of the Inverse Gaussian mixing density needed special attention because one may obtain inadmissible values if the starting values are bad. Good starting values for $\boldsymbol{\beta_i}$ and σ_i, with $i = 1, 2$, were obtained by fitting simple univariate Negative Binomial regression models. Also, a good initial value for γ, which relates to the correlation and overdispersion in the data, was feasible by equating the overdispersion of the model to the average of the observed overdispersion. Furthermore, standard errors were obtained using the standard approach of Louis (1982).

Additionally, ML estimation of the BNB and BPIG models was accomplished via the EM algorithms which were presented in Ghitany et al. (2012). As expected, ML estimation of the BNBIG model, which does not have a closed form density, was more chronologically demanding than that of the BNB and BPIG models both for the cases without and with covariate information. However, taking into account that there were 5186 policies in the sample of MTPL data that was examined in this study, that we used a rather strict stopping criterion for EM iterations and that the expectations involved at the E-Step of the algorithm do not have closed form expressions, the CPU times of the EM algorithm used for ML estimation of both the BNBIG distribution and the BNBIG regression model can be characterized as modest since both cases took a few minutes of CPU time. More importantly, as it will become clear in the following sections, the trade-off between CPU time requirements and the efficiency of the BNBIG distribution/regression model for approximating claim frequencies of MTPL bodily injury and property damage claims in our sample and for deriving a posteriori, or Bonus-Malus, ratemaking mechanisms in a bivariate context is sifted in favour of the latter two.

The ML estimates of the parameters and the corresponding standard errors in parentheses of the BNB, BPIG and BNBIG models[6] are reported in Table 3.4 for the case without covariates[7] and in Table 3.5 for the case with covariates[8]. As we observe from Table 3.5 the values of the estimated regression coefficients of the variables v1, v2 and v3 are almost identical across all three bivariate claim frequency models.

Therefore, the a priori premia per claim type resulting from these models would be almost identical when the net premium principle is used. However, as we are going to see in Subsection 3.5.3, due to the discrepancies in the values of the parameters γ, σ_1 and σ_2, the a posteriori, or Bonus-Malus, premium rates resulting from the BNB, BPIG and BNBIG models will differ with this difference being more noticeable in the case of the premia determined by the last model.

Table 3.4: Parameters estimates and in parenthesis the associated standard errors of the fitted BNB, BPIG and BNBIG distributions

BNB		BPIG		BNBIG	
μ_1	μ_2	μ_1	μ_2	μ_1	μ_2
0.0954	0.0618	0.0954	0.0618	0.0954	0.0618
(0.0542)	(0.0639)	(0.0535)	(0.0633)	(0.0526)	(0.0629)
σ_1	σ_2	σ_1	σ_2	σ_1	σ_2
-	-	-	-	0.785	11.6814
-	-	-	-	(0.0744)	(3.4871)
γ		γ		γ	
0.2612		0.4866		0.6028	
(0.1037)		(0.0554)		(0.0671)	

[6]The parameters of the models are statistically significant at a 5% threshold.

[7]Note that the mean parameters of the BNB, BPIG and BNBIG distributions are denoted by μ_1 and μ_2 and the dispersion parameter is denoted by ϕ.

[8]Note that for larger data sets with more explanatory variables it is crucial to perform variable selection of the proposed model because we may collect a large amount of policyholders' information where not all is useful. To incorporate variable selection, we maximise the penalized log-likelihood function using penalty functions such as the lasso.

Table 3.5: Parameters estimates and in parenthesis the associated standard errors of the fitted BNB, BPIG and BNBIG regression models for each covariate

	BNB		BPIG		BNBIG	
Variable	Coeff. $\boldsymbol{\beta_1}$	Coeff. $\boldsymbol{\beta_2}$	Coeff. $\boldsymbol{\beta_1}$	Coeff. $\boldsymbol{\beta_2}$	Coeff. $\boldsymbol{\beta_1}$	Coeff. $\boldsymbol{\beta_2}$
Intercept	-2.3916 (0.0998)	-2.9249 (0.1134)	-2.3815 (0.1132)	-2.9148 (0.1141)	-2.3651 (0.0966)	-2.9163 (0.1113)
v1 C2	0.0529 (0.0243)	0.1511 (0.0718)	0.0422 (0.0119)	0.1407 (0.0587)	0.0275 (0.0087)	0.1459 (0.0566)
v1 C3	0.1543 (0.0775)	0.1760 (0.0892)	0.1435 (0.0652)	0.1645 (0.0791)	0.1337 (0.0620)	0.1698 (0.0770)
v2 C2	0.0403 (0.0120)	0.1733 (0.0672)	0.0618 (0.0273)	0.1944 (0.0858)	0.0739 (0.0088)	0.1808 (0.0651)
v3 C2	-0.1232 (0.0526)	-0.0216 (0.0080)	-0.1390 (0.0557)	-0.0376 (0.0056)	-0.1537 (0.0494)	-0.0311 (0.0119)
v3 C3	0.1733 (0.0759)	0.1686 (0.0760)	0.1683 (0.0686)	0.1636 (0.0681)	0.1700 (0.0627)	0.1610 (0.0660)
	σ_1	σ_2	σ_1	σ_2	σ_1	σ_2
	- -	- -	- -	- -	0.7834 (0.0721)	11.1952 (3.4842)
	γ		γ		γ	
	0.2643 (0.1011)		0.4892 (0.0538)		0.6068 (0.0644)	

3.5.2 Model comparison

In this subsection we examine the model fit of the BNB, BPIG, BNBIG distributions/regression models employing the Global Deviance (DEV), Akaike Information Criterion

(AIC) and the Schwarz Bayesian Criterion (SBC) which are classic hypothesis/specification tests[9].

The (fitted) DEV is defined as

$$\text{DEV} = -2\hat{l}\left(\hat{\boldsymbol{\theta}}\right), \tag{3.26}$$

where $\hat{l}$ is the maximum of the log-likelihood and $\hat{\boldsymbol{\theta}}$ is the estimated parameter vector of the model. Furthermore, the AIC is given by

$$\text{AIC} = \text{DEV} + 2 \times df \tag{3.27}$$

and the SBC is given by

$$\text{SBC} = \text{DEV} + \log\left(n\right) \times df, \tag{3.28}$$

where df are the degrees of freedom, that is, the number of fitted parameters in the model and n is the number of observations in the sample.

Furthermore, regarding the case with covariates, it is important to compare the performance of the BNB, BPIG and BNBIG regression models, which are constructed in a very similar fashion, with other candidate models designed under different approaches. Therefore, in what follows the DEV, AIC and SBC will be used to compare the fit of the BNB, BPIG and BNBIG regression models with the EC-LRMoE regression model which, as was previously mentioned, exhibits a lot of desirable statistical properties that make it justified for many practical applications.

The pmf of $k_{1,j}, k_{2,j}|\boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}$ in the EC-LRMoE model is given by:

$$P\left(k_{1,j}, k_{2,j}|\boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}\right) = \prod_{i=1}^{2}\sum_{g=1}^{G}\frac{\exp\left(\boldsymbol{x_{i,j}^T\beta_{i,g}}\right)}{\sum_{g=1}^{G}\exp\left(\boldsymbol{x_{i,j}^T\beta_{i,g}}\right)}\exp(-\delta_{i,g})\sum_{b=0}^{m-1}\frac{\delta_{i,g}^{m_{i,g}k_{i,j}+b}}{(m_{i,g}k_{i,j}+b)!} \tag{3.29}$$

[9]Note that for other data sets with more complex features it is good to conduct further analyses to investigate the quality of the proposed models.

where G is the number of latent classes and $m_{i,g}$ and $\delta_{i,g}$ are the parameters of the Erlang Count expert function. The EC-LRMoE regression model is fitted to the data via Expectation-Conditional-Maximisation (ECM) algorithm [10].

The resulting DEV, AIC and SBC values for the BNB, BPIG, BNBIG distributions/regression models and the EC-LRMoE regression model are given in Table 3.6.

Table 3.6: Models comparison based on the DEV, AIC and SBC

Distributions				Regression Models				
Model	*df*	AIC	SBC	Model	*df*	DEV	AIC	SBC
BNB	3	5615	5635	BNB	13	4520	4546	4631
BPIG	3	5541	5561	BPIG	13	4433	4459	4544
BNBIG	5	5432	5465	BNBIG	15	4334	4364	4462
-	-	-	-	EC-LRMoE	14	5330	5358	5450

As is well known, a commonly used rule-of-thumb states that a model significantly outperforms a competitor if the difference in their log-likelihoods exceeds 5, corresponding to a difference in their AIC values of more than 10 and to a difference in their SBC values of more than 5, see Anderson and Burnham (2004) and Raftery (1984) respectively. This means here that, as can be seen from Table 4.3, the best fit is given by the BNBIG distribution/regression model.

At this point we perform a 10-fold cross-validation to check the robustness of the proposed models. We calculated the DEV, AIC and SBC values for the BNB, BPIG, BNBIG regression models and the EC-LRMoE regression model on each of the 10 subsets and in Table 3.7 we report an average of the 10 values that we got from each subset. Furthermore, we calculated the root-mean-square error (RMSE) on each subset. The

[10]For more details regarding the ECM algorithm for parameter estimation and the choice of good starting values, the interested reader can refer to Fung et al. (2019a) and Fung et al. (2019b). Furthermore, an R package for actuarial loss modelling using mixture of experts regression model was developed by Tseung et al. (2020).

RMSEs were calculated as follows:

$$\text{RMSE} = \frac{\sum_{j=1}^{h}\sum_{i=1}^{2}[k_{i,j} - \hat{k}_{i,j}]^2}{h} \tag{3.30}$$

where h is the size of each subset, $k_{i,j}$ for $i = 1, 2$ are the two vectors of responses of the subset and $\hat{k}_{i,j}$ for $i = 1, 2$ are the two vectors of predicted values for the responses obtained with each of the proposed models using as estimates for the parameters the ones calculated on the subset. In Table 3.7 we report a mean of the 10 RMSEs obtained on each subset.

Table 3.7: DEV, AIC, SBC and RMSE with the 10-fold cross-validation

Regression Models					
Model	*df*	DEV	AIC	SBC	RMSE
BNB	13	532	558	613	0.3854
BPIG	13	510	536	591	0.3723
BNBIG	15	441	471	535	0.3710
EC-LRMoE	14	539	567	627	0.3832

At this point it should be noted that, as empirical evidence has shown and as it can be verified by Shared's two crossings theorem, see Shared (1980), the overdispersion phenomenon can be attributed to the excess of zeros and/or heavy upper tails in count data. Therefore, since, as it can be seen from Table 3.2, most of the claim counts in our data set are zero, it would be interesting to consider a numerical example where the number of claims is larger to investigate the performance of the BNB, BPIG, BNBIG and EC-LRMoE regression models in this case.

In particular, we randomly generated a data set of size $n = 5000$ from the bivariate normal copula with weak Kendall's τ dependence ($\tau = 0.2$) and two NB regressions where the two response variables $k_{1,j}$ and $k_{2,j}$ represent bodily injury and property damage claims[11]. For the covariates and regression parameters, we chose two categorical covari-

[11]A similar numerical example can be found in Nikoloulopoulos et al. (2011).

ates with 3 categories that represent the city size and horsepower, a binary covariate that represents the car fuel and one continuous covariate that takes integer values and represents the age of the policyholder. These explanatory variables are summarized in Table 3.8:

Table 3.8: The explanatory variables of the simulated data set and their description

Variables	Categories		
	C1	C2	C3
City size (v1)	Small	Medium	Large
Horsepower of the vehicle (v2)	0-1400 cc	1400-1800 cc	$\geq$ 1800 cc
Car fuel (v3)	Diesel	Gasoline	-
Age of the policyholder (v4)	Continuous, integers from 18 to 73		

The descriptive statistics for the simulated data set are presented in Table 3.9:

Table 3.9: Descriptive statistics for the two responses for the simulated data set

k_1		k_2	
statistic	value	statistic	value
Minimum	0	Minimum	0
Median	1	Median	1
Mean	1.038	Mean	1.016
Variance	1.544834	Variance	1.508858
Maximum	9	Maximum	10
Kendall's τ: 0.1952361			

Table 3.10 depicts the DEV, AIC and SBC values for all of the fitted models.

Table 3.10: Models comparison based on the DEV, AIC and SBC on the simulated data set

Regression Models				
Model	*df*	DEV	AIC	SBC
BNB	15	27490	27520	27618
BPIG	15	27398	27428	27526
BNBIG	17	27295	27329	27440
EC-LRMoE	26	27483	27535	27705

As it can be seen from Table 3.10, in this case as well, the best fitting performances are provided by the BNBIG model. Of course, it should be noted that for other data sets the EC-LRMoE which is fully flexible and can capture different types of correlation and over-and-under-dispersion may perform better than the BNB, BPIG and BNBIG regression models which are only suitable for examining the relationship between positively correlated and overdispersed claim counts. In such cases, the EC-LRMoE model should be preferred over the BNB, BPIG and BNBIG models.

We perform now a 10-fold cross-validation on the simulated dataset and the specification criteria and RMSEs are reported in Table 3.11.

Table 3.11: DEV, AIC, SBC and RMSE with the 10-fold cross-validation on the simulated data set

Regression Models					
Model	*df*	DEV	AIC	SBC	RMSE
BNB	15	2852	2882	2945	0.3823
BPIG	15	2763	2793	2856	0.3711
BNBIG	17	2605	2639	2711	0.3693
EC-LRMoE	26	2801	2853	2963	0.3801

In what follows we will restrict our attention to the BNB, BPIG and BNBIG models which will be used within the Bayesian paradigm for deriving a posteriori, or Bonus-Malus, ratemaking mechanisms, or Bonus-Malus Systems (BMSs) in the next subsection.

3.5.3 Calculation of the a posteriori premia

In this subsection, we examine the response of the BNBIG distribution/regression model to claim experience and we compare it to those of the two bivariate mixed Poisson distributions/regression models which were presented in Section 3.4.

Consider the policyholder j, $j = 1, ...n$, with number of bodily injury and property damage claims $k_{1,j,l}$ and $k_{2,j,l}$ respectively, for the year of coverage l, with $l = 1, ..., t$. Assume that the cumulative number of claims per type $i = 1, 2$ for all the years that the individual j has been registered with the insurance company is denoted as $K_{i,j} = \sum_{l=1}^{t} k_{i,j,l}$. Also, let the unobserved inverse Gaussian random variable take into account individual characteristics.

On the path towards actuarial relevance, the Bayesian view is taken[12] to compute the posterior distribution of $\lambda_{j,t+1}$ for the period $t+1$ given the observations of the reported accidents in the preceding t periods and observable characteristics in the preceding $t+1$ periods and the current period. In fact, as we mentioned in Chapter 1, Bayesian statistics ensures the most accurate BMSs as it grants the optimal and fairer premia estimates. In particular, the posterior distribution of $\lambda_{j,t+1}$ can be derived as follows:

$$f\left(\lambda_{j,t+1}|k_{1,j,1},...,k_{1,j,t};k_{2,j,1},...,k_{2,j,t};\boldsymbol{x_{1,j,1}},...,\boldsymbol{x_{1,j,t+1}};\boldsymbol{x_{2,j,1}},...,\boldsymbol{x_{2,j,t+1}}\right)$$
$$= \frac{\prod_{l=1}^{t} P\left(k_{1,j},k_{2,j}|\boldsymbol{x_{1,j}},\boldsymbol{x_{2,j}},\lambda_j\right) f\left(\lambda_{j,t+1}\right)}{\int_0^\infty P\left(k_{1,j},k_{2,j}|\boldsymbol{x_{1,j}},\boldsymbol{x_{2,j}},\lambda_j\right) f\left(\lambda_{j,t+1}\right) d\lambda_{j,t+1}}, \tag{3.31}$$

where $P\left(k_{1,j},k_{2,j}|\boldsymbol{x_{1,j}},\boldsymbol{x_{2,j}},\lambda_j\right)$ is the bivariate Poisson distribution in the case of the BNB and BPIG models, while it takes the form of the bivariate Negative Binomial in the case of the BNBIG model and where $f(\lambda_{j,t+1})$ is the pdf of the Gamma distribution in the case of the BNB model and the pdf of the Inverse Gaussian distribution in the case of the BPIG and BNBIG models respectively.

Using the net premium principle and the quadratic loss function, one can easily see that the optimal estimator of $\lambda_{j,t+1}$ is the mean of the posterior distribution in Eq. (3.31),

[12]For more details regarding the Bayesian interpretation of Bonus-Malus systems the interested reader can refer, for instance, to Dionne and Vanasse (1992) and Lemaire (1995).

given by

$$\mathbb{E}\left(\lambda_{j,t+1}|k_{1,j,1},...,k_{1,j,t};k_{2,j,1},...,k_{2,j,t};\boldsymbol{x_{1,j,1}},...,\boldsymbol{x_{1,j,t+1}};\boldsymbol{x_{2,j,1}},...,\boldsymbol{x_{2,j,t+1}}\right)$$
$$=\int_0^\infty \lambda_{j,t+1} f\left(\lambda_{j,t+1}|k_{1,j,1},...,k_{1,j,t};k_{2,j,1},...,k_{2,j,t};\boldsymbol{x_{1,j,1}},...,\boldsymbol{x_{1,j,t+1}};\right. \tag{3.32}$$
$$\left.\boldsymbol{x_{2,j,1}},...,\boldsymbol{x_{2,j,t+1}}\right)d\lambda_{j,t+1}.$$

- In the case of the BNB model, Eq. (3.31) is a Gamma distribution with parameters $\gamma+\sum_{i=1}^2 k_{i,j}$ and $\gamma+\sum_{l=1}^t\sum_{i=1}^2\boldsymbol{\varepsilon_{i,j}}$, and hence Eq. (3.32) takes the form:

$$\mathbb{E}\left(\lambda_{j,t+1}|k_{1,j,1},...,k_{1,j,t};k_{2,j,1},...,k_{2,j,t};\boldsymbol{x_{1,j,1}},...,\boldsymbol{x_{1,j,t+1}};\boldsymbol{x_{2,j,1}},...,\boldsymbol{x_{2,j,t+1}}\right)$$
$$=\frac{\gamma+\sum_{i=1}^2 k_{i,j}}{\gamma+\sum_{l=1}^t\sum_{i=1}^2\boldsymbol{\varepsilon_{i,j}}}. \tag{3.33}$$

- In the case of the BPIG model, Eq. (3.31) is a Generalized Inverse Gaussian (GIG) distribution with parameters $\sqrt{\gamma^2+2\sum_{l=1}^t\sum_{i=1}^2\boldsymbol{\varepsilon_{i,j}}}$, γ and $\sum_{i=1}^2 k_{i,j}-\frac{1}{2}$ and thus Eq. (3.32) is given by:

$$\mathbb{E}\left(\lambda_{j,t+1}|k_{1,j,1},...,k_{1,j,t};k_{2,j,1},...,k_{2,j,t};\boldsymbol{x_{1,j,1}},...,\boldsymbol{x_{1,j,t+1}};\boldsymbol{x_{2,j,1}},...,\boldsymbol{x_{2,j,t+1}}\right)$$
$$=\frac{\gamma K_{\sum_{i=1}^2 k_{i,j}+\frac{1}{2}}\left(\gamma\sqrt{\gamma^2+2\sum_{l=1}^t\sum_{i=1}^2\boldsymbol{\varepsilon_{i,j}}}\right)}{\sqrt{\gamma^2+2\sum_{l=1}^t\sum_{i=1}^2\boldsymbol{\varepsilon_{i,j}}}\,K_{\sum_{i=1}^2 k_{i,j}-\frac{1}{2}}\left(\gamma\sqrt{\gamma^2+2\sum_{l=1}^t\sum_{i=1}^2\boldsymbol{\varepsilon_{i,j}}}\right)}. \tag{3.34}$$

- In the case of the the BNBIG model, the expectation in Eq. (3.32) cannot be computed in closed form. However, it can be computed based on either numerical integration or a Monte Carlo approach since both schemes do not rely on the knowledge of the pdf given by Eq. (3.31).

Following the aforementioned methodology, we calculate the Bonus-Malus premia resulting from the BNBIG model and we compare them to those derived by the BNB and BPIG models based only on the number of individual bodily injury and property damage claims, i.e. the a posteriori criteria, and based on the characteristics of the policyholders and their cars, i.e. the a priori criteria. The premium rates will be divided by the premium when $t = 0$, i.e. we calculate the relative premia, since we are interested in the differences between various classes and the results are presented so that the premium for a new policyholder is 100.

Firstly, Table 3.12 depicts comparable relative premia for the BNB, BPIG and BNBIG distributions, assuming that the number of claims $k_{1,j}$ and $k_{2,j}$ ranges from 0 to 3 for each claim type and the age of the policy is $t = 1$, $t = 2$ and $t = 3$ years.

Table 3.12: Comparison of the a posteriori, or Bonus-Malus, premium rates for $t = 1, 2, 3$, bivariate claim frequency distributions

	$t=1$ BNB distribution				$t=1$ BPIG distribution				$t=1$ BNBIG distribution			
$k_{1,j}/k_{2,j}$	0	1	2	3	0	1	2	3	0	1	2	3
0	72.69	350.98	629.27	907.56	79.80	256.06	553.66	892.83	82.97258	233.17	493.12	809.72
1	350.98	629.27	907.56	1185.85	256.06	553.66	892.83	1240.99	200.62	419.84	696.16	1001.87
2	629.27	907.56	1185.85	1464.14	553.66	892.83	1240.99	1591.52	362.47	603.31	876.47	1171.39
3	907.56	1185.85	1464.14	1742.43	892.83	1240.99	1591.52	1942.92	527.58	770.77	1038.14	1324.29
	$t=2$ BNB distribution				$t=2$ BPIG distribution				$t=2$ BNBIG distribution			
$k_{1,j}/k_{2,j}$	0	1	2	3	0	1	2	3	0	1	2	3
0	57.10	275.69	494.28	712.87	68.35	197.65	411.53	657.85	72.17	186.15	380.62	622.19
1	275.69	494.28	712.87	931.46	197.65	411.53	657.85	912.20	165.03	332.97	546.14	784.76
2	494.28	712.87	931.46	1150.05	411.53	657.85	912.20	1168.82	294.61	483.14	697.83	931.01
3	712.87	931.46	1150.05	1368.64	657.85	912.20	1168.82	1426.30	430.85	623.93	836.36	1064.56
	$t=3$ BNB distribution				$t=3$ BPIG distribution				$t=3$ BNBIG distribution			
$k_{1,j}/k_{2,j}$	0	1	2	3	0	1	2	3	0	1	2	3
0	47.01	226.99	406.97	586.96	60.73	162.83	328.94	521.70	64.61	155.89	308.90	500.72
1	226.99	406.97	586.96	766.94	162.83	328.94	521.70	721.75	141.10	275.81	447.09	639.94
2	406.97	586.96	766.94	946.92	328.94	521.70	721.75	923.98	248.54	402.02	576.93	767.35
3	586.96	766.94	946.92	1126.90	521.70	721.75	923.98	1127.06	363.96	522.81	697.36	884.99

Secondly, when both criteria are considered, we examine three risk class profiles that can be classified as Best, Average and Worst according to the mean claim frequencies $\boldsymbol{\varepsilon_{1,j}}$ and $\boldsymbol{\varepsilon_{2,j}}$, with $j = 1, ..., n$, based on the same set of explanatory variables per claim type $i = 1, 2$. Specifically, the Best, Average and Worst profiles, for our data, are determined as such based on category C1 for all three explanatory variables v1, v2 and v3 in the case of the first, category C2 for v1, v2 and v3 in the case of the second, and category C3 for v1 and v3 and C2 for v2 in the case of the third. The results for all three profiles per claim type are presented in Table 3.13 in the case of the BNB, BPIG and BNBIG models respectively.

Table 3.13: Results of the fitted BNB, BPIG and BNBIG regression models for each risk class profile

Regression model	Profile	$\boldsymbol{\varepsilon_{1,j}}$	$\mathrm{Var}(k_{1,j}\vert\boldsymbol{x_{1,j}})$	$\boldsymbol{\varepsilon_{2,j}}$	$\mathrm{Var}(k_{2,j}\vert\boldsymbol{x_{2,j}})$
BNB	Best	0.091483	0.123149	0.053670	0.064569
	Average	0.088779	0.118601	0.072650	0.092620
	Worst	0.132166	0.198256	0.090085	0.120790
BPIG	Best	0.092412	0.128097	0.054215	0.066497
	Average	0.089233	0.122506	0.072100	0.095267
	Worst	0.134270	0.209604	0.091419	0.126341
BNBIG	Best	0.093940	0.159764	0.054134	0.063065
	Average	0.089153	0.148440	0.072752	0.088883
	Worst	0.137038	0.277115	0.090293	0.115140

We observe from Table 3.13 that, as expected, for all three risk profiles small discrepancies lie in the mean values $\boldsymbol{\varepsilon_{1,j}}$ and $\boldsymbol{\varepsilon_{2,j}}$ in the case of the BNB, BPIG and BNBIG regression models respectively. However, when the a posteriori correction will be calculated, we will see that compared to the relative Bonus-Malus premia provided by the two bivariate mixed Poisson models, the premia derived from the BNBIG model will be much less extreme for policyholders with some bodily injury and property damage claim experience.

This can be clearly justified since given the estimates of σ_1 and σ_2 of the BNBIG model, which also appear in the marginal Negative Binomial distributions, see Eq. (3.1), we can assess the extent of marginal overdispersion for the bodily injury and property damage claim distributions of an individual policyholder with any given mean frequency rates per claim type. Therefore, this situation affects the calculation of the Bonus-Malus premium rates.

Table 3.14 shows some premia for the three risk profiles during the years $t = 1$, $t = 2$ and $t = 3$ respectively. Such table can provide a more complete picture to the actuary than Table 3.12, where only the a posteriori criteria were considered, as they include all available information on the level of riskiness of the individual, as assessed by the insurance company.

Table 3.14: A Posteriori, or Bonus-Malus, premium rates for $t = 1, 2, 3$ for the three risk profiles, bivariate claim frequency regression models

	$t=1$ BNB regression model Best profile				$t=1$ BPIG regression model Best profile				$t=1$ BNBIG regression model Best profile			
$k_{1,j}/k_{2,j}$	0	1	2	3	0	1	2	3	0	1	2	3
0	73.83	353.16	632.49	911.83	80.31	260.95	566.63	914.58	83.57	236.80	503.08	827.87
1	353.16	632.49	911.83	1191.16	260.95	566.63	914.58	1271.53	203.03	426.77	709.44	1022.71
2	632.49	911.83	1191.16	1470.49	566.63	914.58	1271.53	1630.83	367.29	612.97	892.18	1194.32
3	911.83	1191.16	1470.49	1749.83	914.58	1271.53	1630.83	1991.00	534.58	782.64	1055.99	1349.02
	$t=2$ BNB regression model Average profile				$t=2$ BPIG regression model Average profile				$t=2$ BNBIG regression model Average profile			
$k_{1,j}/k_{2,j}$	0	1	2	3	0	1	2	3	0	1	2	3
0	56.87	272.05	487.22	702.40	68.18	194.75	403.56	644.34	71.95	181.71	365.84	592.13
1	272.05	487.22	702.40	917.57	194.75	403.56	644.34	893.16	163.14	324.61	526.88	750.85
2	487.22	702.40	917.57	1132.75	403.56	644.34	893.16	1144.28	290.75	471.92	675.62	894.52
3	702.40	917.57	1132.75	1347.93	644.34	893.16	1144.28	1396.27	425.52	610.76	812.12	1026.25
	$t=3$ BNB regression model Worst profile				$t=3$ BPIG regression model Worst profile				$t=3$ BNBIG regression model Worst profile			
$k_{1,j}/k_{2,j}$	0	1	2	3	0	1	2	3	0	1	2	3
0	42.19	201.81	361.43	521.05	58.12	141.20	273.16	427.77	60.81	135.37	258.38	414.87
1	201.81	361.43	521.05	680.67	141.20	273.16	427.77	589.46	122.04	229.40	367.95	526.46
2	361.43	521.05	680.67	840.30	273.16	427.77	589.46	753.45	205.75	328.89	471.60	628.84
3	521.05	680.67	840.29	999.91	427.77	589.46	753.45	918.37	296.22	424.87	568.08	723.48

Overall, from Tables[13] 3.12 and 3.14, we see that the BNB, BPIG and BNBIG distributions/regression models result in a noticeable decrement in the premia that must be paid by the policyholder j when it has a claim free year for both types of claims $i = 1, 2$, whereas if it has one or more claims of type $i = 1, 2$ the premium rates increase, hence resulting in bonus or malus in the former and latter case respectively.

[13] Note that the symmetry of each Table for the bivariate mixed Poisson models is a logical consequence of the common random effects assumption, whereas the premia are distinguishable per claim type under the BNBIG model which is due to its quantiple Poisson-Gamma-Poisson-Gamma-Inverse Gaussian mixture decomposition.

Furthermore, as was previously mentioned, we observe that the BNBIG distribution/regression model returns lower premia for individuals with some claim history per claim type $i = 1, 2$ than the two bivariate mixed Poisson distributions/regression models. For example, under the BNBIG distribution, for a policyholder who had $k_{1,j} = 3$ and $k_{2,j} = 3$ claims, Table 3.12 shows a premium of only 1324.29 for the year of coverage $t = 1$. Meanwhile, for the same number of claims per claim type $i = 1, 2$, we observe that the BNB and the BPIG distributions result in higher premia of 1742.43 and 1942.92 respectively.

Also, similar discrepancies are observed if we incorporate the a priori information from Table 3.13. For example, still for the case when $k_{1,j} = 3$ and $k_{2,j} = 3$, according to Table 3.14, an individual with the Best profile is expected to pay a premium of 1349.02 under the BNBIG model as opposed to the higher premia of 1749.83 and 1991.00 under the BNB and the BPIG regression models respectively for the year of coverage $t = 1$. This characteristic of the BNBIG model can be explained by the fact that this model is constructed by starting with two Negative Binomial models which assume that the individual bodily injury and property damage claim experience will be overdispersed as opposed to the two Poisson models in the BNB and BPIG models.

The overdispersion is larger for policyholders with larger mean claim rates per claim type. Therefore, extreme individual bodily injury and property damage claim counts are more likely under the BNBIG model, resulting in more moderate relative premia than under those models based on the bivariate Poisson mixtures[14] .

The second noticeable difference between the BNBIG and the two bivariate mixed Poisson distributions/regression models in the calculation of Bonus-Malus premia is that the two bivariate Poisson mixtures can only take into account the policyholder's total number of claims, which is computed by aggregating the bodily injury and property damage claims, but are unable to distinguish between the two types of claims. For

[14] Similar findings were reported by Shengwang et al. (1999) and Gómez-Déniz et al. (2008) and Tzougas et al. (2019).

instance, for the case without covariates, we see from Table 3.12 that for $\sum_{i=1}^{2} k_{i,j} = 3$ the policyholder has to pay premia of 907.56 and 892.83 for $t = 1$ under the BNB and BPIG distributions respectively, regardless of the exact frequencies of bodily injury and property damage claims.

Similarly, for the case with covariates, we see from Table 3.14 that for $\sum_{i=1}^{2} k_{i,j} = 3$, an individual with the Best profile has to pay premia of 911.83 and 914.58 for $t = 1$ under the BNB and BPIG regression models respectively, regardless of the exact composition per type of claim. Overall, if we consider all the four cases for $k_{1,j}$ and $k_{2,j}$ in Tables 3.12 and 3.14:

1. $k_{1,j} = 3$ and $k_{2,j} = 0$,
2. $k_{1,j} = 2$ and $k_{2,j} = 1$,
3. $k_{1,j} = 1$ and $k_{2,j} = 2$,
4. $k_{1,j} = 0$ and $k_{2,j} = 3$,

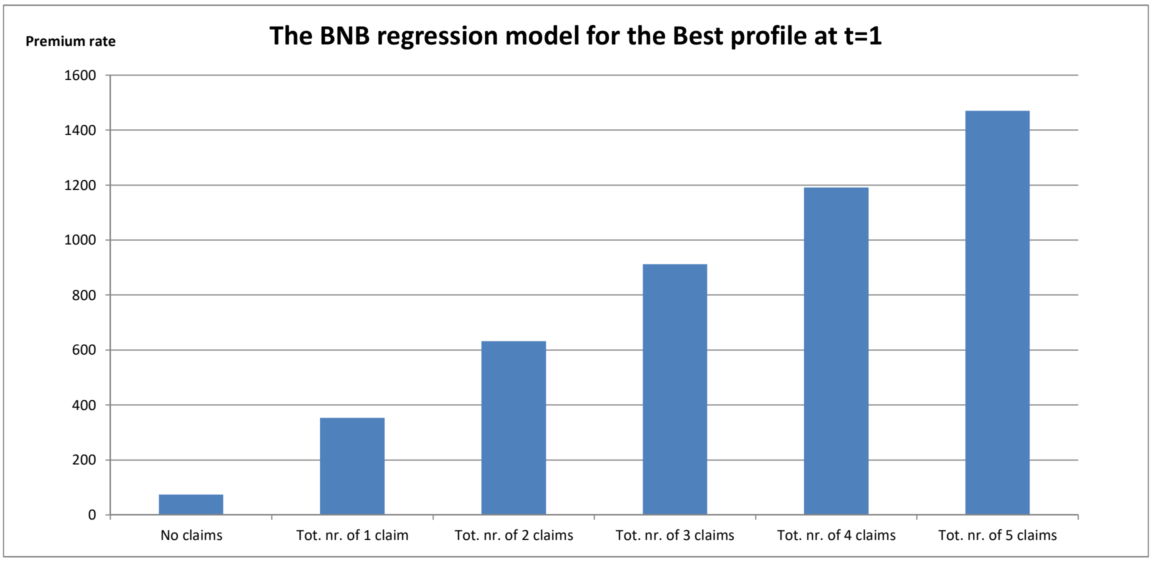
for the same year of coverage t and the same type of risk-profile, the premium rates do not vary per claim type in the case of the two bivariate mixed Poisson models. In particular, in Tables 3.12 and 3.14, the values on the diagonals are always the same for a certain profile and for a certain year of insurance in the case of the BNB and BPIG models. However, these premium rates ought to be evaluated differently than under the two bivariate mixed Poisson models since the two types of claims have different frequencies and hence different means (as seen in Table 3.3, $\mathbb{E}(k_1) = 0.0954$ and $\mathbb{E}(k_2) = 0.0618$). Consequently, the probability of resulting claim free for the first type of claim is not the same as for the second type of claim. Therefore, from a practical business standpoint, these discrepancies in the two responses $k_{1,j}$ and $k_{2,j}$ should be taken into consideration for constructing a bivariate claim frequency model that will be the building block for the a posteriori ratemaking process.

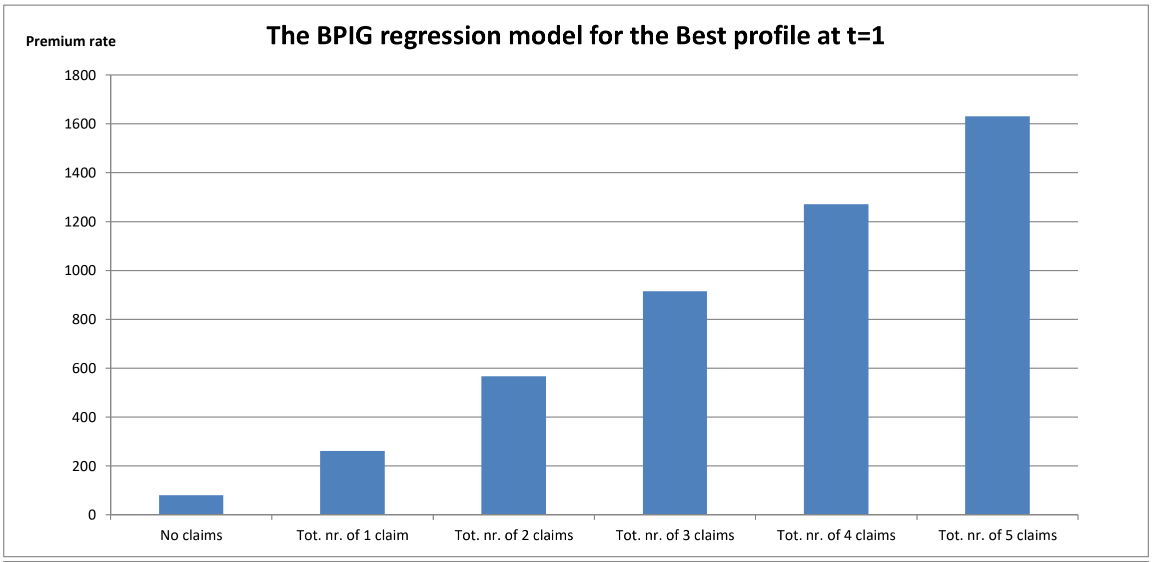
Fortunately, as was previously mentioned, the BNBIG distribution/regression model results in varying premium rates depending on the total number of claims and the composition of claims. This feature of the BNBIG model enhances its validity as a model for the claim numbers $k_{1,j}$ and $k_{2,j}$ as it leads to a premium structure that can be sufficiently explained to policyholders and regulators. In particular, the findings in Tables 3.12 and 3.14 indicate that under the BNBIG distribution/regression model, a high number of claims per type two is discouraged more than per type one in all four above mentioned cases. Thus, since for our data the second response has a smaller mean than the first, this result is consistent with the core principle underpining the design of Bonus-Malus systems which is to promote careful driving.

For instance, for the case without covariates, we see from Table 3.12 that for $t = 1$ the policyholder has to pay premia of 527.58, 603.31, 696.16 and 809.72 for the cases 1, 2, 3 and 4 respectively. The same holds if we include the a priori information, in fact in Table 3.14 we see the same diversification in the premium rates depending on the type of claim. If the policyholder has the Best profile, then it has to pay premia of 534.58, 612.97, 709.44 and 827.87 for the cases 1, 2, 3 and 4 respectively for $t = 1$.

A graphical representation of Table 3.14 is depicted in Figure 3.2 for the case of the Best profile in $t = 1$, in Figure 3.3 for the case of the Average profile in $t = 2$ and in Figure 3.4 for the case of the Worst profile in $t = 3$.

Figure 3.2: Premium rates for the Best profile at $t = 1$, bivariate claim frequency regression models





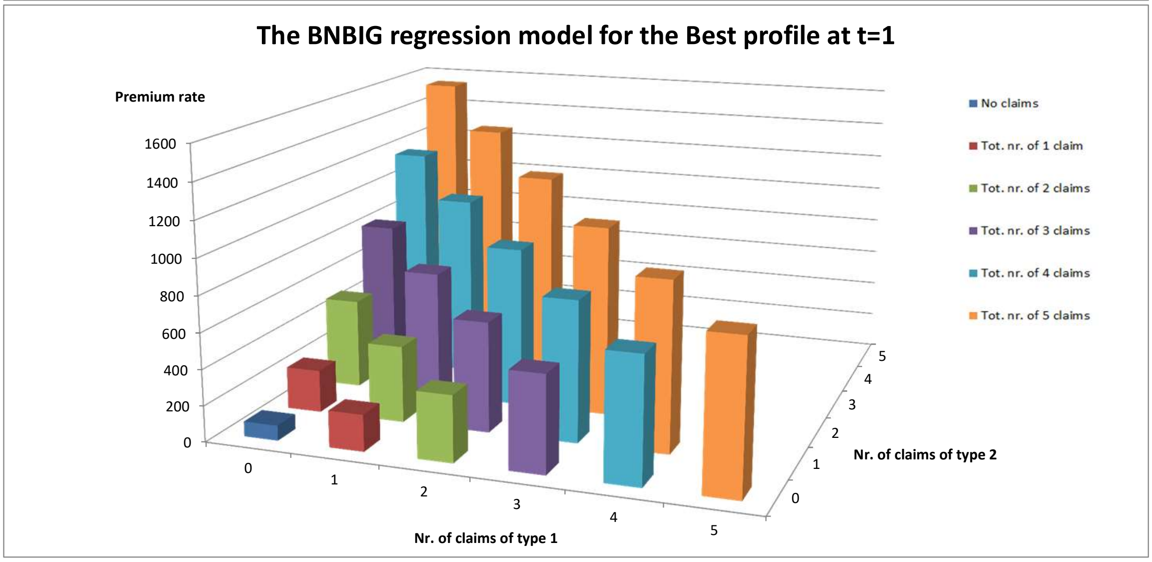
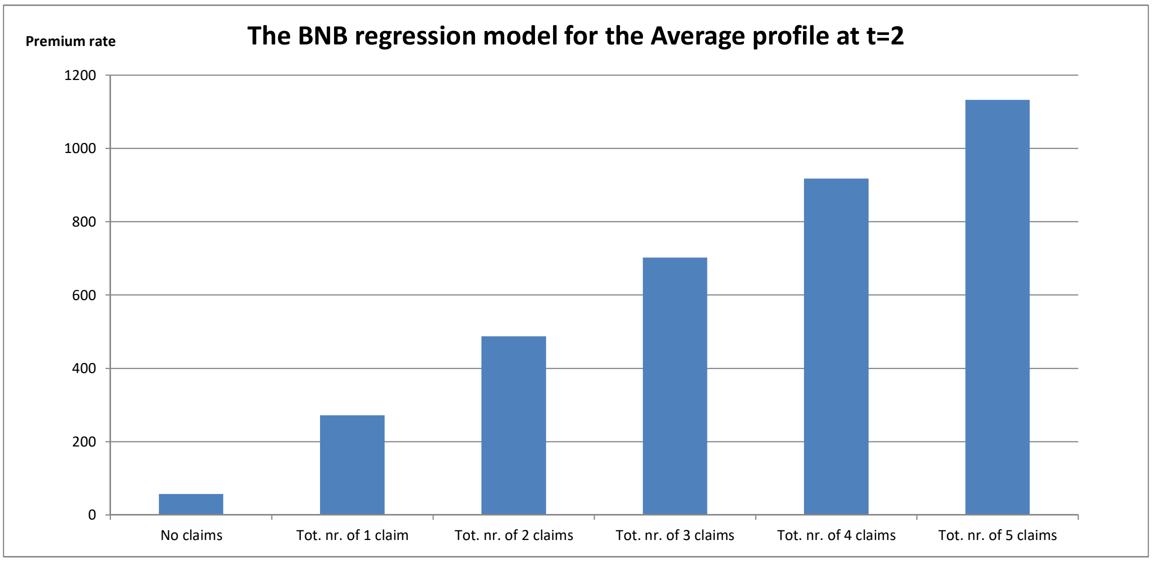
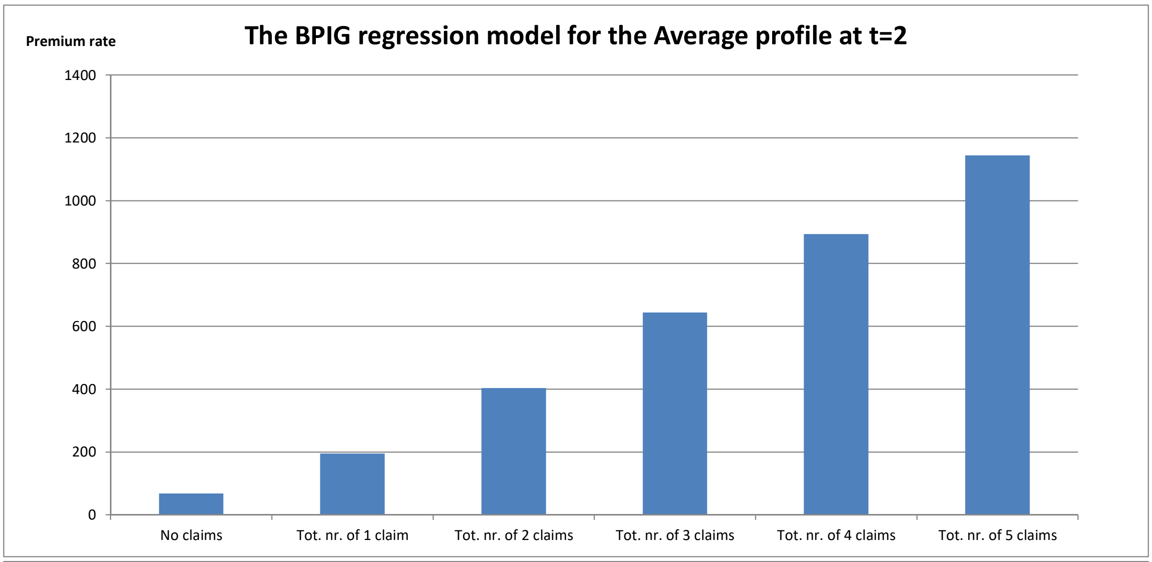


Figure 3.3: Premium rates for the Average profile at $t = 2$, bivariate claim frequency regression models





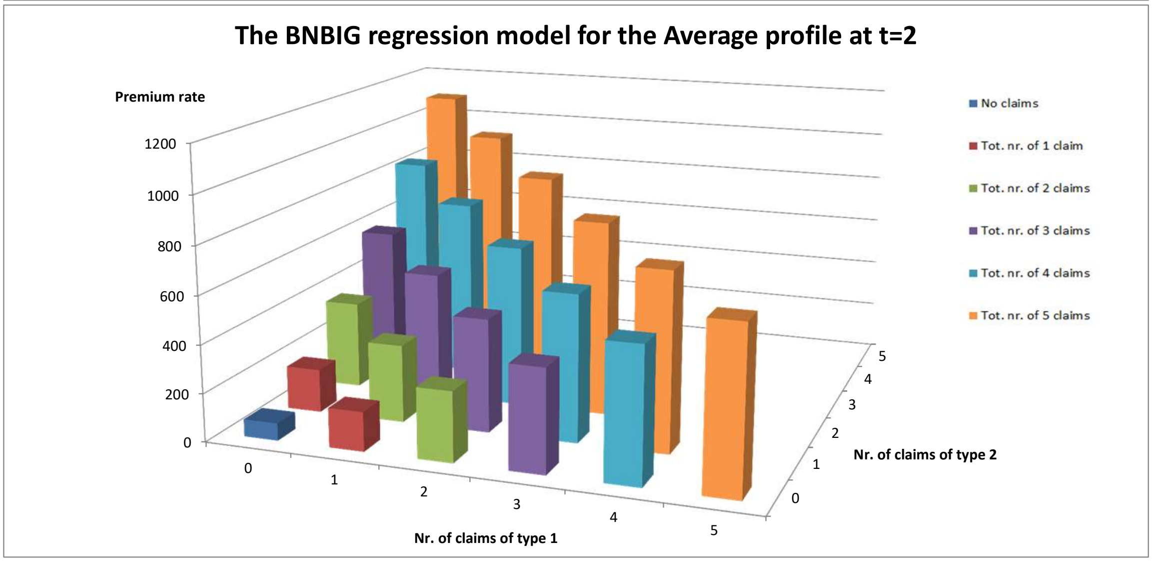
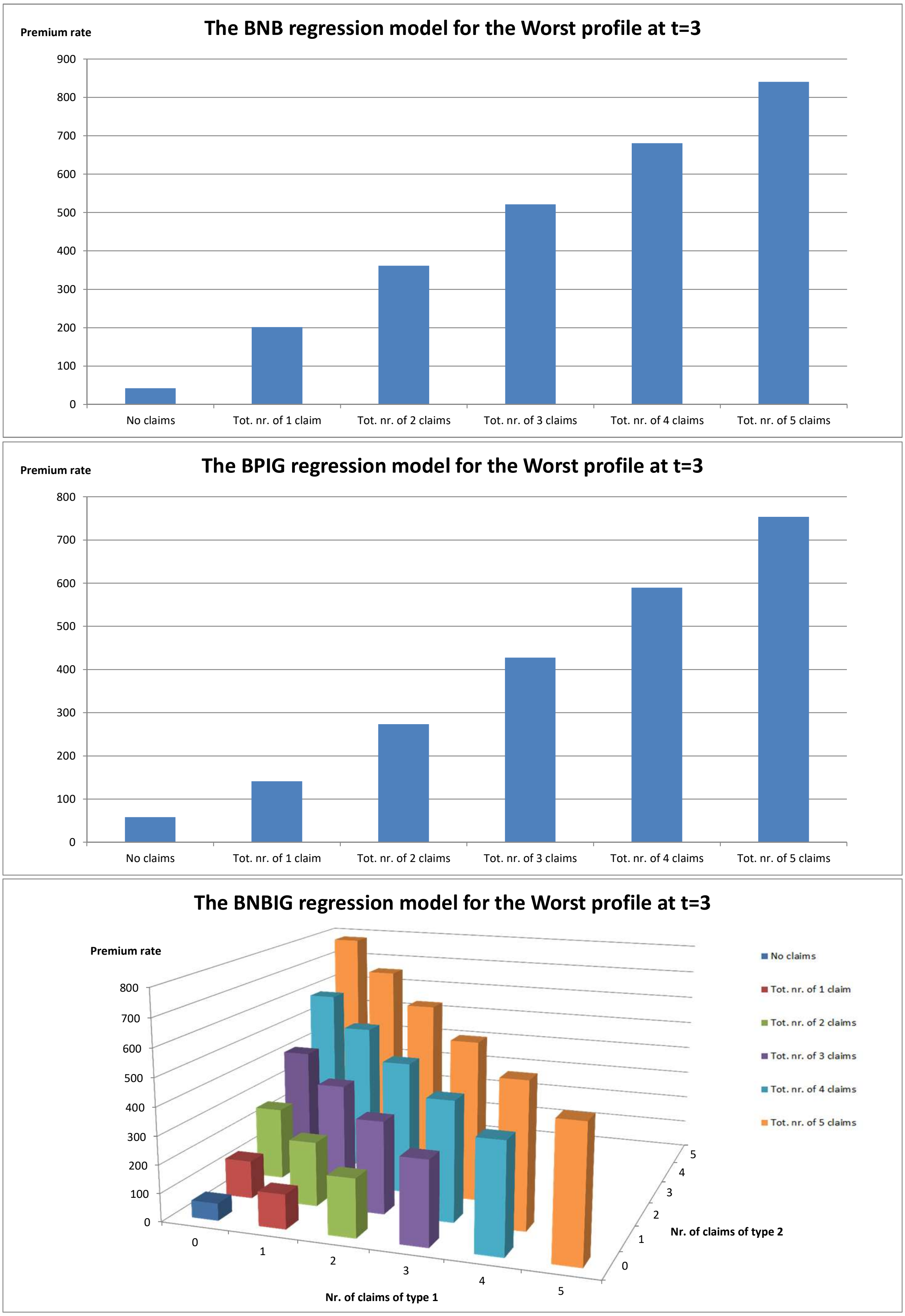


Figure 3.4: Premium rates for the Worst profile at $t = 3$, bivariate claim frequency regression models



3.6 Concluding remarks

In this chapter we demonstrated how to construct the BNBIG count regression model based on the assumption that both marginals follow the NBIG distribution, which was recently proposed by Tzougas et al. (2019) for ratemaking purposes. The BNBIG model accomodates overdispersion and allows for positive correlation structures in two-dimensional count valued data. Thus, the proposed model is suitable for addressing the a posteriori, or Bonus-Malus, ratemaking problem of pricing an automobile insurance contract in the bivariate setting where the dynamics for the premium determination are governed by the interactions of third party bodily injury claims and property damage claims, which are conceivably positively correlated with each other.

Furthermore, an EM type scheme was proposed for ML estimation of the parameters of the BNBIG model which does not have its jpmf in closed form in a computationally parsimonious manner. The ML estimation procedure we developed avoids overflow issues that may be plausible in the bivariate context via alternative numerical maximisation algorithms.

In our numerical illustration, special consideration was put on the comparison of the a posteriori premium rates derived from the BNBIG distribution/regression model with those determined by the BNB and BPIG distributions/regression models. The reason for this is that, in contrast to the numerous studies that have been devoted to univariate experience rating models, the extent to which the a posteriori tariff system is affected when the claim frequency experience consists of detailed information on two different types of insurance claims arising from the same policy has not been fully elucidated in the MTPL pricing literature thus far. The results indicated that the employment of the BNBIG regression which presents the most superior fit for our data is beneficial for the insurance company, since it can enable them to adopt a milder a posteriori pricing strategy for policyholders with some claim experience and it can provide a more complete picture about the extent to which the premia vary according to the frequency of each type of claim.

Furthermore, an interesting line of further research would be to to tackle Bonus-Malus ratemaking based on generalizations of the proposed model. For instance, two random effects which are distributed according to different Inverse Gaussian mixing densities can be added for modelling the unobserved heterogeneity when dealing with different types of claims from different types of coverage, see Bermúdez and Karlis (2017). Also, for example, the BNBIG model can be adapted to take into account both the positive correlation between the MTPL bodily injury and property damage claims and the serial correlation between the observations of the same insured observed over time. This could be done proceeding along similar lines as Bermúdez et al. (2018), who were the first to consider a bivariate INAR(1) regression model which can provide an integrated framework that can take into account both time dependence and cross dependence, which have been commonly treated as separate entities in the ratemaking. Finally, for instance, the regression models for the two marginal means of the BNBIG model can be embedded within neural network architectures in order to explore interactions between feature components beyond multiplications, see Wüthrich (2019).

Acknowledgements:

The authors would like to thank the participants at the 23rd International Congress on "Insurance: Mathematics and Economics" (IME) in Munich, 2019. We are also deeply grateful to Dr Tsz Chai Fung for his very useful feedback and for providing us with the R code for the EC-LRMoE model. Finally, we would like to thank the editor and the two anonymous referees for their very helpful comments and suggestions which have significantly improved this chapter in its submitted version.

Chapter 4

EM estimation for a new class of bivariate mixed Poisson regression models with varying dispersion: an application to a posteriori ratemaking

4.1 Introduction

4.1.1 Multivariate models

The rapid advent of big data over the last few decades has motivated the need for constructing bivariate (and/or multivariate) regression models that can permit inferences about dependence structures which typically arise in high-dimensional count-valued data sets based on explanatory variables. The interested reader is referred to the recent editions of the books of Winkelmann (2008) and Cameron and Trivedi (2013) for thorough reviews of regression models for bivariate (and/or multivariate) count data. In general, the three main classes of models, that have been widely applied in various fields of studies, are the bivariate (and/or multivariate) Poisson models, bivariate (and/or multivariate) mixed Poisson models and copula-based models.

The literature on bivariate (and/or multivariate) Poisson distributions started growing nine decades ago, see M'Kendrick (1925), and has now reached solid and extensive foundations. In particular, as is mentioned in Cameron and Trivedi (2013), the number of possible ways in which the univariate Poisson distribution can be generalized to a bivariate (and/or multivariate) Poisson distribution is not exhaustive. Also, many references for bivariate (and/or multivariate) Poisson distributions, including historical remarks, can be found in Johnson et al. (1997) and Krummenauer (1998). In the context of regression analysis, different versions of the bivariate (and/or multivariate) Poisson model have also been studied by many authors, see, for instance, Jung and Winkelmann (1993), Ho and Singer (2001), Kocherlakota and Kocherlakota (2001), Karlis and Meligkotsidou (2005), p. 205 of Winkelmann (2008) and Famoye (2010).

The bivariate (and/or multivariate) mixed Poisson models, which belong to the second class, can permit for overdispersion in the data. The bivariate (and/or multivariate) extensions of the Negative Binomial, Poisson-Inverse Gaussian and Poisson-Lognormal regression model have been the most popular choices. The literature along this line includes, for example, the works of Stein and Juritz (1987), Stein et al. (1987) and Kocherlakota (1988) for the case without explanatory variables. Also, for instance, Munkin and Trivedi (1999), Gurmu and Elder (2000), Chib and Winkelmann (2001), Wang (2003), Alfò and Trovato (2004), Park and Lord (2007), Ma et al. (2008), El-Basyouny and Sayed (2009), Aguero-Valverde and Jovanis (2009), Famoye (2012), Ghitany et al. (2012), Zhan et al. (2015) and Silva et al. (2017) considered extensions with regression specifications for the marginal means.

Finally, a multivariate count distribution can be viewed as a continuous copula distribution paired with discrete marginals. The copula functions can fully specify the dependence structure separately from the univariate marginals, see, for example, Section 1.6 of Joe (1997). The literature on copula-based regression models includes, among others, Lee (1999), Cameron et al. (2004), Zimmer and Trivedi (2006), Nikoloulopoulos and Karlis (2009b), Cook et al. (2010), Nikoloulopoulos and Karlis (2010), Nikoloulopoulos (2013a), Nikoloulopoulos (2013b) and Nikoloulopoulos (2016). At this point it is worth noting that some members of the second and the third classes can allow for either positive or negative correlations between two variables and hence can adequately describe diverse data situations in numerous bivariate (and/or multivariate) domains including, but not limited to, marketing, epidemiology, medical science and finance.

4.1.2 Multivariate models in non-life insurance

In non-life insurance practice, it is common for the actuary to observe the existence of dependence structures between different types of claims and their associated claim counts. Nevertheless, most of research endeavours have been traditionally confined on univariate mixed Poisson count regression models which can only be used to specify a separate model for different claim types, whereas the effort towards relaxing the independence assumption is still sparse, even if such an assumption may not be realistic.
As far as ratemaking (which is the main focus of this work) is concerned, notable exceptions are the articles by Bermúdez (2009), Bermúdez and Karlis (2011), Bermúdez and Karlis (2012) and Shi and Valdez (2014), who introduced different bivariate (and/or multivariate) regression and copula based models and also pointed out the existence of a positive correlation between claim counts of two (and/or multiple) types of claims. Also, Bermúdez and Karlis (2017) were the first to take the Bayesian view for constructing two bivariate experience rating models, which integrate the a priori ratemaking based on bivariate Poisson regression models, extending the existing literature in the bivariate setting which was confined on ratemaking models that were derived via the credibility

approach.

Additionally, it should be noted that recently many alternative approaches have been proposed in the literature for constructing flexible bivariate (and/or multivariate) insurance claim frequency regression models, see, for instance, Abdallah et al. (2016), Bermúdez et al. (2018), Pechon et al. (2018), Pechon et al. (2019b), Pechon et al. (2019a), Bolancé and Vernic (2019), Denuit et al. (2019), Fung et al. (2019a), Fung et al. (2019b) and Bolancé et al. (2020).

4.1.3 The class of multivariate mixed Poisson regression models with varying dispersion

The present study, is concerned with introducing a new class of bivariate mixed Poisson regression models with varying dispersion for modelling jointly bodily injury and property damage claim frequencies in Motor Third Party Liability (MTPL) insurance. This class is based on a mixing between two marginal Poisson distributions and a unit mean continuous prior, or mixing, distribution which belongs to a general distribution family including those which do not belong to the natural Exponential family and/or are not conjugate to the Poisson. Within the framework introduced here, both marginal mean parameters and the dispersion parameter of the two-dimensional response variable are modelled jointly as parametric functions of explanatory variables.

In what follows, we provide a detailed discussion of our contributions putting special emphasis on the suitability of the proposed family of models when dealing with MTPL claim count data in the bivariate setting, the Expectation-Maximisation (EM) type algorithm we developed for maximum likelihood (ML) estimation of the bivariate mixed Poisson model and practical application aspects in the context of a posteriori ratemaking.

Firstly, there are many factors in the MTPL insurance line that can simultaneously affect the joint dynamics of bodily injury and property damage claims which are conceivably positively correlated and may also lead to extra variation occurring in their associated claim counts. As empirical evidence has shown, these factors are observable variables concerning the policyholders and their vehicles and differences among policyholders (which cannot be observed by the actuary) and give rise to marginal overdispersion, which can be attributed to the excess of zeros and/or heavy upper tails, see Shared (1980), in MTPL bodily injury and property damage count data.

Moreover, as these factors vary from one country to another, significant differences are also observed in the frequency of MTPL bodily injury and property damage claims. For instance, the frequency of claims involving bodily injury for EU member states ranged from 0.13% in the Czech Republic through 0.98% in Italy to 1.28% in Turkey, see Insurance Europe (2019). Therefore, in order to capture the influence of risk factors and unobserved heterogeneity to a good approximation, it is important to have all the necessary due diligence in place when constructing a bivariate claim frequency regression model and ensure that it is a suitable candidate for modelling the relationship between MTPL bodily injury and property damage claims and a set of covariates.

Otherwise, a potential distribution misspecification and other issues, such as its failure to accurately account for the degree of dependency between the two MTPL claim types and/or its inability to accommodate marginal overdispersion per MTPL claim type, may result in biased and unreliable parameter estimates, which, in turn, can have a profound impact on how accurately insurers carry our different tasks such as pricing the policies and setting the appropriate level of reserves and reinsurance.

More importantly, due to the economic importance of MTPL insurance, it can be clearly understood that such shortcomings can subsequently lead to non-negligible financial implications for the company.

4.1.4 Importance of the proposed class of models

The family of bivariate mixed Poisson regression models with varying dispersion, which we present in this chapter, can efficiently capture the complex features of two-dimensional MTPL data. In particular, it allows for positive dependencies between bivariate responses, which is what we expect from this data, in very flexible manner since it is assumed that all the parameters of the bivariate mixed Poisson model can be modelled as functions of important risk factors. Moreover, this results in an improved risk evaluation since it allows to better quantify the extent of the impact of risk factors on the body and the tail areas of the marginal distributions, which might not necessarily be of the same magnitude, and allows to more effectively model the changes in the skewness of the marginal distributions because it depends on the marginal mean and the dispersion parameters.

Furthermore, as was previously mentioned, our model class permits for a variety of different distributional assumptions for the mixing density which measures the level of unobservable risk associated with each policy. Thus, since the thickness of the tail of a mixed Poisson distribution resembles that of its mixing density, our general approach can enable the actuary to fit more representative models that can match the tail behavior of MTPL bodily injury and property damage claim counts and hence can handle different levels of marginal overdispersion. For example, mixed Poisson models resulting from less heavy-tailed mixing densities have a more promising shape for zero and near zero values in the left tail area, whereas, those stemming from more heavy-tailed mixing distributions are more suitable for overdispersed claim counts with a long tail.

In this work, following the literature which is devoted to bivariate (and/or multivariate) mixed Poisson models, we emphasize the utility and generality of our approach by extending the setup of all the models we discussed above, namely the bivariate Negative Binomial (BNB), Bivariate Poisson-Inverse Gaussian (BPIG) and bivariate Poisson-

Lognormal (BPLN) models to allow for regression specifications on both of their mean parameters and their dispersion parameter.

Secondly, it is worth noting that the development of ML estimation procedures for joint modelling of all the parameters of mixed Poisson distributions in terms of covariate information remains a largely uncharted territory even within the univariate regression analysis context in majority of both statistical and actuarial applications.

In particular, regarding the statistical setting, this approach has only been explored so far by Rigby and Stasinopoulos (2005) and Barreto-Souza and Simas (2016). Rigby and Stasinopoulos (2005) proposed the generalized additive models for location, scale and shape (GAMLSS). The GAMLSS is a general regression framework which allows for every parameter of the distribution of discrete and/or continuous response distributions to be modelled as parametric and/or as additive nonparametric functions of covariates and/or random-effects terms including many well known univariate mixed Poisson distributions such as the Negative Binomial (NB) and Poisson-Inverse Gaussian (PIG) distributions and their zero-inflated versions for handling data sets that contain a large number of zeros.

The ML estimation of these regression type models can be carried out either by using the RS algorithm, which is based on the algorithm of Rigby and Stasinopoulos (1996a) and Rigby and Stasinopoulos (1996b), or the CG algorithm, which is based on the algorithm by Cole and Green (1992). Furthermore, Barreto-Souza and Simas (2016) used the EM algorithm for fitting a general family of mixed Poisson regression models with varying dispersion. In their application they focused on the estimation of the NB and PIG regression models with regression structures on both their mean and dispersion parameters.

Regarding the actuarial setting, Tzougas and Karlis (2020) implemented the EM algorithm for estimating the parameters of mixed Exponential regression models with varying dispersion, which can be used for approximating heavy-tailed losses in non-life insurance, and Tzougas (2020) employed the EM algorithm for ML estimation in the Poisson-Inverse Gamma (PIGA) regression model with varying dispersion, which can be regarded as a plausible model for deriving ratemaking mechanisms for heavy-tailed and overdispersed claim counts.

However, using the ML estimation procedure for the case when all the parameters of bivariate mixed Poisson distributions are allowed to vary through covariates has not yet been addressed in the statistical or actuarial literature. The reason for this is because the log-likelihood of the mixed Poisson model becomes more complicated in the two-dimensional setting and hence allowing for regressors on every parameter further increases the computational burden especially for the majority of members of the mixed Poisson family which, as is well known, have complicated densities that are either expressed in terms of special functions or cannot be written in a tractable closed form. Such examples are, for instance, the BPIG and BPLN models for which direct maximisation of their log-likelihoods via traditional optimization routines may suffer from computational instability and overflow issues when we allow for regressors on all their parameters.

4.1.5 Contribution to the literature

The main achievement of this chapter is that it demonstrates that ML estimation for our class of bivariate mixed Poisson regression models with varying dispersion can be accomplished via an efficient and easily implementable EM type algorithm which exploits the latent structure that is implied by the mixture representation of the bivariate mixed Poisson model and thus it reduces the problem of maximising its joint likelihood function to the problem of maximising the likelihood function of its mixing distribution. Moreover, the proposed algorithm can produce the information matrix of the bivariate mixed Poisson model as a by-product while it is computationally parsimonious.

Finally, following the setup of Bermúdez and Karlis (2017), the proposed class of mixed Poisson regression models with varying dispersion will be used within the Bayesian paradigm for deriving a posteriori ratemaking mechanisms, or Bonus-Malus Systems (BMSs). At this point we would like to call attention to the fact that, because the posterior claim frequency distribution is expressed in terms of both mean parameters and the dispersion parameter of the bivariate mixed Poisson model, using regressors on every parameter results in better risk adjusted a posteriori, or Bonus-Malus, premia. More importantly, since the motor insurance market is highly competitive, our family of models is well justified to be used in practice, as it can enable the actuary to set fair and equitable premia based on a sound risk measuring basis. These tailor-made to the risk involved premia are calculated based on the expected value and variance principles[1] providing the company with useful alternative tariff structures.

4.1.6 Outline

The remainder of this chapter proceeds as follows: in Section 4.2, we provide an in depth description of the proposed class of bivariate mixed Poisson regression models with varying dispersion. Also, we derive the joint probability mass function (jpmf) of the BNB, BPIG and BPLN regression models with varying dispersion. Section 4.3, describes the ML estimation via the EM algorithm. Furthermore, we consider detailed EM algorithms for the BNB, BPIG and BPLN regression models with varying dispersion. In Section 4.4 we explain how to calculate the a posteriori premia according to the expected value and variance principles. A real data application based on MTPL data

[1]Note that Lemaire (1995), Heilmann (1989), Gómez-Déniz et al. (2000) and Gómez-Déniz et al. (2002) used the variance principle for deriving BMSs in the univariate context based only on the a posteriori criteria, while Tzougas et al. (2018) proposed its use for developing such sytems based on both the a priori and the a posteriori criteria. The variance is an important risk measure and the difference in the Bonus-Malus premia that it implies can act as a cushion against adverse experience. However, the use of the variance principle for computing Bonus-Malus premia in a way that takes into consideration the positive correlation between MTPL bodily injury and property damage claims has not yet been proposed and thus this work expands on this setup as well.

is presented in Section 4.5. In Subsection 4.5.4, we describe the computational issues regarding the use of the EM algorithm for fitting the BNB, BPIG and BPLN regression models with varying dispersion. Finally, concluding remarks are given in Section 4.6.

4.2 Description of the bivariate mixed Poisson regression model with varying dispersion

The general class of bivariate mixed Poisson regression models with varying dispersion, which we consider in this chapter, can be described as follows. Assume that the individual claim frequencies $k_{i,j}$, where $i = 1$ denotes the MTPL bodily injury claims and $i = 2$ denotes the MTPL property damage claims, arising from a policyholder j, $j = 1, ..., n$, are independent per j and consider that given the random variables $\lambda_j > 0$, $k_{i,j}|\lambda_j$ per claim type $i = 1, 2$, are distributed according to a Poisson distribution with probability mass function (pmf) given by

$$P\left(k_{i,j}|\boldsymbol{x_{i,j}}, \boldsymbol{x_{3,j}}, \lambda_j\right) = \frac{\exp(-(\boldsymbol{\varepsilon_{i,j}}\lambda_j))(\boldsymbol{\varepsilon_{i,j}}\lambda_j)^{k_{i,j}}}{k_{i,j}!}, \tag{4.1}$$

for $k_{i,j} = 0, 1, 2, 3, ...$, where $\boldsymbol{\varepsilon_{i,j}}, \lambda_j > 0$ where $\mathbb{E}(k_{i,j}|\boldsymbol{x_{i,j}}, \boldsymbol{x_{3,j}}, \lambda_j) = \boldsymbol{\varepsilon_{i,j}}\lambda_j$ and $\mathrm{Var}(k_{i,j}|\boldsymbol{x_{i,j}}, \boldsymbol{x_{3,j}}, \lambda_j) = \boldsymbol{\varepsilon_{i,j}}\lambda_j$. Furthermore, suppose that λ_j are random variables from a continuous and at least twice differentiable mixing distribution with probability density function (pdf) $f(\lambda_j; \boldsymbol{\gamma_j})$, where we assume that $\mathbb{E}(\lambda_j) = 1$ as this ensures that the model is identifiable and where $\boldsymbol{\gamma_j} > 0$ is the dispersion parameter. Therefore, considering the previous assumptions, we can easily see that the unconditional distribution of $k_{i,j}$ is a bivariate mixed Poisson distribution with joint probability mass function (jpmf) given by

$$P\left(k_{1,j}, k_{2,j}|\boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \boldsymbol{x_{3,j}}\right) = \int_0^\infty \prod_{i=1}^{2} P\left(k_{i,j}|\lambda_j\right) f\left(\lambda_j; \boldsymbol{\gamma_j}\right) d\lambda_j. \tag{4.2}$$

To allow the two mean parameters and the dispersion parameter to be modelled in terms of explanatory variables with parametric linear functional forms we consider that

$$\boldsymbol{\varepsilon_{1,j}} = \exp(\boldsymbol{x_{1,j}^T \beta_1}) \tag{4.3}$$

$$\boldsymbol{\varepsilon_{2,j}} = \exp(\boldsymbol{x_{2,j}^T \beta_2}) \tag{4.4}$$

$$\boldsymbol{\gamma_j} = \exp(\boldsymbol{x_{3,j}^T \beta_3}) \tag{4.5}$$

where $\boldsymbol{x_{1,j}}$, $\boldsymbol{x_{2,j}}$ and $\boldsymbol{x_{3,j}}$ are vectors of covariates with dimensions $p_1 \times 1$, $p_2 \times 1$ and $p_3 \times 1$ respectively, with $(\beta_{1,1}, ..., \beta_{1,p_1})^T$, $(\beta_{2,1}, ..., \beta_{2,p_2})^T$ and $(\beta_{3,1}, ..., \beta_{3,p_3})^T$ the corresponding parameter vectors and where it is assumed that the matrices $\mathbf{X_1}$, $\mathbf{X_2}$ and $\mathbf{X_3}$, with rows given by $x_{1,i}$, $x_{2,i}$ and $x_{3,i}$ respectively, are of full rank. Finally, in context of MTPL insurance, the following useful properties associated with the bivariate mixed Poisson models with varying dispersion (which were already presented in Section 3.4 for the case where the dispersion parameter was not expressed in terms of covariates) are provided below.

1. The marginal distribution of $k_{i,j}$, for $i = 1, 2$, is the same mixed Poisson distribution as its bivariate counterpart. Also, the mean and the variance of $k_{i,j}$ are:

$$\mathbb{E}\left(k_{i,j}|\boldsymbol{x_{i,j}}\right) = \boldsymbol{\varepsilon_{i,j}} \tag{4.6}$$

$$\mathrm{Var}\left(k_{i,j}|\boldsymbol{x_{i,j}}, \boldsymbol{x_{3,j}}\right) = \boldsymbol{\varepsilon_{i,j}}\left[1 + \boldsymbol{\varepsilon_{i,j}}\mathrm{Var}(\lambda_j)\right]. \tag{4.7}$$

2. The covariance (Cov) and the correlation (Corr) between $k_{1,j}$ and $k_{2,j}$ are given by

$$\mathrm{Cov}\left(k_{1,j}, k_{2,j}|\boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \boldsymbol{x_{3,j}}\right) = \boldsymbol{\varepsilon_{1,j}\varepsilon_{2,j}}\mathrm{Var}(\lambda_j) \quad i = 1 \neq i = 2. \tag{4.8}$$

and

$$\text{Corr}\left(k_{1,j}, k_{2,j} | \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \boldsymbol{x_{3,j}}\right) = \frac{\text{Var}(\lambda_j)\sqrt{\boldsymbol{\varepsilon_{1,j}}\boldsymbol{\varepsilon_{2,j}}}}{\sqrt{\Big(1+\boldsymbol{\varepsilon_{1,j}}\text{Var}(\lambda_j)\Big)\Big(1+\boldsymbol{\varepsilon_{2,j}}\text{Var}(\lambda_j)\Big)}}. \tag{4.9}$$

3. The generalized variance ratio (GVR) between a bivariate mixed Poisson model with varying dispersion, i.e. $k_{i,j} \sim \text{Poisson}(\boldsymbol{\varepsilon_{i,j}}\lambda_j)$, with $\lambda_j \sim f(\lambda_j; \boldsymbol{\gamma_j})$ which is the pdf of the mixing density, and a simple Poisson model, i.e. $y_{i,j} \sim \text{Poisson}(\boldsymbol{\varepsilon_{i,j}})$, is given by

$$\text{GVR}\left(k_{1,j}, k_{2,j} | \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \boldsymbol{x_{3,j}}\right) = \frac{\sum_{i=1}^{2}\text{Var}\left(k_{i,j}\right) + 2\sum_{i<l}\text{Cov}\left(k_{i,j}, k_{l,j}\right)}{\sum_{i=1}^{2}\text{Var}\left(y_{i,j}\right)}$$

$$= 1 + \text{Var}(\lambda_j)\sum_{i=1}^{2}\boldsymbol{\varepsilon_{i,j}}. \tag{4.10}$$

As it can be seen from Eqs. (4.9 and 4.10), $\text{Corr}\left(k_{1,j}, k_{2,j} | \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \boldsymbol{x_{3,j}}\right) > 0$ and $\text{GVR}\left(k_{1,j}, k_{2,j} | \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \boldsymbol{x_{3,j}}\right) > 1$. Also, the GVR increases as the variance of the mixing distribution increases. Thus, as was previously mentioned, the bivariate mixed Poisson regression model allows for the positive correlation between the MTPL bodily injury and property damage claims and accommodates overdispersion. Consequently, as highlighted before, our models cannot capture underdispersion and negative dependencies, therefore if dealing with a data set which exhibits these characteristics, other models such as the ones by Fung et al. (2019a) and Fung et al. (2019b) are more advisable.

In what follows, different bivariate mixed Poisson distributions with regression structures on every parameter are used to describe the behaviour of the number of bodily injury and property damage claims as a function of the explanatory variables including the bivariate Negative Binomial (BNB), bivariate Poisson-Inverse Gaussian (BPIG) and bivariate Poisson-Lognormal (BPLN) distributions. It must be noted that the BNB and BPIG models were already briefly presented in Section 3.4 for comparative purposes with the BNBIG model, but there the common heterogeneity factor could only be expressed in terms of a mixing distribution (i.e. the Gamma in the case of the BNB and the Inverse Gaussian in case of the BPIG), while here it is a regression itself.

4.2.1 BNB regression model with varying dispersion

Let λ_j follow a Gamma distribution with a pdf

$$f(\lambda_j; \boldsymbol{\gamma_j}) = \frac{\boldsymbol{\gamma_j}^{\gamma_j}}{\Gamma(\boldsymbol{\gamma_j})} \lambda_j^{\gamma_j - 1} \exp(-\boldsymbol{\gamma_j} \lambda_j), \tag{4.11}$$

where $\boldsymbol{\gamma_j}, \lambda_j > 0$, with mean and variance:

$$\mathbb{E}(\lambda_j) = 1$$

$$\text{Var}(\lambda_j) = 1/\boldsymbol{\gamma_j}, \tag{4.12}$$

for $j = 1, ..., n$. Thus, based on Eqs. (4.1 and 4.11) it is easy to see that the resulting distribution is the BNB distribution with jpmf

$$P(k_{1,j}, k_{2,j} | \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \boldsymbol{x_{3,j}}) = \frac{\Gamma(\boldsymbol{\gamma_j} + \sum_{i=1}^{2} k_{i,j})}{\Gamma(\boldsymbol{\gamma_j}) \prod_{i=1}^{2} k_{i,j}!} \frac{\boldsymbol{\gamma_j}^{\gamma_j} \prod_{i=1}^{2} (\boldsymbol{\varepsilon_{i,j}})^{k_{i,j}}}{(\boldsymbol{\gamma_j} + \boldsymbol{\varepsilon_{i,j}})^{\gamma_j + \sum_{i=1}^{2} k_{i,j}}}. \tag{4.13}$$

4.2.2 BPIG regression model with varying dispersion

Let λ_j follow an Inverse Gaussian distribution with a pdf of the form

$$f(\lambda_j; \boldsymbol{\gamma_j}) = \frac{\boldsymbol{\gamma_j}}{\sqrt{2\pi}} \lambda_j^{-3/2} \exp\left[\boldsymbol{\gamma_j^2} - \frac{\boldsymbol{\gamma_j^2}}{2}\left(\frac{1}{\lambda_j} + \lambda_j\right)\right], \tag{4.14}$$

where $\boldsymbol{\gamma_j}, \lambda_j > 0$, with mean and variance:

$$\mathbb{E}(\lambda_j) = 1$$

$$\text{Var}(\lambda_j) = 1/\boldsymbol{\gamma_j}^2, \tag{4.15}$$

for $j = 1, ..., n$. Therefore, considering the assumptions in Eqs. (4.1 and 4.14) it can be verified that the resulting distribution is the BPIG distribution with jpmf

$$P(k_{1,j}, k_{2,j}|\boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \boldsymbol{x_{3,j}}) = \frac{2\boldsymbol{\gamma_j} \exp(\boldsymbol{\gamma_j}^2)}{\sqrt{2\pi}} K_{\sum_{i=1}^{2} k_{i,j} - \frac{1}{2}}(\boldsymbol{\gamma_j}\Delta_j) \left(\frac{\boldsymbol{\gamma_j}}{\Delta_j}\right)^{\sum_{i=1}^{2} k_{i,j} - \frac{1}{2}} \prod_{i=1}^{2} \frac{\boldsymbol{\varepsilon_{i,j}}^{k_{i,j}}}{k_{i,j}!}, \tag{4.16}$$

where $\Delta_j = \sqrt{\boldsymbol{\gamma_j}^2 + 2\sum_{i=1}^{2} \boldsymbol{\varepsilon_{i,j}}}$ and $K_\nu(\omega)$ denotes the modified Bessel function of the third kind of order ν and argument ω.

4.2.3 BPLN regression model with varying dispersion

Let λ_j follow an Lognormal distribution with a pdf of the form

$$f(\lambda_j; \boldsymbol{\gamma_j}) = \frac{\exp\left[-\frac{(\log(\lambda_j) + \boldsymbol{\gamma_j}^2/2)^2}{2\boldsymbol{\gamma_j}^2}\right]}{\sqrt{2\pi}\boldsymbol{\gamma_j}\lambda_j}, \tag{4.17}$$

where $\boldsymbol{\gamma_j}, \lambda_j > 0$, with mean and variance:

$$\mathbb{E}(\lambda_j) = 1$$

$$\text{Var}(\lambda_j) = \exp(\boldsymbol{\gamma_j}^2) - 1, \tag{4.18}$$

for $j = 1, ..., n$. Thus, based on Eqs. (4.1 and 4.17) it is easy to see that the resulting distribution is the BNB distribution with jpmf

$$P(k_{1,j}, k_{2,j} | \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \boldsymbol{x_{3,j}}) = \int_0^\infty \prod_{i=1}^{2} \frac{\exp(-(\boldsymbol{\varepsilon_{i,j}}\lambda_j))(\boldsymbol{\varepsilon_{i,j}}\lambda_j)^{k_{i,j}}}{k_{i,j}!} \frac{\exp\left[-\frac{(\log(\lambda_j)+\boldsymbol{\gamma_j}^2/2)^2}{2\boldsymbol{\gamma_j}^2}\right]}{\sqrt{2\pi}\boldsymbol{\gamma_j}\lambda_j} d\lambda_j, \tag{4.19}$$

which could not be written in closed form and hence numerical integration is required.

4.3 The EM algorithm

In this section we describe how the EM algorithm (see, Dempster et al. (1977) and McLachlan and Krishnan (2007)) can be employed for facilitating ML estimation of the parameters of the bivariate mixed Poisson regression model for marginal means and dispersion which was described in Section 4.2.

Let $(k_{1,j}, k_{2,j}, \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \boldsymbol{x_{3,j}})$, $j = 1, ..., n$, be a sample of independent observations, where $k_{1,j}$ and $k_{2,j}$ are the response variables and $\boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}$ and $\boldsymbol{x_{3,j}}$ are the vectors of covariate information with dimensions $p_1 \times 1$, $p_2 \times 1$ and $p_3 \times 1$ respectively. Also, suppose that the data are produced according to the bivariate mixed Poisson model: then, the log-likelihood of the model can be written as

$$l(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log\left(P\left(k_{1,j}, k_{2,j} | \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \boldsymbol{x_{3,j}}\right)\right), \tag{4.20}$$

where $\boldsymbol{\theta} = \left(\boldsymbol{\beta_1^T}, \boldsymbol{\beta_2^T}, \boldsymbol{\beta_3^T}\right)^T$ is the vector of the parameters and where $P\left(k_{1,j}, k_{2,j} | \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \boldsymbol{x_{3,j}}\right)$ is the jpmf of the bivariate mixed Poisson model, which is given by Eq. (4.2).

Direct maximisation of Eq. (4.20) with respect to the vector of parameters $\boldsymbol{\theta}$ is cumbersome because the log-likelihood of the bivariate mixed Poisson model is not usually tractable. Moreover, when both mean parameters and the dispersion parameter are modelled as functions of explanatory variables this raises additional computational challenges.

Fortunately, ML estimation can be accomplished relatively easily via an EM type algorithm which is specifically tailored to ML estimation for univariate and bivariate (and/or multivariate) mixed Poisson models (see, for instance, Karlis (2001) and Karlis (2005), Ghitany et al. (2012), Barreto-Souza and Simas (2016) and Tzougas (2020)) since their stochastic mixture representation involving a non-observable random variable, denoted by λ_j herein, can be considered to produce missing data. In our case, if one augments the unobserved data λ_j to the observed data $(k_{1,j}, k_{2,j}, \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \boldsymbol{x_{3,j}})$, then the complete data log-likelihood decomposes into two parts

$$l_c(\theta) = \sum_{i=1}^{2} \sum_{j=1}^{n} \left[-\boldsymbol{\varepsilon_{i,j}} \lambda_j + k_{i,j} \log(\boldsymbol{\varepsilon_{i,j}} \lambda_j) - \log(k_{i,j}!) \right] + \sum_{j=1}^{n} \log\left(f\left(\lambda_j; \boldsymbol{\gamma_j}\right)\right), \qquad (4.21)$$

for $i = 1, 2$ and $j = 1, ..., n$, where $f(\lambda_j; \boldsymbol{\gamma_j})$ is the pdf of the mixing distribution and where $\boldsymbol{\varepsilon_{i,j}}$ and $\boldsymbol{\gamma_j}$ are given by Eqs. (4.3, 4.4 and 4.5) respectively. The E- and the M-Steps of our EM type algorithm procedure for the bivariate mixed Poisson regression model with varying dispersion are described below, including a few comments for each step.

- **E-Step:**

  The Q-function, which is the conditional expectation of the complete data log-likelihood in Eq. (4.21), is calculated in a general way so as to elucidate its features for our general class of bivariate mixed Poisson regression models with varying dispersion

$$
\begin{aligned}
Q\left(\boldsymbol{\theta};\boldsymbol{\theta}^{(r)}\right) &\equiv \mathbb{E}_{\lambda_j}\left(l_c\left(\boldsymbol{\theta}\right)|k_{i,j},\boldsymbol{x_{i,j}},\boldsymbol{x_{3,j}};\boldsymbol{\theta}^{(r)}\right) \propto \\
&\propto \sum_{i=1}^{2}\sum_{j=1}^{n}\left[-\boldsymbol{\varepsilon_{i,j}}\mathbb{E}_{\lambda_j}\left[\lambda_j|k_{i,j},\boldsymbol{x_{i,j}},\boldsymbol{x_{3,j}};\boldsymbol{\theta}^{(r)}\right]+k_{i,j}\log\left(\boldsymbol{\varepsilon_{i,j}}\right)\right] \\
&\quad +\sum_{j=1}^{n}\mathbb{E}_{\lambda_j}\left[\log\left(f\left(\lambda_j;\boldsymbol{\gamma_j}^{(r)}\right)\right)\right],
\end{aligned}
$$

where $\boldsymbol{\theta}^{(r)}$ is the estimate of $\boldsymbol{\theta}$ at the r-th iteration in the E-Step of our EM algorithm. Then, compute the pseudo-values $w_j = \mathbb{E}_{\lambda_j}\left[\lambda_j|k_{i,j},\boldsymbol{x_{i,j}},\boldsymbol{x_{3,j}};\boldsymbol{\theta}^{(r)}\right]$ and $\omega_{k,j} = \mathbb{E}_{\lambda_j}\left[s_k\left(\lambda_j\right)|k_{i,j},\boldsymbol{x_{i,j}},\boldsymbol{x_{3,j}};\boldsymbol{\theta}^{(r)}\right]$, for $i = 1,2,\ j = 1,..,n$ and $k = 1,....,\nu$, where $s_k\left(.\right)$ are certain functions[2] which are involved in the terms needed for maximising the part of the Q-function which corresponds to the conditional expectation of the log-likelihood of $f\left(\lambda_j;\boldsymbol{\gamma_j}\right)$.

- **M-Step:**

  Using the pseudo-values w_j and $\omega_{k,j}$ from the E-Step and the Newton-Raphson algorithm three times[3], find the maximum global point $\boldsymbol{\theta}^{(r+1)}$ of the Q-function, i.e. obtain the updated estimates $\boldsymbol{\beta_1}^{(r+1)}$,$\boldsymbol{\beta_2}^{(r+1)}$ and $\boldsymbol{\beta_3}^{(r+1)}$.

  – Firstly, differentiating the Q-function with respect to $\boldsymbol{\beta_1}$ gives:

$$
h_1\left(\boldsymbol{\beta_1}\right) = \frac{\partial Q\left(\boldsymbol{\theta};\boldsymbol{\theta}^{(r)}\right)}{\partial\beta_{1,l}} = \sum_{j=1}^{n}\left(k_{1,j}-\boldsymbol{\varepsilon}_{\boldsymbol{1,j}}^{(r)}w_j\right)x_{1,j,l}, \tag{4.22}
$$

[2]Note that, as it will be demonstrated in what follows, if $s_k\left(\lambda_j\right)$ is a linear function, then the conditional posterior expectations can be computed in an easy and accurate way. However, for more complicated functions, for which an exact solution is not available, one can use Taylor approximations, or numerical approximations, including numerical integration, and/or simulation based approximations.

[3]Note also that this procedure can be used for every continuous and at least twice differentiable mixing distribution, i.e. similar to those we considered in this work. Therefore, we provide a complete estimation tool for our class of bivariate mixed Poisson regression models with varying dispersion. However, for some other mixing distributions a special iterative scheme or another EM algorithm inside the M-Step may be more appropriate.

$$H_1\left(\boldsymbol{\beta_1}\right)=\frac{\partial^2 Q\left(\boldsymbol{\theta};\boldsymbol{\theta}^{(r)}\right)}{\partial\beta_{1,l}\partial\beta_{1,l}^T}=\sum_{j=1}^{n}\left(-\boldsymbol{\varepsilon}_{\mathbf{1,j}}^{(r)}w_j\right)x_{1,j,l}x_{1,j,l}^T=\mathbf{X_1^T W_1 X_1}, \tag{4.23}$$

for $j=1,...,n$ and $l=1,...,p_1$ and where $\mathbf{W_1}=\text{diag}\{-\boldsymbol{\varepsilon}_{\mathbf{1,j}}^{(r)}w_j\}$. Then, the iterative procedure for the Newton-Raphson algorithm for $\boldsymbol{\beta_1}$ is as follows:

$$\boldsymbol{\beta}_{\mathbf{1}}^{(r+1)}\equiv\boldsymbol{\beta}_{\mathbf{1}}^{(r)}-\left[H_1\left(\boldsymbol{\beta}_{\mathbf{1}}^{(r)}\right)\right]^{-1}h_1\left(\boldsymbol{\beta}_{\mathbf{1}}^{(r)}\right). \tag{4.24}$$

– Secondly, differentiating the Q-function with respect to $\boldsymbol{\beta_2}$ gives:

$$h_2\left(\boldsymbol{\beta_2}\right)=\frac{\partial Q\left(\boldsymbol{\theta};\boldsymbol{\theta}^{(r)}\right)}{\partial\beta_{2,l}}=\sum_{j=1}^{n}\left(k_{2,j}-\boldsymbol{\varepsilon}_{\mathbf{2,j}}^{(r)}w_j\right)x_{2,j,l}, \tag{4.25}$$

$$H_2\left(\beta_2\right)=\frac{\partial^2 Q\left(\theta;\theta^{(r)}\right)}{\partial\beta_{2,l}\partial\beta_{2,l}^T}=\sum_{j=1}^{n}\left(-\boldsymbol{\varepsilon}_{\mathbf{2,j}}^{(r)}w_j\right)x_{2,j,l}x_{2,j,l}^T=\mathbf{X_2^T W_2 X_2}, \tag{4.26}$$

for $j=1,...,n$ and $l=1,...,p_2$ and where $\mathbf{W_2}=\text{diag}\{-\boldsymbol{\varepsilon}_{\mathbf{2,j}}^{(r)}w_j\}$. Then, the iterative procedure for the Newton-Raphson algorithm for $\boldsymbol{\beta_2}$ is as follows:

$$\boldsymbol{\beta}_{\mathbf{2}}^{(r+1)}\equiv\boldsymbol{\beta}_{\mathbf{2}}^{(r)}-\left[H_2\left(\boldsymbol{\beta}_{\mathbf{2}}^{(r)}\right)\right]^{-1}h_2\left(\boldsymbol{\beta}_{\mathbf{2}}^{(r)}\right). \tag{4.27}$$

– Thirdly, differentiating the Q-function with respect to $\boldsymbol{\beta_3}$ gives

$$h_3\left(\boldsymbol{\beta_3}\right)=\frac{\partial Q\left(\boldsymbol{\theta};\boldsymbol{\theta}^{(r)}\right)}{\partial\beta_{3,l}}=\sum_{j=1}^{n}\frac{\partial\mathbb{E}_{\lambda_j}\left[\log\left(f\left(\lambda_j;\boldsymbol{\gamma}_{\boldsymbol{j}}^{(r)}\right)\right)\right]}{\partial\beta_{3,l}} \tag{4.28}$$

$$H_3\left(\boldsymbol{\beta_3}\right)=\frac{\partial^2 Q\left(\boldsymbol{\theta};\boldsymbol{\theta}^{(r)}\right)}{\partial\beta_{3,l}\partial\beta_{3,j}^T}=\sum_{j=1}^{n}\frac{\partial^2\mathbb{E}_{\lambda_j}\left[\log\left(f\left(\lambda_j;\boldsymbol{\gamma}_{\boldsymbol{j}}^{(r)}\right)\right)\right]}{\partial\beta_{3,l}\partial\beta_{3,l}^T}, \tag{4.29}$$

where for calculating $h_3\left(\boldsymbol{\beta_3}\right)$ and $H_3\left(\boldsymbol{\beta_3}\right)$ one needs to use the pseudo-values $\omega_{k,j}$ for $j=1,..,n$ and $k=1,....,\nu$ since in this case the maximisation of the Q-function reduces to the maximisation of the conditional expectation of the log-likelihood of $f(\lambda_j;\boldsymbol{\gamma_j})$. Then, the Newton-Raphson iterative algorithm for $\boldsymbol{\beta_3}$ is as follows:

$$\boldsymbol{\beta_3}^{(r+1)} \equiv \boldsymbol{\beta_3}^{(r)} - \left[H_3\left(\boldsymbol{\beta_3}^{(r)}\right)\right]^{-1} h_3\left(\boldsymbol{\beta_3}^{(r)}\right), \tag{4.30}$$

for $j = 1, .., n$ and $l = 1, ..., p_3$.

- Finally, iterate between the E- and the M-Steps until some convergence criterion is satisfied, for example

$$\left|\frac{l^{(r+1)} - l^{(r)}}{l^{(r)}}\right| < tol,$$

where $l^{(r)}$ is the value of the log-likelihood after the r-th iteration and where tol is a small number usually of the form 10^{-m}, where $m \in \mathbb{Z}$. If this stopping criterion which refers to the progress of the likelihood function, i.e., its convergence, is satisfied, the EM algorithm stops iterating and the estimate of $\boldsymbol{\theta}$ is $\boldsymbol{\theta}^{(r+1)}$. Otherwise, $\boldsymbol{\theta}$ is updated by $\boldsymbol{\theta}^{(r+1)}$ and the algorithm goes back to the E-Step.

- Note that when the regression specifications for both mean parameters and the dispersion parameter of the model are limited to the constants $\beta_{1,0}$, $\beta_{2,0}$ and $\beta_{3,0}$, this EM type algorithm can be employed for the ML estimation of the "univariate", without regression components, model. In what follows, we describe in detail the E- and the M-Steps of our EM type algorithm for the BNB, BPIG and BPLN regression models with varying dispersion.

4.3.1 BNB regression model with varying dispersion

In the case of the Gamma mixing distribution with pdf given by Eq. (4.11) we have that the posterior distribution of $\lambda_j|k_{i,j}, \boldsymbol{x_{i,j}}, \boldsymbol{x_{3,j}}; \boldsymbol{\theta}$ is a Gamma with parameters $\boldsymbol{\gamma_j} + \sum_{i=1}^{2} k_{i,j}$ and $\boldsymbol{\gamma_j} + \sum_{i=1}^{2} \boldsymbol{\varepsilon_{i,j}}$, for $i = 1, ..., n$.

Then, the EM algorithm goes as follows:

- **E-Step:**

  Calculate for all $j = 1, .., n$,

$$w_j = \mathbb{E}_{\lambda_j}\left[\lambda_j | k_{i,j}, \boldsymbol{x_{i,j}}, \boldsymbol{x_{3,j}}; \boldsymbol{\theta}^{(r)}\right] = \frac{\boldsymbol{\gamma_j^{(r)}} + \sum_{i=1}^{2} k_{i,j}}{\boldsymbol{\gamma_j^{(r)}} + \sum_{i=1}^{2} \boldsymbol{\varepsilon_{i,j}^{(r)}}} \tag{4.31}$$

$$\omega_j = \mathbb{E}_{\lambda_j}\left[\log(\lambda_j) | k_{i,j}, \boldsymbol{x_{i,j}}, \boldsymbol{x_{3,j}}; \boldsymbol{\theta}^{(r)}\right] = \Psi\left(\boldsymbol{\gamma_j^{(r)}} + \sum_{i=1}^{2} k_{i,j}\right) - \log\left(\boldsymbol{\gamma_j^{(r)}} + \sum_{i=1}^{2} \boldsymbol{\varepsilon_{i,j}^{(r)}}\right), \tag{4.32}$$

  where $\Psi(\cdot)$ is the digamma function and where $\boldsymbol{\varepsilon_{1,j}^{(r)}} = \exp\left(\boldsymbol{x_{1,j}^T}\boldsymbol{\beta_1^{(r)}}\right)$, $\boldsymbol{\varepsilon_{2,j}^{(r)}} = \exp\left(\boldsymbol{x_{2,j}^T}\boldsymbol{\beta_2^{(r)}}\right)$ and $\boldsymbol{\gamma_j^{(r)}} = \exp\left(\boldsymbol{x_{3,j}^T}\boldsymbol{\beta_3^{(r)}}\right)$ are the estimates obtained after r-th iteration.

- **M-Step**:

  - Update the regression parameters $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$ using the pseudo-values w_j, which are given by Eq. (4.31), and the Newton-Raphson algorithms in Eqs. (4.22, 4.23 and 4.24) and Eqs. (4.25, 4.26 and 4.27) respectively.

  - Update the regression parameters $\boldsymbol{\beta_3}$ using the pseudo-values w_j and ω_j, which are given by Eqs. (4.31 and 4.32) respectively, and the Newton-Raphson algorithm which, in the case of the Gamma mixing distribution, is as follows

$$h_3(\boldsymbol{\beta_3}) = \boldsymbol{\gamma_j^{(r)}}\left[\log\left(\boldsymbol{\gamma_j^{(r)}}\right) - \Psi\left(\boldsymbol{\gamma_j^{(r)}}\right) - w_j + \omega_j + 1\right] x_{3,j,l}, \tag{4.33}$$

$$
\begin{aligned}
H_3\left(\boldsymbol{\beta_3}\right) &= \sum_{j=1}^{n} \boldsymbol{\gamma_j}^{(r)} \left[\log\left(\boldsymbol{\gamma_j}^{(r)}\right) - \Psi\left(\boldsymbol{\gamma_j}^{(r)}\right) - w_j + \omega_j \right. \\
& \left. -\Psi_3\left(\boldsymbol{\gamma_j}^{(r)}\right)\boldsymbol{\gamma_j}^{(r)} + 2\right] x_{3,j,l} x_{3,j,l}^T \\
&= \mathbf{X_3}^T \mathbf{W_3} \mathbf{X_3}, \qquad (4.34)
\end{aligned}
$$

for $j = 1, .., n$ and $l = 1, ..., p_3$, where $\Psi_3(.)$ is the trigamma function and where
$\mathbf{W_3} = \text{diag}\{\boldsymbol{\gamma_j}^{(r)} \log\left(\boldsymbol{\gamma_j}^{(r)}\right) - \boldsymbol{\gamma_j}^{(r)} \Psi\left(\boldsymbol{\gamma_j}^{(r)}\right) - \boldsymbol{\gamma_j}^{(r)} w_j + \boldsymbol{\gamma_j}^{(r)} \omega_j - \Psi_3\left(\boldsymbol{\gamma_j}^{(r)}\right)\left(\boldsymbol{\gamma_j}^2\right)^{(r)} + 2\boldsymbol{\gamma_j}^{(r)}\}$.
Then, we can obtain the updated estimates of $\boldsymbol{\beta_3}^{(r)}$ using Eq. (4.30).

4.3.2 BPIG regression model with varying dispersion

In the case of the Inverse Gaussian mixing distribution with pdf given by Eq. (4.14), we have that the posterior distribution of $\lambda_j | k_{i,j}, \boldsymbol{x_{i,j}}, \boldsymbol{x_{3,j}}; \boldsymbol{\theta}$ is a Generalized Inverse Gaussian (GIG) distribution with pdf

$$
f\left(\lambda_j | k_{i,j}, \boldsymbol{x_{i,j}}, \boldsymbol{x_{3,j}}; \boldsymbol{\theta}\right) = \frac{\left(\frac{\psi_j}{\chi_j}\right)^{\nu_j}}{2K_{\nu_j}\left(\psi_j \chi_j\right)} \lambda_j^{\nu_j - 1} \exp\left[-\frac{1}{2}\left(\frac{\chi_j^2}{\lambda_j} + \psi_j^2 \lambda_j\right)\right], \qquad (4.35)
$$

where $\psi_j = \sqrt{\boldsymbol{\gamma_j}^2 + 2\sum_{i=1}^{2} \boldsymbol{\varepsilon_{i,j}}} > 0$, $\chi_j = \boldsymbol{\gamma_j} > 0$, and $\nu_j = \sum_{i=1}^{2} k_{i,j} - \frac{1}{2} \in \Re$ for $j = 1, ..., n$.

Then, the EM algorithm is as follows:

- **E-Step:**

  Calculate for all $j = 1, .., n$,

$$w_j = \mathbb{E}_{\lambda_j}\left[\lambda_j | k_{i,j}, \boldsymbol{x_{i,j}}, \boldsymbol{x_{3,j}}; \boldsymbol{\theta}^{(r)}\right]$$

$$= \frac{\boldsymbol{\gamma_j}^{(r)}}{\sqrt{\left(\boldsymbol{\gamma_j^2}\right)^{(r)} + 2\sum_{i=1}^{2}\boldsymbol{\varepsilon}_{\boldsymbol{1,j}}^{(r)}}} \frac{K_{\sum_{i=1}^{2} k_{i,j}+\frac{1}{2}}\left(\boldsymbol{\gamma_j}^{(r)}\sqrt{\left(\boldsymbol{\gamma_j^2}\right)^{(r)} + 2\sum_{i=1}^{2}\boldsymbol{\varepsilon}_{\boldsymbol{i,j}}^{(r)}}\right)}{K_{\sum_{i=1}^{2} k_{i,j}-\frac{1}{2}}\left(\boldsymbol{\gamma_j}^{(r)}\sqrt{\left(\boldsymbol{\gamma_j^2}\right)^{(r)} + 2\sum_{i=1}^{2}\boldsymbol{\varepsilon}_{\boldsymbol{i,j}}^{(r)}}\right)} \tag{4.36}$$

  and

$$\omega_j = \mathbb{E}_{\lambda_j}\left[\frac{1}{\lambda_j} | k_{i,j}, \boldsymbol{x_{i,j}}, \boldsymbol{x_{3,j}}; \boldsymbol{\theta}^{(r)}\right]$$

$$= \frac{\sqrt{\left(\boldsymbol{\gamma_j^2}\right)^{(r)} + 2\sum_{i=1}^{2}\boldsymbol{\varepsilon}_{\boldsymbol{i,j}}^{(r)}}}{\boldsymbol{\gamma_j}^{(r)}} \frac{K_{\sum_{i=1}^{2} k_{i,j}-\frac{3}{2}}\left(\boldsymbol{\gamma_j}^{(r)}\sqrt{\left(\boldsymbol{\gamma_j^2}\right)^{(r)} + 2\sum_{i=1}^{2}\boldsymbol{\varepsilon}_{\boldsymbol{i,j}}^{(r)}}\right)}{K_{\sum_{i=1}^{2} k_{i,j}-\frac{1}{2}}\left(\boldsymbol{\gamma_j}^{(r)}\sqrt{\left(\boldsymbol{\gamma_j^2}\right)^{(r)} + 2\sum_{i=1}^{2}\boldsymbol{\varepsilon}_{\boldsymbol{i,j}}^{(r)}}\right)}, \tag{4.37}$$

  where $\boldsymbol{\varepsilon}_{\boldsymbol{1,j}}^{(r)} = \exp\left(\boldsymbol{x}_{\boldsymbol{1,j}}^T \boldsymbol{\beta}_{\boldsymbol{1}}^{(r)}\right)$, $\boldsymbol{\varepsilon}_{\boldsymbol{2,j}}^{(r)} = \exp\left(\boldsymbol{x}_{\boldsymbol{2,j}}^T \boldsymbol{\beta}_{\boldsymbol{2}}^{(r)}\right)$ and $\boldsymbol{\gamma_j}^{(r)} = \exp\left(\boldsymbol{x}_{\boldsymbol{3,j}}^T \boldsymbol{\beta}_{\boldsymbol{3}}^{(r)}\right)$ are the estimates obtained after r-th iteration.

- **M-Step**:

  - Update the regression parameters $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$ using the pseudo-values w_j, which are given by Eq. (4.36), and the Newton-Raphson algorithms in Eqs. (4.22, 4.23 and 4.24) and Eqs. (4.25, 4.26 and 4.27) respectively.
  - Update the regression parameters $\boldsymbol{\beta_3}$ using the pseudo-values w_j and ω_j, which are given by Eqs. (4.36 and 4.37) respectively, and the Newton-Raphson algorithm which, in the case of the Inverse Gaussian mixing dis-

tribution, is as follows

$$h_3\left(\boldsymbol{\beta_3}\right)=\left[2\left(\boldsymbol{\gamma_j^2}\right)^{(r)}-w_j\left(\boldsymbol{\gamma_j^2}\right)^{(r)}-\omega_j\left(\boldsymbol{\gamma_j^2}\right)^{(r)}+1\right]x_{3,j,l}, \tag{4.38}$$

$$\begin{aligned} H_3\left(\boldsymbol{\beta_3}\right) &= \sum_{j=1}^{n}\left[\left(\boldsymbol{\gamma_j^2}\right)^{(r)}-2w_j\left(\boldsymbol{\gamma_j^2}\right)^{(r)}-2\omega_j\left(\boldsymbol{\gamma_j^2}\right)^{(r)}\right]x_{3,j,l}x_{3,j,l}^T \\ &= \mathbf{X_3^T W_3 X_3}, \end{aligned} \tag{4.39}$$

for $j=1,..,n$ and $l=1,...,p_3$, where
$\mathbf{W_3}=\text{diag}\{\left(\boldsymbol{\gamma_j^2}\right)^{(r)}-2w_j\left(\boldsymbol{\gamma_j^2}\right)^{(r)}-2\omega_j\left(\boldsymbol{\gamma_j^2}\right)^{(r)}\}$.
Then, we can obtain the updated estimates of $\boldsymbol{\beta_3^{(r)}}$ using Eq. (4.30).

4.3.3 BPLN regression model with varying dispersion

The EM algorithm can also be employed to find the ML estimates of the BPLN model which was defined in Eq. (4.19). In this case, the complete data log-likelihood takes the form:

$$\begin{aligned} l_c\left(\boldsymbol{\theta}\right) &= \sum_{i=1}^{2}\sum_{j=1}^{n}\left[-\boldsymbol{\varepsilon_{i,j}}\lambda_j+k_{i,j}\log\left(\boldsymbol{\varepsilon_{i,j}}\lambda_j\right)-\log\left(k_{i,j}!\right)\right]+ \\ & \sum_{j=1}^{n}\left[-\frac{1}{2}\log\left(2\pi\right)-\log\left(\boldsymbol{\gamma_j}\right)-\log\left(\lambda_j\right)-\frac{\left(\log\left(\lambda_j\right)+\frac{\boldsymbol{\gamma_j^2}}{2}\right)^2}{2\boldsymbol{\gamma_j^2}}\right], \end{aligned} \tag{4.40}$$

for $i=1,2$ and $j=1,...,n$. Thus, the expectations needed for the M-Step are $\mathbb{E}_{\lambda_j}\left[\lambda_j|k_{i,j},\boldsymbol{x_{i,j}},\boldsymbol{x_{3,j}};\boldsymbol{\theta}^{(r)}\right]$ and $\mathbb{E}_{\lambda_j}\left[\left(\log\left(\lambda_j\right)\right)^2|k_{i,j},\boldsymbol{x_{i,j}},\boldsymbol{x_{3,j}};\boldsymbol{\theta}^{(r)}\right]$.

Therefore, the algorithm can be written as follows:

- **E-Step:**

  Calculate for all $j = 1, .., n$,

$$w_j = \mathbb{E}_{\lambda_j}\left[\lambda_j | k_{i,j}, \boldsymbol{x_{i,j}}, \boldsymbol{x_{3,j}}; \boldsymbol{\theta}^{(r)}\right]$$

$$= \frac{\int_0^\infty \lambda_j \prod_{i=1}^{2} \frac{\exp(-\boldsymbol{\varepsilon}_{\boldsymbol{i,j}}^{(r)}\lambda_j)\left(\boldsymbol{\varepsilon}_{\boldsymbol{i,j}}^{(r)}\lambda_j\right)^{k_{i,j}}}{k_{i,j}!} \frac{\exp\left[-\frac{\left(\log\left(\lambda_j\right)+\frac{\left(\boldsymbol{\gamma_j^2}\right)^{(r)}}{2}\right)^2}{2\left(\boldsymbol{\gamma_j^2}\right)^{(r)}}\right]}{\sqrt{2\pi}\boldsymbol{\gamma_j}^{(r)}\lambda_j} d\lambda_j}{\int_0^\infty \prod_{i=1}^{2} \frac{\exp(-\boldsymbol{\varepsilon}_{\boldsymbol{i,j}}^{(r)}\lambda_j)\left(\boldsymbol{\varepsilon}_{\boldsymbol{i,j}}^{(r)}\lambda_j\right)^{k_{i,j}}}{k_{i,j}!} \frac{\exp\left[-\frac{\left(\log\left(\lambda_j\right)+\frac{\left(\boldsymbol{\gamma_j^2}\right)^{(r)}}{2}\right)^2}{2\left(\boldsymbol{\gamma_j^2}\right)^{(r)}}\right]}{\sqrt{2\pi}\boldsymbol{\gamma_j}^{(r)}\lambda_j} d\lambda_j}, \tag{4.41}$$

$$\omega_j = \mathbb{E}_{\lambda_j}\left[\left(\log\left(\lambda_j\right)\right)^2 | k_{i,j}, \boldsymbol{x_{i,j}}, \boldsymbol{x_{3,j}}; \boldsymbol{\theta}^{(r)}\right]$$

$$= \frac{\int_0^\infty \left(\log\left(\lambda_j\right)\right)^2 \prod_{i=1}^{2} \frac{\exp(-\boldsymbol{\varepsilon}_{\boldsymbol{i,j}}^{(r)}\lambda_j)\left(\boldsymbol{\varepsilon}_{\boldsymbol{i,j}}^{(r)}\lambda_j\right)^{k_{i,j}}}{k_{i,j}!} \frac{\exp\left[-\frac{\left(\log\left(\lambda_j\right)+\frac{\left(\boldsymbol{\gamma_j^2}\right)^{(r)}}{2}\right)^2}{2\left(\boldsymbol{\gamma_j^2}\right)^{(r)}}\right]}{\sqrt{2\pi}\boldsymbol{\gamma_j}^{(r)}\lambda_j} d\lambda_j}{\int_0^\infty \prod_{i=1}^{2} \frac{\exp(-\boldsymbol{\varepsilon}_{\boldsymbol{i,j}}^{(r)}\lambda_j)\left(\boldsymbol{\varepsilon}_{\boldsymbol{i,j}}^{(r)}\lambda_j\right)^{k_{i,j}}}{k_{i,j}!} \frac{\exp\left[-\frac{\left(\log\left(\lambda_j\right)+\frac{\left(\boldsymbol{\gamma_j^2}\right)^{(r)}}{2}\right)^2}{2\left(\boldsymbol{\gamma_j^2}\right)^{(r)}}\right]}{\sqrt{2\pi}\boldsymbol{\gamma_j}^{(r)}\lambda_j} d\lambda_j}, \tag{4.42}$$

  where $\boldsymbol{\varepsilon}_{\boldsymbol{1,j}}^{(r)} = \exp\left(\boldsymbol{x}_{\boldsymbol{1,j}}^T \boldsymbol{\beta}_{\boldsymbol{1}}^{(r)}\right)$, $\boldsymbol{\varepsilon}_{\boldsymbol{2,j}}^{(r)} = \exp\left(\boldsymbol{x}_{\boldsymbol{2,j}}^T \boldsymbol{\beta}_{\boldsymbol{2}}^{(r)}\right)$ and $\boldsymbol{\gamma}_{\boldsymbol{j}}^{(r)} = \exp\left(\boldsymbol{x}_{\boldsymbol{3,j}}^T \boldsymbol{\beta}_{\boldsymbol{3}}^{(r)}\right)$ are the estimates obtained after r-th iteration. Note that the expectations in Eqs. (4.41 and 4.42) do not have closed form expressions and thus have to be evaluated numerically. Alternatively, a Monte Carlo approach is also possible using a rejection algorithm. This approach leads to variants of the EM algorithm such as the Monte Carlo EM (MCEM) algorithm (see, for instance, Booth and

Hobert (1999), Booth et al. (2001), Karlis (2001) and Karlis (2005)) which do not require knowledge of the jpmf $P\left(k_{1,j}, k_{2,j}|\boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \boldsymbol{x_{3,j}}\right)$ but it suffices to be able to simulate from the posterior density $f\left(\lambda_j|k_{i,j}, \boldsymbol{x_{i,j}}, \boldsymbol{x_{3,j}}; \boldsymbol{\theta}\right)$.

- **M-Step**:
  - Update the regression parameters $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$ using the pseudo-values w_j, which are given by Eq. (4.41), and the Newton-Raphson algorithms in Eqs. (4.22, 4.23 and 4.24) and Eqs. (4.25, 4.26 and 4.27) respectively.
  - Update the regression parameters $\boldsymbol{\beta_3}$ using the pseudo-values ω_j, which are given by Eq. (4.42) and the Newton-Raphson algorithm which, in the case of the Lognormal mixing distribution, goes as follows

$$h_3\left(\boldsymbol{\beta_3}\right) = \left[\frac{\omega_j}{\left(\boldsymbol{\gamma_j^2}\right)^{(r)}} - \frac{\left(\boldsymbol{\gamma_j^2}\right)^{(r)}}{4} - 1\right] x_{3,j,l}, \tag{4.43}$$

$$\begin{aligned} H_3\left(\boldsymbol{\beta_3}\right) &= \sum_{j=1}^{n}\left[\frac{-2\omega_j}{\left(\boldsymbol{\gamma_j^2}\right)^{(r)}} - \frac{\left(\boldsymbol{\gamma_j^2}\right)^{(r)}}{2}\right] x_{3,j,l} x_{3,j,l}^T \\ &= \mathbf{X_3^T W_3 X_3}, \end{aligned} \tag{4.44}$$

for $j = 1, .., n$ and $l = 1, ..., p_3$, where $\Psi_3\left(.\right)$ is the trigamma function and where $\mathbf{W}_3 = \text{diag}\left\{\frac{-2\omega_j}{\left(\boldsymbol{\gamma_j^2}\right)^{(r)}} - \frac{\left(\boldsymbol{\gamma_j^2}\right)^{(r)}}{2}\right\}$. Then, we can obtain the updated estimates of $\boldsymbol{\beta_3^{(r)}}$ using Eq. (4.30).

4.4 Calculation of the premia according to the expected value and variance principles

Similarly to what we saw in Subsection 3.5.3, consider the policyholder j, $j = 1, ...n$, with number of bodily injury and property damage claims $k_{1,j,l}$ and $k_{2,j,l}$ respectively, for the year of coverage l, with $l = 1, ..., t$. Also, assume that, for all the years that the individual j has been registered with the insurance company, its cumulative number of claims per type $i = 1, 2$ is given by $K_{i,j} = \sum_{l=1}^{t} k_{i,j,l}$. Then, employing Bayes theorem, we can easily compute the posterior distribution of $\lambda_{j,t+1}$ for the period $t + 1$ given the observations of the reported accidents in the preceding t periods and observable characteristics in the preceding $t + 1$ periods and the current period. In particular, the posterior distribution of $\lambda_{j,t+1}$, analogously as we did in Eq. (3.31), can be derived as follows:

$$
\begin{aligned}
& f\left(\lambda_{j,t+1} | k_{1,j,1}, ..., k_{1,j,t}; k_{2,j,1}, ..., k_{2,j,t}; \boldsymbol{x_{1,j,1}}, ..., \boldsymbol{x_{1,j,t+1}}; \boldsymbol{x_{2,j,1}}, ..., \boldsymbol{x_{2,j,t+1}}; \right. \\
& \qquad \left. \boldsymbol{x_{3,j,1}}, ..., \boldsymbol{x_{3,j,t+1}}\right) \\
& = \frac{\prod_{l=1}^{t} P\left(k_{1,j}, k_{2,j} | \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \boldsymbol{x_{3,j}}, \lambda_j\right) f\left(\lambda_{j,t+1}; \boldsymbol{\gamma_j}\right)}{\int_0^{\infty} P\left(k_{1,j}, k_{2,j} | \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \boldsymbol{x_{3,j}}, \lambda_j\right) f\left(\lambda_{j,t+1}; \boldsymbol{\gamma_j}\right) d\lambda_{j,t+1}},
\end{aligned}
\tag{4.45}
$$

where $P\left(k_{1,j}, k_{2,j} | \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \boldsymbol{x_{3,j}}, \lambda_j\right)$ is the bivariate Poisson distribution and where $f(\lambda_{j,t+1}; \boldsymbol{\gamma_j})$ is the pdf of the mixing distribution.

4.4.1 Expected value principle

The a posteriori, or Bonus-Malus, premia calculated according to the expected value principle are given by

$$P_1 = (1+\omega_1)\mathbb{E}\left(\lambda_{j,t+1}|k_{1,j,1},...,k_{1,j,t};k_{2,j,1},...,k_{2,j,t};\boldsymbol{x_{1,j,1}},...,\boldsymbol{x_{1,j,t+1}};\right.$$
$$\left.\boldsymbol{x_{2,j,1}},...,\boldsymbol{x_{2,j,t+1}};\boldsymbol{x_{3,j,1}},...,\boldsymbol{x_{3,j,t+1}}\right), \tag{4.46}$$

where $\omega_1 > 0$ is a risk load and where the expectation in Eq. (4.46) is that of the posterior distribution given by Eq. (4.45), similarly to the expectation we had in Eq. (3.32).

- In the case of the BNB model, Eq. (4.45) is a Gamma distribution with parameters $\boldsymbol{\gamma_j} + \sum_{i=1}^{2} k_{i,j}$ and $\boldsymbol{\gamma_j} + \sum_{l=1}^{t}\sum_{i=1}^{2}\varepsilon_{i,j,l}$, like we had seen in Eq. (3.33) for the case of the BNB with no covariates in the heterogeneity factor, and hence Eq. (4.46) takes the form
$$P_1 = (1+\omega_1)\frac{\boldsymbol{\gamma_j} + \sum_{i=1}^{2} k_{i,j}}{\boldsymbol{\gamma_j} + \sum_{l=1}^{t}\sum_{i=1}^{2}\varepsilon_{i,j,l}}. \tag{4.47}$$

- In the case of the BPIG model, Eq. (4.45) is a Generalized Inverse Gaussian (GIG) distribution with parameters $\sqrt{\boldsymbol{\gamma_j^2} + 2\sum_{l=1}^{t}\sum_{i=1}^{2}\varepsilon_{i,j,l}}$, $\boldsymbol{\gamma_j}$ and $\sum_{i=1}^{2} k_{i,j} - \frac{1}{2}$, like we had seen in Eq. (3.34) for the case on the BPIG with no covariates in the heterogeneous component, and thus Eq. (4.46) is given by

$$P_1 = \frac{(1+\omega_1)\boldsymbol{\gamma_j} K_{\sum_{i=1}^{2} k_{i,j}+\frac{1}{2}}\left(\boldsymbol{\gamma_j}\sqrt{\boldsymbol{\gamma_j^2} + 2\sum_{l=1}^{t}\sum_{i=1}^{2}\varepsilon_{i,j,l}}\right)}{\sqrt{\boldsymbol{\gamma_j^2} + 2\sum_{l=1}^{t}\sum_{i=1}^{2}\varepsilon_{i,j,l}}\, K_{\sum_{i=1}^{2} k_{i,j}-\frac{1}{2}}\left(\boldsymbol{\gamma_j}\sqrt{\boldsymbol{\gamma_j^2} + 2\sum_{l=1}^{t}\sum_{i=1}^{2}\varepsilon_{i,j,l}}\right)}. \tag{4.48}$$

- In the case of the BPLN model, the posterior expectation in Eq. (4.46) cannot be calculated in closed form but it can be computed via numerical integration which does not require the knowledge of the pdf given by Eq. (4.45). Thus, P_1 can be calculated without any special effort as:

$$P_1 = (1+\omega_1) \int_0^{\infty} \lambda_{j,t+1} \frac{\prod_{l=1}^{t} \prod_{i=1}^{2} \frac{\exp(-(\boldsymbol{\varepsilon_{i,j}}\lambda_j))(\boldsymbol{\varepsilon_{i,j}}\lambda_j)^{k_{i,j}}}{k_{i,j}!} \frac{\exp\left[-\frac{(\log(\lambda_{j,t+1})+\boldsymbol{\gamma_j}^2/2)^2}{2\boldsymbol{\gamma_j}^2}\right]}{\sqrt{2\pi}\boldsymbol{\gamma_j}\lambda_{j,t+1}}}{\int_0^{\infty} \prod_{i=1}^{2} \frac{\exp(-(\boldsymbol{\varepsilon_{i,j}}\lambda_j))(\boldsymbol{\varepsilon_{i,j}}\lambda_j)^{k_{i,j}}}{k_{i,j}!} \frac{\exp\left[-\frac{(\log(\lambda_{j,t+1})+\boldsymbol{\gamma_j}^2/2)^2}{2\boldsymbol{\gamma_j}^2}\right]}{\sqrt{2\pi}\boldsymbol{\gamma_j}\lambda_{j,t+1}} d\lambda_{j,t+1}} d\lambda_{j,t+1}. \tag{4.49}$$

4.4.2 Variance principle

The a posteriori, or Bonus-Malus, premia calculated according to the variance principle are given by

$$\begin{aligned} P_2 = {} & (1+\omega_2)\mathbb{E}\,(\lambda_{j,t+1}|k_{1,j,1},...,k_{1,j,t};k_{2,j,1},...,k_{2,j,t};\boldsymbol{x_{1,j,1}},...,\boldsymbol{x_{1,j,t+1}};\boldsymbol{x_{2,j,1}},...,\boldsymbol{x_{2,j,t+1}}; \\ & \boldsymbol{x_{3,j,1}},...,\boldsymbol{x_{3,j,t+1}}) \\ & + \omega_2 \Big[\mathrm{Var}\,(\lambda_{j,t+1}|k_{1,j,1},...,k_{1,j,t};k_{2,j,1},...,k_{2,j,t};\boldsymbol{x_{1,j,1}},...,\boldsymbol{x_{1,j,t+1}};\boldsymbol{x_{2,j,1}},...,\boldsymbol{x_{2,j,t+1}}; \\ & \boldsymbol{x_{3,j,1}},...,\boldsymbol{x_{3,j,t+1}})\Big]. \end{aligned} \tag{4.50}$$

where $\omega_2 > 0$ is a risk load and where the expectation and the variance in Eq. (4.50) are those of the posterior distribution in Eq.(4.45).

- In the case of the BNB model, using the result in Eq. (4.47), Eq. (4.50) becomes

$$P_2 = (1+\omega_2)\frac{\boldsymbol{\gamma_j} + \sum_{i=1}^{2} k_{i,j}}{\boldsymbol{\gamma_j} + \sum_{l=1}^{t}\sum_{i=1}^{2} \varepsilon_{i,j,l}} + \omega_2 \frac{\boldsymbol{\gamma_j} + \sum_{i=1}^{2} k_{i,j}}{\left(\boldsymbol{\gamma_j} + \sum_{l=1}^{t}\sum_{i=1}^{2} \varepsilon_{i,j,l}\right)^2}. \tag{4.51}$$

- In the case of the BPIG model, using the result in Eq. (4.48), Eq. (4.50) becomes

$$P_2 = \frac{(1+\omega_2)\boldsymbol{\gamma_j} K_{\sum_{i=1}^{2} k_{i,j}+\frac{1}{2}}\left(\boldsymbol{\gamma_j}\sqrt{\boldsymbol{\gamma_j}^2 + 2\sum_{l=1}^{t}\sum_{i=1}^{2}\varepsilon_{i,j,l}}\right)}{\sqrt{\boldsymbol{\gamma_j}^2 + 2\sum_{l=1}^{t}\sum_{i=1}^{2}\varepsilon_{i,j,l}}\, K_{\sum_{i=1}^{2} k_{i,j}-\frac{1}{2}}\left(\boldsymbol{\gamma_j}\sqrt{\boldsymbol{\gamma_j}^2 + 2\sum_{l=1}^{t}\sum_{i=1}^{2}\varepsilon_{i,j,l}}\right)}$$
$$+ \omega_2\left(\frac{\boldsymbol{\gamma_j}^2}{\boldsymbol{\gamma_j}^2 + 2\sum_{l=1}^{t}\sum_{i=1}^{2}\varepsilon_{i,j,l}}\right)\left[\frac{K_{\sum_{i=1}^{2} k_{i,j}+\frac{3}{2}}(\boldsymbol{\gamma_j}\sqrt{\boldsymbol{\gamma_j}^2 + 2\sum_{l=1}^{t}\sum_{i=1}^{2}\varepsilon_{i,j,l}})}{K_{\sum_{i=1}^{2} k_{i,j}-\frac{1}{2}}(\boldsymbol{\gamma_j}\sqrt{\boldsymbol{\gamma_j}^2 + 2\sum_{l=1}^{t}\sum_{i=1}^{2}\varepsilon_{i,j,l}})}\right.$$
$$\left. - \left(\frac{K_{\sum_{i=1}^{2} k_{i,j}+\frac{1}{2}}(\boldsymbol{\gamma_j}\sqrt{\boldsymbol{\gamma_j}^2 + 2\sum_{l=1}^{t}\sum_{i=1}^{2}\varepsilon_{i,j,l}})}{K_{\sum_{i=1}^{2} k_{i,j}-\frac{1}{2}}(\boldsymbol{\gamma_j}\sqrt{\boldsymbol{\gamma_j}^2 + 2\sum_{l=1}^{t}\sum_{i=1}^{2}\varepsilon_{i,j,l}})}\right)^2\right]. \tag{4.52}$$

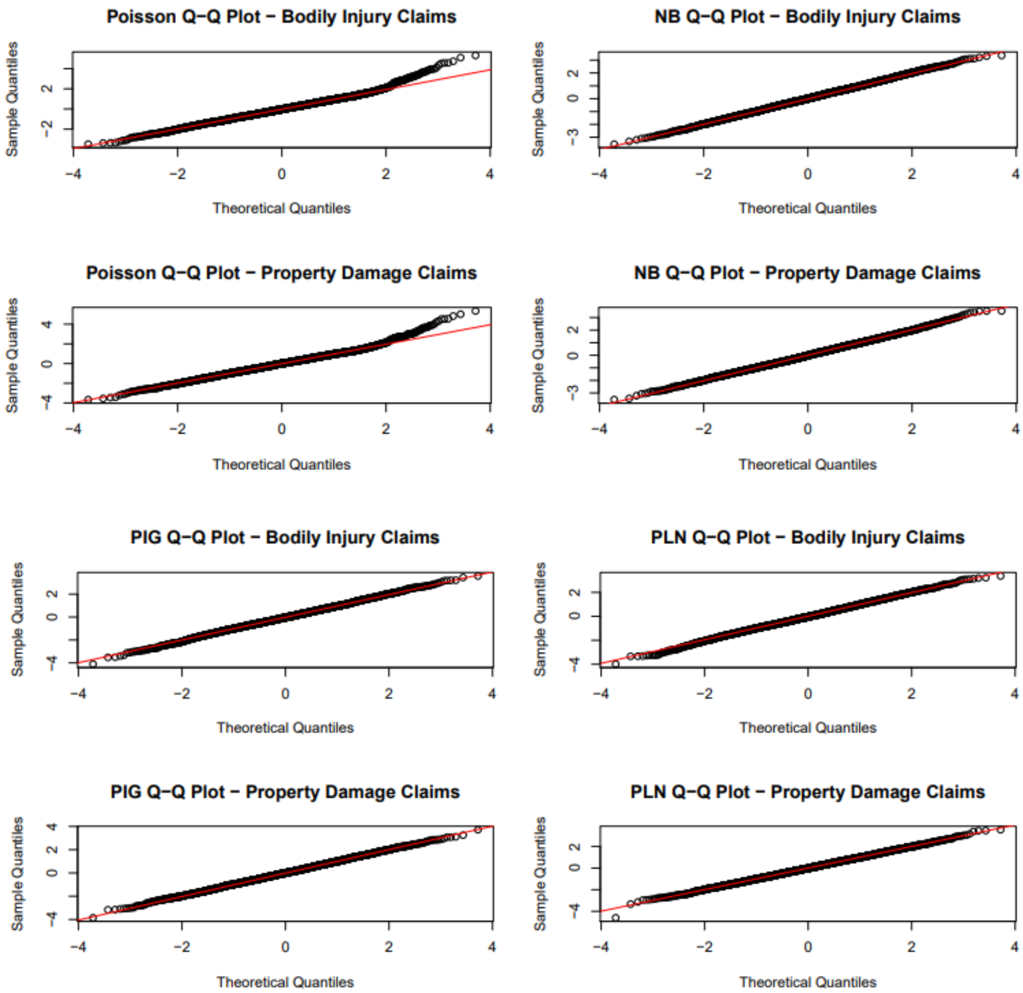
- In the case of the BPLN model, the posterior mean and the posterior variance in Eq. (4.50) cannot be calculated in closed form. However, both can be calculated based on numerical integration which does not rely on the knowledge of the pdf given by Eq. (4.45) and hence P_2 can be easily computed as

$$
\begin{aligned}
P_2 =&(1+\omega_2)\int_0^\infty \lambda_{j,t+1}\frac{\prod\limits_{l=1}^{t}\prod\limits_{i=1}^{2}\frac{\exp(-(\boldsymbol{\varepsilon_{i,j}}\lambda_j))(\boldsymbol{\varepsilon_{i,j}}\lambda_j)^{k_{i,j}}}{k_{i,j}!}\frac{\exp\left[-\frac{(\log(\lambda_{j,t+1})+\boldsymbol{\gamma_j^2}/2)^2}{2\boldsymbol{\gamma_j^2}}\right]}{\sqrt{2\pi}\boldsymbol{\gamma_j}\lambda_{j,t+1}}}{\int\limits_0^\infty\prod\limits_{i=1}^{2}\frac{\exp(-(\boldsymbol{\varepsilon_{i,j}}\lambda_j))(\boldsymbol{\varepsilon_{i,j}}\lambda_j)^{k_{i,j}}}{k_{i,j}!}\frac{\exp\left[-\frac{(\log(\lambda_{j,t+1})+\boldsymbol{\gamma_j^2}/2)^2}{2\boldsymbol{\gamma_j^2}}\right]}{\sqrt{2\pi}\boldsymbol{\gamma_j}\lambda_{j,t+1}}d\lambda_{j,t+1}}d\lambda_{j,t+1}\\
&+\omega_2\Bigg[\int_0^\infty \lambda_{j,t+1}^2\frac{\prod\limits_{l=1}^{t}\prod\limits_{i=1}^{2}\frac{\exp(-(\boldsymbol{\varepsilon_{i,j}}\lambda_j))(\boldsymbol{\varepsilon_{i,j}}\lambda_j)^{k_{i,j}}}{k_{i,j}!}\frac{\exp\left[-\frac{(\log(\lambda_{j,t+1})+\boldsymbol{\gamma_j^2}/2)^2}{2\boldsymbol{\gamma_j^2}}\right]}{\sqrt{2\pi}\boldsymbol{\gamma_j}\lambda_{j,t+1}}}{\int\limits_0^\infty\prod\limits_{i=1}^{2}\frac{\exp(-(\boldsymbol{\varepsilon_{i,j}}\lambda_j))(\boldsymbol{\varepsilon_{i,j}}\lambda_j)^{k_{i,j}}}{k_{i,j}!}\frac{\exp\left[-\frac{(\log(\lambda_{j,t+1})+\boldsymbol{\gamma_j^2}/2)^2}{2\boldsymbol{\gamma_j^2}}\right]}{\sqrt{2\pi}\boldsymbol{\gamma_j}\lambda_{j,t+1}}d\lambda_{j,t+1}}d\lambda_{j,t+1}\\
&-\left(\int_0^\infty \lambda_{j,t+1}\frac{\prod\limits_{l=1}^{t}\prod\limits_{i=1}^{2}\frac{(\exp(-(\boldsymbol{\varepsilon_{i,j}}\lambda_j))\boldsymbol{\varepsilon_{i,j}}\lambda_j)^{k_{i,j}}}{k_{i,j}!}\frac{\exp\left[-\frac{(\log(\lambda_{j,t+1})+\boldsymbol{\gamma_j^2}/2)^2}{2\boldsymbol{\gamma_j^2}}\right]}{\sqrt{2\pi}\boldsymbol{\gamma_j}\lambda_{j,t+1}}}{\int\limits_0^\infty\prod\limits_{i=1}^{2}\frac{\exp(-(\boldsymbol{\varepsilon_{i,j}}\lambda_j))(\boldsymbol{\varepsilon_{i,j}}\lambda_j)^{k_{i,j}}}{k_{i,j}!}\frac{\exp\left[-\frac{(\log(\lambda_{j,t+1})+\boldsymbol{\gamma_j^2}/2)^2}{2\boldsymbol{\gamma_j^2}}\right]}{\sqrt{2\pi}\boldsymbol{\gamma_j}\lambda_{j,t+1}}d\lambda_{j,t+1}}d\lambda_{j,t+1}\right)^2\Bigg].
\end{aligned}
\tag{4.53}
$$

4.5 Numerical illustration

The study is based on the real data set described in Section 3.5 from an MTPL insurance portfolio observed during the year 2017 from a major European insurance company. Now we fit the univariate Negative Binomial (NB), Poisson-Inverse Gaussian (PIG) and Poisson-Lognormal (PLN) regression models with varying dispersion for claim frequencies using the univariate version of the EM algorithm which was presented in Section 4.2. Additionally, the simple Poisson regression model was fitted for comparison purposes. As we did in Section 3.5, the normalized randomized quantile residuals, see Dunn and Smyth (1996), are used as a graphical tool to help us assess the adequacy of the fit of the competing models for both bodily injury and property damage claim counts $k_{1,j}$ and $k_{2,j}$. The normalized (random) quantiles for the Poisson, NB, PIG and PLN models are presented in Figure 4.1 per claim type $i = 1, 2$. From Figure 4.1, we observe that the NB, PIG and PLN are better assumptions than the Poisson model which does not capture the tails of the claim frequency distributions of $k_{1,j}$ and $k_{2,j}$, observation which had already been made clear from Figure 3.1. In particular, the residuals of the three mixed Poisson models are close to the diagonal and indicate a good fit to the distributions of both $k_{1,j}$ and $k_{2,j}$, whereas the sample quantiles of the Poisson model, due to the equidispersion constraint, near the tail end of the distributions of both $k_{1,j}$ and $k_{2,j}$, are significantly higher than the theoretical quantiles.

Figure 4.1: Normalized quantiles for the Poisson, NB, PIG and PLN regression models with varying dispersion



4.5.1 Modelling results

This subsection describes the modelling results of the BNB, BPIG and BPLN distributions and regression models with varying dispersion. The ML estimates of their parameters and the corresponding standard errors in parentheses are presented in Table 4.1 for the distributions[4] and in Table 4.2 for the regression models with varying dispersion respectively. Note that in the latter case, for illustrative purposes we considered that the two location parameters $\boldsymbol{\varepsilon_{1,j}}$, $\boldsymbol{\varepsilon_{2,j}}$ and the dispersion parameter $\boldsymbol{\gamma_j}$, $j = 1, ..., n$, of the aforementioned models are modelled using all three available explanatory variables. However, it should be noted that for larger data sets variable selection can start with the examination of the two mean parameters of the bivariate mixed Poisson regression model with varying dispersion.

This can be achieved by adding all available covariates and testing whether the exclusion of each one lowers the Global Deviance (DEV), Akaike Information Criterion (AIC) and the Schwartz Bayesian Criterion (SBC) values. Then, after having selected the best predictors for the two mean parameters, we can continue in determining the remaining predictors by testing which rating variable between those used in the two mean parameters would lead to a further decrease of the DEV, AIC and SBC values when inserted in the dispersion parameter of the bivariate claim frequency model with varying dispersion.

Additionally, if between the same bivariate mixed Poisson distribution with different parameter specifications, several models have similar DEV, AIC and SBC values, the simpler model can be used in order to avoid overfitting. Therefore, in such cases, it should be expected that the dispersion parameters of the bivariate mixed Poisson regression model with varying dispersion may have fewer predictors than the two mean parameters.

[4] Note that the mean parameters of the BNB, BPIG and BPLN distributions are denoted by μ_1 and μ_2 and the dispersion parameter is denoted by γ.

Table 4.1: Parameters estimates and in parenthesis the associated standard errors of the fitted BNB, BPIG and BPLN distributions

BNB		BPIG		BPLN	
μ_1	μ_2	μ_1	μ_2	μ_1	μ_2
0.0954	0.0618	0.0954	0.0618	0.0954	0.0618
(0.0542)	(0.0639)	(0.0535)	(0.0633)	(0.0546)	(0.0643)
γ		γ		γ	
0.2612		0.4866		1.4072	
(0.1037)		(0.0554)		(0.0448)	

Table 4.2: Parameters estimates and in parenthesis the associated standard errors of the fitted BNB, BPIG and BPLN regression models with varying dispersion for each covariate

	BNB			BPIG			BPLN		
Variable	Coeff. $\boldsymbol{\beta_1}$	Coeff. $\boldsymbol{\beta_2}$	Coeff. $\boldsymbol{\beta_3}$	Coeff. $\boldsymbol{\beta_1}$	Coeff. $\boldsymbol{\beta_2}$	Coeff. $\boldsymbol{\beta_3}$	Coeff. $\boldsymbol{\beta_1}$	Coeff. $\boldsymbol{\beta_2}$	Coeff. $\boldsymbol{\beta_3}$
Intercept	-2.3933 (0.0981)	-2.9262 (0.1121)	-1.1296 (0.2126)	-2.3950 (0.0997)	-2.9279 (0.1010)	-0.5908 (0.1418)	-2.3839 (0.0898)	-2.9167 (0.1374)	0.2380 (0.0474)
v1 C2	0.0524 (0.0238)	0.1504 (0.0711)	-0.1912 (0.0940)	0.0535 (0.0232)	0.1518 (0.0698)	-0.1157 (0.0564)	0.0708 (0.0329)	0.1691 (0.0808)	0.0905 (0.0436)
v1 C3	0.1556 (0.0798)	0.1770 (0.0938)	-0.2303 (0.1268)	0.1587 (0.0804)	0.1793 (0.0939)	-0.1364 (0.0732)	0.1941 (0.1003)	0.2143 (0.1149)	0.1223 (0.0685)
v2 C2	0.0452 (0.0169)	0.1780 (0.0719)	0.3627 (0.1611)	0.0465 (0.0170)	0.1790 (0.0704)	0.1959 (0.0859)	0.0348 (0.0137)	0.1669 (0.0704)	-0.1198 (0.0571)
v3 C2	-0.1216 (0.0542)	-0.0203 (0.0093)	-0.3144 (0.1473)	-0.1203 (0.0525)	-0.0190 (0.0085)	-0.1769 (0.0819)	-0.1000 (0.0459)	-0.021 (0.0098)	0.1230 (0.0608)
v3 C3	0.1767 (0.0793)	0.1712 (0.0786)	-0.0883 (0.0426)	0.1784 (0.0787)	0.1731 (0.0776)	-0.0716 (0.0341)	0.1934 (0.0893)	0.1882 (0.0895)	0.0674 (0.0341)

As we can see from Table 4.2, the values of the estimated regression coefficients of the variables v1, v2, and v3 are almost identical for $\boldsymbol{\varepsilon_{1,j}}$ and $\boldsymbol{\varepsilon_{2,j}}$ across all three bivariate mixed Poisson distributions, whereas they differ hugely for the dispersion parameter $\boldsymbol{\gamma_j}$. Additionally, we observe that the same explanatory variables always have the same effect (positive and/or negative) on the parameter $\boldsymbol{\gamma_j}$ in the case of the BNB and BPIG models but have a different effect for $\boldsymbol{\gamma_j}$ in the case of the BPLN model.

4.5.2 Model comparison

In this subsection we compare the fit of the BNB, BPIG and BPLN distributions/regression models with varying dispersion based on the classic hypothesis/specification tests DEV, AIC and SBC. These specification criteria were already presented in Subsection 3.5.2, in Eqs. (3.26, 3.27 and 3.28). The values of the DEV, AIC and SBC for the competing bivariate mixed Poisson distributions/regression models with varying dispersion are given in Table 4.3.

As mentioned in Subsection 3.5.2, according to a very well known rule of thumb, two models can be considered to be significantly different if the difference in the log-likelihoods exceeds five, corresponding to a difference in their respective AIC and SBC values greater than ten and five respectively, see Anderson and Burnham (2004) and Raftery (1995) respectively. Therefore, in this case we see that the best fitting performances are provided by the BPIG distribution/regression model with varying dispersion.

Table 4.3: Models comparison based on the DEV, AIC and SBC

Distributions				Regression Models				
Model	*df*	AIC	SBC	Model	*df*	DEV	AIC	SBC
BNB	3	5615	5635	BNB	18	4388	4424	4542
BPIG	3	5541	5561	BPIG	18	4249	4285	4403
BPLN	3	5684	5704	BPLN	18	4513	4549	4667

It is also to be noted that, compared to Table 3.6, the BNB and BPIG models with varying dispersion fit significantly better than the corresponding BNB and BPIG with fixed dispersion, proving to us once more that considering the use of covariates also on the dispersion parameter is of crucial importance.

At this point we perform a 10-fold cross-validation to check the robustness of the proposed models. We calculated the DEV, AIC and SBC values for the BNB, BPIG and BPLN regression models on each of the 10 subsets and in Table 4.4 we report an average of the 10 values that we got from each subset. Furthermore, we calculated the root-mean-square error (RMSE) on each subset using Eq. (3.30) and in Table 4.4 we report a mean of the 10 RMSEs obtained on each subset.

Table 4.4: DEV, AIC, SBC and RMSE with the 10-fold cross-validation

Regression Models					
Model	*df*	DEV	AIC	SBC	RMSE
BNB	18	450	486	563	0.3941
BPIG	18	431	467	544	0.3699
BPLN	18	461	497	574	0.3730

4.5.3 Application to ratemaking

In this subsection, following the current methodology, as presented in Section 4.4, we calculate the a posteriori, or Bonus-Malus, premia resulting from the three BNB, BPIG and BPLN distributions/regression models with varying dispersion using the expected value and the variance principles. The premium rates will be divided by the premium when $t = 0$, i.e. we calculate the relative premia, since we are interested in the differences between various classes and the results are presented so that the premium for a new policyholder is 100.

Thus, in what follows, when the expected value principle is used, note the disappearance of the factor $(1 + \omega_1)$ from Eqs. (4.47, 4.48 and 4.49). Also, when the variance principle is used, following and extending to the bivariate case the framework of Lemaire (1995) and Tzougas et al. (2018), we consider that $\omega_2 = 0.235$ in Eqs. (4.51, 4.52 and 4.53), which corresponds to a safety loading of 25% of the net premium.

Firstly, assuming that the number of individual bodily injury and property damage claims, $k_{1,j}$ and $k_{2,j}$ respectively, with $j = 1, ..., n$, range from 0 to 3 and the age of the policy is $t = 1$, $t = 2$ and $t = 3$ years, we computed comparable relative premia for the three bivariate mixed Poisson distributions. Tables 4.5 and 4.6, present the premia rates calculated according to the expected value and variance principles respectively.

Table 4.5: Comparison of the a posteriori, or Bonus-Malus, premium rates for $t = 1, 2, 3$, bivariate claim frequency distributions under the expected value principle

	$t = 1$ BNB distribution				$t = 1$ BPIG distribution				$t = 1$ BPLN distribution			
$k_{1,j}/k_{2,j}$	0	1	2	3	0	1	2	3	0	1	2	3
0	72.69	350.98	629.27	907.56	79.80	256.06	553.66	892.83	81.47	244.10	510.00	847.30
1	350.98	629.27	907.56	1185.85	256.06	553.66	892.83	1240.99	244.10	510.00	847.30	1227.45
2	629.27	907.56	1185.85	1464.14	553.66	892.83	1240.99	1591.52	510.00	847.30	1227.45	1633.51
3	907.56	1185.85	1464.14	1742.43	892.83	1240.99	1591.52	1942.92	847.30	1227.45	1633.51	2056.12
	$t = 2$ BNB distribution				$t = 2$ BPIG distribution				$t = 2$ BPLN distribution			
$k_{1,j}/k_{2,j}$	0	1	2	3	0	1	2	3	0	1	2	3
0	57.10	275.69	494.28	712.87	68.35	197.65	411.53	657.85	70.44	193.99	382.91	614.90
1	275.69	494.28	712.87	931.46	197.65	411.53	657.85	912.20	193.99	382.91	614.90	872.58
2	494.28	712.87	931.46	1150.05	411.53	657.85	912.20	1168.82	382.91	614.90	872.58	1145.94
3	712.87	931.46	1150.05	1368.64	657.85	912.20	1168.82	1426.30	614.90	872.58	1145.94	1429.44
	$t = 3$ BNB distribution				$t = 3$ BPIG distribution				$t = 3$ BPLN distribution			
$k_{1,j}/k_{2,j}$	0	1	2	3	0	1	2	3	0	1	2	3
0	47.01	226.99	406.97	586.96	60.73	162.83	328.94	521.70	62.81	163.41	310.83	488.29
1	226.99	406.97	586.96	766.94	162.83	328.94	521.70	721.75	163.41	310.83	488.29	683.62
2	406.97	586.96	766.94	946.92	328.94	521.70	721.75	923.98	310.83	488.29	683.62	889.94
3	586.96	766.94	946.92	1126.90	521.70	721.75	923.98	1127.06	488.29	683.62	889.94	1103.44

Table 4.6: Comparison of the a posteriori, or Bonus-Malus, premium rates for $t = 1, 2, 3$, bivariate claim frequency distributions under the variance principle

	$t = 1$ BNB distribution				$t = 1$ BPIG distribution				$t = 1$ BPLN distribution			
$k_{1,j}/k_{2,j}$	0	1	2	3	0	1	2	3	0	1	2	3
0	66.48	321.01	575.54	830.07	72.36	261.00	585.67	951.89	74.79	248.08	553.01	953.33
1	321.01	575.54	830.07	1084.61	261.00	585.67	951.89	1325.81	248.08	553.01	953.33	1411.35
2	575.54	830.07	1084.61	1339.14	585.67	951.89	1325.81	1701.58	553.01	953.33	1411.35	1903.99
3	830.07	1084.61	1339.14	1593.67	951.89	1325.81	1701.58	2078.02	953.33	1411.35	1903.99	2418.52
	$t = 2$ BNB distribution				$t = 2$ BPIG distribution				$t = 2$ BPLN distribution			
$k_{1,j}/k_{2,j}$	0	1	2	3	0	1	2	3	0	1	2	3
0	49.44	238.71	427.98	617.26	59.00	186.12	399.90	644.15	62.04	182.94	376.80	620.13
1	238.71	427.98	617.26	806.53	186.12	399.90	644.15	895.13	182.94	376.80	620.13	893.03
2	427.98	617.26	806.53	995.80	399.90	644.15	895.13	1147.87	376.80	620.13	893.03	1183.86
3	617.26	806.53	995.80	1185.08	644.15	895.13	1147.87	1401.27	620.13	893.03	1183.86	1486.19
	$t = 3$ BNB distribution				$t = 3$ BPIG distribution				$t = 3$ BPLN distribution			
$k_{1,j}/k_{2,j}$	0	1	2	3	0	1	2	3	0	1	2	3
0	39.22	189.39	339.56	489.73	50.89	146.12	303.31	484.57	53.94	147.64	289.72	463.44
1	189.39	339.56	489.73	639.89	146.12	303.31	484.57	671.85	147.64	289.72	463.44	655.99
2	339.56	489.73	639.89	790.06	303.31	484.57	671.85	860.82	289.72	463.44	655.99	860.06
3	489.73	639.89	790.06	940.23	484.57	671.85	860.82	1050.43	463.44	655.99	860.06	1071.59

Secondly, when both the a posteriori and the a priori criteria, i.e. the characteristics of the policyholders and their cars, are considered, we analyse three risk class profiles that we classify (as we did in Chapter 3) as Best, Average and Worst according to the values of the mean claim frequencies $\boldsymbol{\varepsilon_{1,j}}$ and $\boldsymbol{\varepsilon_{2,j}}$, which are calculated using the same set of explanatory variables per claim type $i = 1, 2$ in the case of the BNB, BPIG and BPLN models respectively. More specifically, for our data, we characterize the Best, Average and Worst profiles as such based on category C1 for all three explanatory variables v1, v2 and v3 in the case of the first, category C2 for v1, v2 and v3 in the case of the second, and category C3 for v1 and v3 and C2 for v2 in the case of the third. Also, the dispersion parameter $\boldsymbol{\gamma_j}$ of the BNB, BPIG and BPLN models is computed for each risk class profile. The results are shown in Table 4.7.

Table 4.7: Results of the fitted BNB, BPIG and BPLN regression models with varying dispersion for each risk class profile

Regression model	Profile	$\varepsilon_{1,j}$	$\varepsilon_{2,j}$	γ_j
BNB	Best	0.091327	0.053599	0.323154
	Average	0.089161	0.072942	0.280116
	Worst	0.133208	0.090727	0.337726
BPIG	Best	0.091173	0.053509	0.553884
	Average	0.089341	0.073087	0.502832
	Worst	0.133801	0.091036	0.547222
BPLN	Best	0.092194	0.054112	1.268773
	Average	0.092705	0.074147	1.393263
	Worst	0.140632	0.095639	1.360708

From Table 4.7 we observe that for all three risk profiles small differences lie in the mean values $\boldsymbol{\varepsilon_{1,j}}$ and $\boldsymbol{\varepsilon_{2,j}}$ of the BNB, BPG and BPLN models. On the contrary, as previously mentioned, more significant differences are noticed across the three risk profiles in the values of the dispersion parameters $\boldsymbol{\gamma_j}$ of the bivariate mixed Poisson models. Due to these discrepancies, the a posteriori, or Bonus-Malus, premium rates that will result from the three models by updating their posterior mean and the posterior variance will be better distinguished under different distributional assumptions. Thus, as explained before, the proposed modelling framework leads to a better tariffication than the assumption of a constant dispersion γ, which is what we had in Chapter 3. In what follows, Tables 4.8, 4.9 and 4.10 depict the premia computed under the expected value principle for the three risk profiles during the years $t = 1$, $t = 2$ and $t = 3$ respectively. Furthermore, Tables 4.11, 4.12 and 4.13 present the premia calculated via the variance principle for the same risk profiles and years of insurance policy.

Table 4.8: Comparison of the a posteriori, or Bonus-Malus, premium rates for $t = 1$ under the expected value principle, bivariate claim frequency regression models with varying dispersion

	$t=1$ BNB regression model Best profile				$t=1$ BPIG regression model Best profile				$t=1$ BPLN regression model Best profile			
$k_{1,j}/k_{2,j}$	0	1	2	3	0	1	2	3	0	1	2	3
0	76.36	312.65	548.94	785.23	82.05	239.47	500.38	800.56	83.78	223.13	451.24	749.12
1	312.65	548.94	785.23	1021.52	239.47	500.38	800.56	1110.36	223.13	451.24	749.12	1093.54
2	548.94	785.23	1021.52	1257.81	500.38	800.56	1110.36	1422.85	451.24	749.12	1093.54	1467.98
3	785.23	1021.52	1257.81	1494.10	800.56	1110.36	1422.85	1736.37	749.12	1093.54	1467.98	1862.30

	$t=1$ BNB regression model Average profile				$t=1$ BPIG regression model Average profile				$t=1$ BPLN regression model Average profile			
$k_{1,j}/k_{2,j}$	0	1	2	3	0	1	2	3	0	1	2	3
0	73.18	334.41	595.64	856.88	80.00	247.48	528.29	849.50	81.26	237.30	489.54	808.55
1	334.41	595.64	856.88	1118.11	247.48	528.29	849.50	1179.87	237.30	489.54	808.55	1167.96
2	595.64	856.88	1118.11	1379.35	528.29	849.50	1179.87	1512.72	489.54	808.55	1167.96	1551.97
3	856.88	1118.11	1379.35	1640.58	849.50	1179.87	1512.72	1846.48	808.55	1167.96	1551.97	1951.77

	$t=1$ BNB regression model Worst profile				$t=1$ BPIG regression model Worst profile				$t=1$ BPLN regression model Worst profile			
$k_{1,j}/k_{2,j}$	0	1	2	3	0	1	2	3	0	1	2	3
0	71.49	283.19	494.88	706.58	79.05	210.99	425.43	674.37	79.41	209.86	408.55	653.84
1	283.19	494.88	706.58	918.27	210.99	425.43	674.37	932.83	209.86	408.55	653.84	927.77
2	494.88	706.58	918.27	1129.97	425.43	674.37	932.83	1194.13	408.55	653.84	927.77	1219.52
3	706.58	918.27	1129.97	1341.66	674.37	932.83	1194.13	1456.54	653.84	927.77	1219.52	1522.90

Table 4.9: Comparison of the a posteriori, or Bonus-Malus, premium rates for $t = 2$ under the expected value principle, bivariate claim frequency regression models with varying dispersion

	$t=2$ BNB regression model Best profile				$t=2$ BPIG regression model Best profile				$t=2$ BPLN regression model Best profile			
$k_{1,j}/k_{2,j}$	0	1	2	3	0	1	2	3	0	1	2	3
0	61.76	252.87	443.97	635.08	71.23	189.88	382.67	606.51	73.58	181.78	347.13	554.85
1	252.87	443.97	635.08	826.19	189.88	382.67	606.51	838.91	181.78	347.13	554.85	790.34
2	443.97	635.08	826.19	1017.30	382.67	606.51	838.91	1073.89	347.13	554.85	790.34	1043.81
3	635.08	826.19	1017.30	1208.41	606.51	838.91	1073.89	1309.87	554.85	790.34	1043.81	1309.31

	$t=2$ BNB regression model Average profile				$t=2$ BPIG regression model Average profile				$t=2$ BPLN regression model Average profile			
$k_{1,j}/k_{2,j}$	0	1	2	3	0	1	2	3	0	1	2	3
0	57.70	263.68	469.66	675.64	68.60	191.76	394.00	627.70	70.14	188.56	367.80	587.30
1	263.68	469.66	675.64	881.63	191.76	394.00	627.70	869.56	188.56	367.80	587.30	831.00
2	469.66	675.64	881.63	1087.61	394.00	627.70	869.56	1113.77	367.80	587.30	831.00	1089.59
3	675.64	881.63	1087.61	1293.59	627.70	869.56	1113.77	1358.89	587.30	831.00	1089.59	1357.84

	$t=2$ BNB regression model Worst profile				$t=2$ BPIG regression model Worst profile				$t=2$ BPLN regression model Worst profile			
$k_{1,j}/k_{2,j}$	0	1	2	3	0	1	2	3	0	1	2	3
0	55.64	220.37	385.11	549.85	67.41	163.36	315.66	494.13	67.61	165.48	306.02	474.57
1	220.37	385.11	549.85	714.58	163.36	315.66	494.13	680.83	165.48	306.02	474.57	660.27
2	385.11	549.85	714.58	879.32	315.66	494.13	680.83	870.20	306.02	474.57	660.27	856.74
3	549.85	714.58	879.32	1044.06	494.13	680.83	870.20	1060.64	474.57	660.27	856.74	1060.34

Table 4.10: Comparison of the a posteriori, or Bonus-Malus, premium rates for $t = 3$ under the expected value principle, bivariate claim frequency regression models with varying dispersion

	$t=3$ BNB regression model Best profile				$t=3$ BPIG regression model Best profile				$t=3$ BPLN regression model Best profile			
$k_{1,j}/k_{2,j}$	0	1	2	3	0	1	2	3	0	1	2	3
0	51.84	212.28	372.71	533.14	63.81	159.01	311.21	489.09	66.31	155.60	286.21	446.43
1	212.28	372.71	533.14	693.57	159.01	311.21	489.09	674.74	155.60	286.21	446.43	625.90
2	372.71	533.14	693.57	854.00	311.21	489.09	674.74	862.85	286.21	446.43	625.90	817.88
3	533.14	693.57	854.00	1014.44	489.09	674.74	862.85	1051.94	446.43	625.90	817.88	1018.29
	$t=3$ BNB regression model Average profile				$t=3$ BPIG regression model Average profile				$t=3$ BPLN regression model Average profile			
$k_{1,j}/k_{2,j}$	0	1	2	3	0	1	2	3	0	1	2	3
0	47.63	217.65	387.67	557.69	61.01	158.38	315.63	498.68	62.47	158.82	298.70	466.63
1	217.65	387.67	557.69	727.71	158.38	315.63	498.68	689.11	158.82	298.70	466.63	651.41
2	387.67	557.69	727.71	897.73	315.63	498.68	689.11	881.80	298.70	466.63	651.41	846.62
3	557.69	727.71	897.73	1067.75	498.68	689.11	881.80	1075.38	466.63	651.41	846.62	1048.66
	$t=3$ BNB regression model Worst profile				$t=3$ BPIG regression model Worst profile				$t=3$ BPLN regression model Worst profile			
$k_{1,j}/k_{2,j}$	0	1	2	3	0	1	2	3	0	1	2	3
0	45.54	180.36	315.19	450.02	59.75	135.14	252.57	391.05	59.64	138.67	248.03	376.88
1	180.36	315.19	450.02	584.85	135.14	252.57	391.05	536.82	138.67	248.03	376.88	517.66
2	315.19	450.02	584.85	719.68	252.57	391.05	536.82	685.10	248.03	376.88	517.66	665.98
3	450.02	584.85	719.68	854.51	391.05	536.82	685.10	834.43	376.88	517.66	665.98	819.34

Table 4.11: Comparison of the a posteriori, or Bonus-Malus, premium rates for $t = 1$ under the variance principle, bivariate claim frequency regression models with varying dispersion

	$t = 1$				$t = 1$				$t = 1$			
	BNB regression model				BPIG regression model				BPLN regression model			
	Best profile				Best profile				Best profile			
$k_{1,j}/k_{2,j}$	0	1	2	3	0	1	2	3	0	1	2	3
0	71.14	291.28	511.43	731.57	75.56	246.18	534.73	863.49	77.96	228.96	496.95	862.56
1	291.28	511.43	731.57	951.72	246.18	534.73	863.49	1200.74	228.96	496.95	862.56	1294.49
2	511.43	731.57	951.72	1171.86	534.73	863.49	1200.74	1540.17	496.95	862.56	1294.49	1769.17
3	731.57	951.72	1171.86	1392.01	863.49	1200.74	1540.17	1880.38	862.56	1294.49	1769.17	2271.92
	$t = 1$				$t = 1$				$t = 1$			
	BNB regression model				BPIG regression model				BPLN regression model			
	Average profile				Average profile				Average profile			
$k_{1,j}/k_{2,j}$	0	1	2	3	0	1	2	3	0	1	2	3
0	67.27	307.42	547.58	787.73	72.87	251.97	558.09	904.78	74.80	240.29	526.97	901.62
1	307.42	547.58	787.73	1027.88	251.97	558.09	904.78	1259.41	240.29	526.97	901.62	1329.89
2	547.58	787.73	1027.88	1268.04	558.09	904.78	1259.41	1616.01	526.97	901.62	1329.89	1790.58
3	787.73	1027.88	1268.04	1508.19	904.78	1259.41	1616.01	1973.31	901.62	1329.89	1790.58	2271.89
	$t = 1$				$t = 1$				$t = 1$			
	BNB regression model				BPIG regression model				BPLN regression model			
	Worst profile				Worst profile				Worst profile			
$k_{1,j}/k_{2,j}$	0	1	2	3	0	1	2	3	0	1	2	3
0	66.34	262.76	459.19	655.61	73.04	211.66	440.86	704.99	73.95	208.77	424.18	696.33
1	262.76	459.19	655.61	852.04	211.66	440.86	704.99	977.73	208.77	424.18	696.33	1003.48
2	459.19	655.61	852.04	1048.46	440.86	704.99	977.73	1252.87	424.18	696.33	1003.48	1332.26
3	655.61	852.04	1048.46	1244.89	704.99	977.73	1252.87	1528.92	696.33	1003.48	1332.26	1675.09

Table 4.12: Comparison of the a posteriori, or Bonus-Malus, premium rates for $t = 2$ under the variance principle, bivariate claim frequency regression models with varying dispersion

	$t=2$ BNB regression model Best profile				$t=2$ BPIG regression model Best profile				$t=2$ BPLN regression model Best profile			
$k_{1,j}/k_{2,j}$	0	1	2	3	0	1	2	3	0	1	2	3
0	54.93	224.92	394.90	564.89	62.74	181.82	378.71	605.61	66.00	174.25	348.58	573.77
1	224.92	394.90	564.89	734.87	181.82	378.71	605.61	839.91	174.25	348.58	573.77	832.60
2	394.90	564.89	734.87	904.86	378.71	605.61	839.91	1076.28	348.58	573.77	832.60	1113.15
3	564.89	734.87	904.86	1074.84	605.61	839.91	1076.28	1313.42	573.77	832.60	1113.15	1408.11
	$t=2$ BNB regression model Average profile				$t=2$ BPIG regression model Average profile				$t=2$ BPLN regression model Average profile			
$k_{1,j}/k_{2,j}$	0	1	2	3	0	1	2	3	0	1	2	3
0	50.36	230.12	409.89	589.66	59.61	180.97	383.57	615.93	62.05	177.77	360.91	589.90
1	230.12	409.89	589.66	769.43	180.97	383.57	615.93	855.20	177.77	360.91	589.90	846.51
2	409.89	589.66	769.43	949.20	383.57	615.93	855.20	1096.33	360.91	589.90	846.51	1120.00
3	589.66	769.43	949.20	1128.97	615.93	855.20	1096.33	1338.15	589.90	846.51	1120.00	1404.37
	$t=2$ BNB regression model Worst profile				$t=2$ BPIG regression model Worst profile				$t=2$ BPLN regression model Worst profile			
$k_{1,j}/k_{2,j}$	0	1	2	3	0	1	2	3	0	1	2	3
0	49.39	195.63	341.87	488.12	59.96	154.14	305.77	482.55	60.91	155.65	295.83	466.36
1	195.63	341.87	488.12	634.36	154.14	305.77	482.55	666.61	155.65	295.83	466.36	655.49
2	341.87	488.12	634.36	780.60	305.77	482.55	666.61	852.92	295.83	466.36	655.49	856.22
3	488.12	634.36	780.60	926.84	482.55	666.61	852.92	1040.11	466.36	655.49	856.22	1064.60

Table 4.13: Comparison of the a posteriori, or Bonus-Malus, premium rates for $t = 3$ under the variance principle, bivariate claim frequency regression models with varying dispersion

	$t=3$				$t=3$				$t=3$			
	BNB regression model				BPIG regression model				BPLN regression model			
	Best profile				Best profile				Best profile			
$k_{1,j}/k_{2,j}$	0	1	2	3	0	1	2	3	0	1	2	3
0	44.63	182.73	320.83	458.94	54.65	145.58	293.19	464.71	58.13	143.33	272.77	434.72
1	182.73	320.83	458.94	597.04	145.58	293.19	464.71	642.82	143.33	272.77	434.72	617.91
2	320.83	458.94	597.04	735.14	293.19	464.71	642.82	822.90	272.77	434.72	617.91	814.87
3	458.94	597.04	735.14	873.25	464.71	642.82	822.90	1003.75	434.72	617.91	814.87	1021.05
	$t=3$				$t=3$				$t=3$			
	BNB regression model				BPIG regression model				BPLN regression model			
	Average profile				Average profile				Average profile			
$k_{1,j}/k_{2,j}$	0	1	2	3	0	1	2	3	0	1	2	3
0	40.12	183.35	326.59	469.82	51.52	142.72	292.14	465.05	53.94	143.75	278.37	442.40
1	183.35	326.59	469.82	613.05	142.72	292.14	465.05	644.11	143.75	278.37	442.40	624.08
2	326.59	469.82	613.05	756.28	292.14	465.05	644.11	824.96	278.37	442.40	624.08	816.63
3	469.82	613.05	756.28	899.52	465.05	644.11	824.96	1006.48	442.40	624.08	816.63	1016.26
	$t=3$				$t=3$				$t=3$			
	BNB regression model				BPIG regression model				BPLN regression model			
	Worst profile				Worst profile				Worst profile			
$k_{1,j}/k_{2,j}$	0	1	2	3	0	1	2	3	0	1	2	3
0	39.26	155.51	271.75	388.00	51.97	122.99	234.97	366.53	52.68	126.40	230.59	354.57
1	155.51	271.75	388.00	504.24	122.99	234.97	366.53	504.45	126.40	230.59	354.57	490.66
2	271.75	388.00	504.24	620.49	234.97	366.53	504.45	644.48	230.59	354.57	490.66	634.38
3	388.00	504.24	620.49	736.73	366.53	504.45	644.48	785.36	354.57	490.66	634.38	783.16

From all Tables 4.5, 4.6, 4.8, 4.9, 4.10, 4.11, 4.12 and 4.13 we see that if the policyholder j has a claim free year, the premium rates reduce, whereas if they have one or more claims, the premium rates increase, resulting in bonus or malus respectively. For example, for the case when the expected value principle is used, we observe from 4.5 that a claim free policyholder for both types of claims $i = 1, 2$ will receive bonuses of 27.31%, 20.20% and 18.53% in the year $t = 1$ in the case of the BNB, BPIG and BPLN distributions respectively. Furthermore, the insureds who had $k_{1,j} = 2$ and $k_{2,j} = 1$ claims in the year $t = 1$ will have to pay a malus of 807.56%, 792.83% and 747.30% in the case of the BNB, BPIG and BPLN distributions respectively.

Regarding the case with covariates, we see from Table 4.10 that claim free individuals per claim type $i = 1, 2$ in the year $t = 3$ will receive bonuses of 48.16%, 36.19% and 33.69% with the Best profile, of 52.37%, 38.99% and 37.53% with the Average profile and 54.46%, 40.25% and 40.36% with the Worst profile in the case of the BNB, BPIG and BPLN regression models with varying dispersion respectively. Additionally, we see from Table 4.10 that policyholders who had $k_{1,j} = 2$ and $k_{2,j} = 1$ claims in the year $t = 3$ will have to pay maluses of 433.14%, 389.09% and 346.43% with the Best profile, of 457.69%, 398.68% and 366.63% with the Average profile and 350.02%, 291.05% and 276.88% with the Worst profile in the case of the BNB, BPIG and BPLN regression models with varying dispersion respectively.

Similarly, for the case when the variance principle is used, we observe from 4.6 that a claim free insured for both types of claims $i = 1, 2$ will receive a bonus of 60.78%, 49.11% and 46.06% in the year $t = 3$ in the case of the BNB, BPIG and BPLN distributions respectively. Also, the individuals who had $k_{1,j} = 2$ and $k_{2,j} = 3$ claims in the year $t = 3$ will have to pay a malus of 690.06%, 760.82% and 760.06% in the case of the BNB, BPIG and BPLN distributions respectively.

Regarding the case with covariates, we see from Table 4.12 that claim free insureds per claim type $i = 1, 2$ in the year $t = 2$ will receive bonuses of 45.07%, 37.26% and 34.00% with the Best profile, of 49.64%, 40.39% and 37.95% with the Average profile and of 50.61%, 40.04% and 39.09% with the Worst profile in the case of the BNB, BPIG and BPLN regression models with varying dispersion respectively. Furthermore, we see from Table 4.12 that policyholders who had $k_{1,j} = 2$ and $k_{2,j} = 3$ claims in the year $t = 2$ will have to pay maluses of 804.86%, 976.28% and 1013.15% with the Best profile, of 849.20%, 996.33% and 1020.00% with the Average profile and 680.60%, 752.92% and 756.22% with the Worst profile in the case of the BNB, BPIG and BPLN regression models with varying dispersion respectively.

Finally, it is worth noting that Tables 4.8, 4.9, 4.10, 4.11, 4.12 and 4.13 provide a more complete picture to the insurance company than 4.5 and 4.6, when only the a posteriori criteria were considered, since all the important a priori and a posteriori information for the number of bodily injury and property damage claims, $k_{1,j}$ and $k_{2,j}$ respectively, of policyholder j, are considered in order to estimate their risk of having an accident and thus they permit the differentiation of the a posteriori, or Bonus-Malus, premia for various numbers of bodily injury and property damage claims by updating the posterior mean and the posterior variance based on all available information on the level of riskiness of this individual.

4.5.4 Computational aspects

This subsection discusses the computational issues related to the implementation of the EM algorithm for the BNB, BPIG and BPLN regression models with varying dispersion. All computing was made using the programming language R. A rather strict criterion was used and it took the algorithm, for both the cases with and without covariate information, a quite large number of iterations to converge. In particular, the stopping criterion was set as $tol = 10^{-12}$.

We also call attention to the fact that, because the M-Step involves three Newton-Raphson iterations, the choice of meaningful initial values for the vectors of regression coefficients $\boldsymbol{\beta_1}$, $\boldsymbol{\beta_2}$ and $\boldsymbol{\beta_3}$ of all three bivariate mixed Poisson models is important, as it can influence the speed of convergence of the algorithm and its ability to locate the global maximum. Good starting values for the regression parameters $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$ were obtained by fitting two simple Poisson regressions. Alternatively, the initial values can be obtained based on the data as follows: (i) calculate $\mathbb{E}\left(k_{i,j}\right)$, with $i = 1, 2$ and $j = 1, ..., n$, for the 18 different risk classes, which can be formed by dividing the portfolio into clusters defined by the combinations of the available explanatory variables in Table 3.1 and (ii) assuming log-link functions for $\boldsymbol{\varepsilon_{i,j}}$, see Eqs. (4.3 and 4.4), solve Eq.(4.6) with respect to $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$, in the case $i = 1$ and $i = 2$ respectively, since, under the parameterisation we adopted, the marginal means are explicit parameters of the bivariate mixed Poisson models with varying dispersion.

Furthermore, meaningful initial values for the regression parameters $\boldsymbol{\beta_3}$ were obtained by: (i) calculating $\text{Corr}\left(k_{1,j}, k_{2,j} | \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \boldsymbol{x_{3,j}}\right)$ for the 18 different risk classes based on all observations $j = 1, ..., n$, (ii) calculating $\mathbb{E}\left(k_{i,j} | \boldsymbol{x_{i,j}}\right)$ with $i = 1, 2$ for the 18 different risk classes (or alternatively calculating $\mu_{i,j}$, for $i = 1, 2$, based on the initial values for $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$ and on the log-link functions given by Eqs. (4.3 and 4.4)), (iii) solving Eq. (4.9) with respect to $\text{Var}(\lambda_j) > 0$ and subsequently (iv) solving the Eqs. (4.12, 4.15 and 4.18) with respect to $\boldsymbol{\gamma_j}$ and using the log-link function for $\boldsymbol{\gamma_j}$, see Eq. (4.5) in the case of the BNB, BPIG and BPLN models respectively.

Additionally, the standard errors were obtained using the standard approach of Louis (1982) for the EM algorithm. Finally, in terms of computational time requirements, the BNB and BPIG distributions/regression models with varying dispersion were significantly less demanding than the BPLN distribution/regression model with varying dispersion because the numerical evaluation of the integrals is time consuming especially when regression structures are used for all the parameters of the model.

4.6 Concluding remarks

In this chapter, we introduced a general class of bivariate mixed Poisson regression models with varying dispersion which can efficiently capture overdispersion and accurately account for the strength of the positive correlation between MTPL bodily injury and property damage claims by offering full flexibility in the choice of marginals and by utilizing all the available information from important risk factors through regression specifications for both mean parameters and the dispersion parameter of the models.

Our main contribution is that we developed an EM type algorithm which can reduce the computational burden for ML estimation for our family of models, the majority of which have cumbersome densities. In order to illustrate the versatility of the EM estimation scheme we presented, we fitted three members of this family, the BNB, BPIG and BPLN regression models with varying dispersion, on two-dimensional MTPL data from a European insurance company. Also, reliable estimates for the standard errors of the parameters of these models were obtained through expressions which were directly produced by the EM algorithm for the observed information matrix of each model.

Furthermore, the proposed family of models combined with the adopted modeling framework can provide insurance companies with a useful tool (from a practical business point of view) for pricing motor insurance contracts when the dynamics for premium determination are governed by the interactions of the different types of MTPL claims. In our real data application, the Bonus-Malus premia resulting from the BNB, BPIG and BPLN models were computed via the expected value and variance principles, providing alternative options to the insurer when deciding on their ratemaking strategies.

Additionally, it is worth noting that this family of models is suitable for applications not only on bivariate MTPL insurance ratemaking purposes but also in various multivariate domains, as these models can be easily generalized to any vector size response variable providing thus a very flexible way of modelling overdispersed high-dimensional count valued data which contain variables that exhibit complex positive dependencies.

Finally, an interesting future research direction would be to tackle Bonus-Malus ratemaking based on generalizations of the proposed family of models, such as, for example, by adding different random effects for modelling the unobserved heterogeneity when dealing with different types of claims from different types of coverage, see Bermúdez and Karlis (2017), or, for instance, by including time series components to take into account both cross dependence between different types of claims and time dependence, see Bermúdez et al. (2018).

Bibliography

Abdallah, A., Boucher, J.-P., and Cossette, H. (2016). Sarmanov family of multivariate distributions for bivariate dynamic claim counts model. *Insurance: Mathematics and Economics*, 68:120–133.

Aguero-Valverde, J. and Jovanis, P. P. (2009). Bayesian multivariate Poisson Lognormal models for crash severity modeling and site ranking. *Transportation Research Record*, 2136(1):82–91.

Ahmad, I. S., Hassan, A., and Ahmad, P. B. (2019). Negative Binomial-reciprocal Inverse Gaussian distribution: Statistical properties with applications. *arXiv preprint arXiv:1906.03224*.

Aitchison, J. and Ho, C. (1989). The multivariate Poisson-Lognormal distribution. *Biometrika*, 76(4):643–653.

Al-Hussaini, E. and Ahmad, K. (1981). On the identifiability of finite mixtures of distributions (corresp.). *IEEE Transactions on Information Theory*, 27(5):664–668.

Alfò, M. and Trovato, G. (2004). Semiparametric mixture models for multivariate count data, with application. *The Econometrics Journal*, 7(2):426–454.

Ambrosio, L., Fusco, N., and Pallara, D. (2000). *Functions of bounded variation and free discontinuity problems*, volume 254. Clarendon Press Oxford.

Anderson, D. and Burnham, K. (2004). Model selection and multi-model inference. *Second. NY: Springer-Verlag*, 63(2020):10.

Arnold, B. C. and Brockett, P. L. (1983). Identifiability for dependent multiple decrement/competing risk models. *Scandinavian Actuarial Journal*, 1983(2):117–127.

Arnold, B. C., Castillo, E., Sarabia, J.-M., and Sarabia, J. M. (1999). *Conditional specification of statistical models*. Springer Science & Business Media.

Avanzi, B., Cassar, L. C., and Wong, B. (2011). Modelling dependence in insurance claims processes with Lévy copulas. *UNSW Australian School of Business Research Paper*, (2011ACTL01).

Balakrishnan, N. (2000). *Continuous Multivariate Distributions: Vol. 1. Models and Applications*. Wiley.

Balakrishnan, N. (2014). Discrete Multivariate Distributions. *Wiley StatsRef: Statistics Reference Online*.

Ballotta, L. and Bonfiglioli, E. (2016). Multivariate Asset models using Lévy processes and applications. *The European Journal of Finance*, 22(13):1320–1350.

Ballotta, L., Deelstra, G., and Rayée, G. (2015a). Quanto implied correlation in a multi-Lévy framework.

Ballotta, L., Deelstra, G., and Rayée, G. (2017). Multivariate FX models with jumps: Triangles, quantos and implied correlation. *European Journal of Operational Research*, 260(3):1181–1199.

Ballotta, L. and Fusai, G. (2015). Counterparty credit risk in a multivariate structural model with jumps. *Finance*, 36(1):39–74.

Ballotta, L., Fusai, G., Loregian, A., and Perez, M. F. (2019a). Estimation of multivariate asset models with jumps. *Journal of Financial and Quantitative Analysis*, 54(5):2053–2083.

Ballotta, L., Fusai, G., and Marazzina, D. (2019b). Integrated structural approach to credit value adjustment. *European Journal of Operational Research*, 272(3):1143–1157.

Ballotta, L., Loregian, A., and Fusai, G. (2015b). Multivariate Lévy models by linear combination: Estimation.

Barndorff-Nielsen, O. (1965). Identifiability of Mixtures of Exponential Families. *Journal of Mathematical Analysis and Applications*, 12(1):115–121.

Barndorff-Nielsen, O. E. and Lindner, A. M. (2007). Lévy copulas: dynamics and transforms of Upsilon type. *Scandinavian Journal of Statistics*, 34(2):298–316.

Barreto-Souza, W. and Simas, A. B. (2016). General mixed Poisson regression models with varying dispersion. *Statistics and Computing*, 26(6):1263–1280.

Bäuerle, N. and Blatter, A. (2011). Optimal control and dependence modeling of insurance portfolios with Lévy dynamics. *Insurance: Mathematics and Economics*, 48(3):398–405.

Bäuerle, N., Blatter, A., and Müller, A. (2008). Dependence properties and comparison results for Lévy processes. *Mathematical Methods of Operations Research*, 67(1):161–186.

Bermúdez, L. (2009). A priori ratemaking using bivariate Poisson regression models. *Insurance: Mathematics and Economics*, 44(1):135–141.

Bermúdez, L., Guillén, M., and Karlis, D. (2018). Allowing for time and cross dependence assumptions between claim counts in ratemaking models. *Insurance: Mathematics and Economics*, 83:161–169.

Bermúdez, L. and Karlis, D. (2011). Bayesian multivariate Poisson models for insurance ratemaking. *Insurance: Mathematics and Economics*, 48(2):226–236.

Bermúdez, L. and Karlis, D. (2012). A finite mixture of bivariate Poisson regression models with an application to insurance ratemaking. *Computational Statistics & Data Analysis*, 56(12):3988–3999.

Bermúdez, L. and Karlis, D. (2017). A posteriori ratemaking using bivariate Poisson models. *Scandinavian Actuarial Journal*, 2017(2):148–158.

Bertoin, J. (1996). *Lévy processes*, volume 121. Cambridge University Press.

Böcker, K. and Klüppelberg, C. (2008). Modelling and measuring multivariate operational risk with Lévy copulas. *J. Operational Risk*, 3(2):3–27.

Böcker, K. and Klüppelberg, C. (2010). Multivariate models for operational risk. *Quantitative Finance*, 10(8):855–869.

Bolancé, C., Guillen, M., and Pitarque, A. (2020). A Sarmanov distribution with Beta marginals: An application to motor insurance pricing. *Mathematics*, 8(11).

Bolancé, C. and Vernic, R. (2019). Multivariate count data generalized linear models: Three approaches based on the Sarmanov distribution. *Insurance: Mathematics and Economics*, 85:89–103.

Bollerslev, T., Todorov, V., and Li, S. Z. (2013). Jump tails, extreme dependencies, and the distribution of stock returns. *Journal of Econometrics*, 172(2):307–324.

Booth, J. G. and Hobert, J. P. (1999). Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(1):265–285.

Booth, J. G., Hobert, J. P., and Jank, W. (2001). A survey of Monte Carlo algorithms for maximizing the likelihood of a two-stage hierarchical model. *Statistical Modelling*, 1(4):333–349.

Boucher, J.-P. and Denuit, M. (2008). Credibility premiums for the zero-inflated Poisson model and new hunger for bonus interpretation. *Insurance: Mathematics and Economics*, 42(2):727–735.

Boucher, J.-P., Denuit, M., and Guillen, M. (2008). Models of insurance claim counts with time dependence based on generalization of Poisson and Negative Binomial distributions. *Variance*, 2(1):135–162.

Brouhns, N., Guillén, M., Denuit, M., and Pinquet, J. (2003). Bonus-Malus scales in segmented tariffs with stochastic migration between segments. *Journal of Risk and Insurance*, 70(4):577–599.

Bücher, A., Vetter, M., et al. (2013). Nonparametric inference on Lévy measures and copulas. *Annals of Statistics*, 41(3):1485–1515.

Calderín-Ojeda, E. and Gómez-Déniz, E. (2019). The multivariate Negative Binomial-Lindley distribution: Properties and new representation for the univariate case. *Journal of Computational and Applied Mathematics*, 347:36–48.

Cameron, A. C., Li, T., Trivedi, P. K., and Zimmer, D. M. (2004). Modelling the differences in counted outcomes using bivariate copula models with application to mismeasured counts. *The Econometrics Journal*, 7(2):566–584.

Cameron, A. C. and Trivedi, P. K. (2013). *Regression analysis of count data*, volume 53. Cambridge University Press.

Carr, P., Geman, H., Madan, D. B., and Yor, M. (2002). The fine structure of asset returns: An empirical investigation. *The Journal of Business*, 75(2):305–332.

Chib, S. and Winkelmann, R. (2001). Markov chain Monte Carlo analysis of correlated count data. *Journal of Business & Economic Statistics*, 19(4):428–435.

Cole, T. J. and Green, P. J. (1992). Smoothing reference centile curves: the LMS method and penalized likelihood. *Statistics in Medicine*, 11(10):1305–1319.

Cook, R. J., Lawless, J. F., and Lee, K.-A. (2010). A copula-based mixed Poisson model for bivariate recurrent events under event-dependent censoring. *Statistics in Medicine*, 29(6):694–707.

Deelstra, G. and Simon, M. (2017). Multivariate European option pricing in a Markov-modulated Lévy framework. *Journal of Computational and Applied Mathematics*, 317:171–187.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22.

Denuit, M., Guillen, M., Trufin, J., et al. (2019). Multivariate credibility modelling for usage-based motor insurance pricing with behavioural data. *Annals of Actuarial Science*, 13(2):378–399.

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

Dionne, G. and Vanasse, C. (1992). Automobile insurance ratemaking in the presence of asymmetrical information. *Journal of Applied Econometrics*, 7(2):149–165.

Dunn, P. K. and Smyth, G. K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, 5(3):236–244.

Durante, F. and Sempi, C. (2010). Copula theory: an introduction. In *Copula theory and its applications*, pages 3–31. Springer.

El-Bassiouny, A. and Jones, M. (2009). A bivariate F distribution with marginals on arbitrary numerator and denominator degrees of freedom, and related bivariate Beta and t distributions. *Statistical Methods and Applications*, 18(4):465–481.

El-Basyouny, K. and Sayed, T. (2009). Collision prediction models using multivariate Poisson-Lognormal regression. *Accident Analysis & Prevention*, 41(4):820–828.

Esmaeili, H. and Klüppelberg, C. (2011). Parametric estimation of a bivariate stable Lévy process. *Journal of Multivariate Analysis*, 102(5):918–930.

Esmaeili, H. and Klüppelberg, C. (2013). Two-Step Estimation of a Multi-Variate Lévy process. *Journal of Time Series Analysis*, 34(6):668–690.

Famoye, F. (2010). A new bivariate Generalized Poisson distribution. *Statistica Neerlandica*, 64(1):112–124.

Famoye, F. (2012). Comparisons of some bivariate regression models. *Journal of Statistical Computation and Simulation*, 82(7):937–949.

Fang, K.-T., Kotz, S., and Ng, K. W. (2018). *Symmetric multivariate and related distributions*. Chapman and Hall/CRC.

Féron, R. (1956). Sur les tableaux de corrélation dont les marges sont données, cas de l'espace à trois dimensions. *Publ. Inst. Statist. Univ. Paris*, 5:3–12.

Fiorani, F. (1999). The Variance-Gamma process for option pricing. *Journal of Computational Finance*, 2(4):61–73.

Fiorani, F. (2004). Option pricing under the Variance Gamma process. *Available at SSRN 1411741*.

Fréchet, M. (1951). Sur les tableaux de corrélation dont les marges sont données. *Ann. Univ. Lyon, 3ˆ e serie, Sciences, Sect. A*, 14:53–77.

Fréchet, M. (1956). Sur les tableaux de corrélation dont les marges sont données. *Comptes Rendus Hebdomadaires Des Seances De L Academie Des Sciences*, 242(20):2426–2428.

Fung, T. C., Badescu, A. L., and Lin, X. S. (2019a). A class of mixture of experts models for general insurance: Application to correlated claim frequencies. *ASTIN Bulletin: The Journal of the IAA*, 49(3):647–688.

Fung, T. C., Badescu, A. L., and Lin, X. S. (2019b). A class of mixture of experts models for general insurance: Theoretical developments. *Insurance: Mathematics and Economics*, 89:111–127.

Gençtürk, Y. and Yiğiter, A. (2016). Modelling claim number using a new mixture model: Negative Binomial Gamma distribution. *Journal of Statistical Computation and Simulation*, 86(10):1829–1839.

Genest, C., Gendron, M., and Bourdeau-Brien, M. (2009). The advent of Copulas in Finance. *The European Journal of Finance*, 15(7-8):609–618.

Ghitany, M., Karlis, D., Al-Mutairi, D., and Al-Awadhi, F. (2012). An EM algorithm for multivariate mixed Poisson regression models and its application. *Applied Mathematical Sciences*, 6(137):6843–6856.

Gómez-Déniz, E. and Calderín-Ojeda, E. (2017). An alternative representation of the Negative Binomial-Lindley distribution. New results and applications. *arXiv preprint arXiv:1703.04812*.

Gómez-Déniz, E. and Calderín-Ojeda, E. (2018). Multivariate credibility in Bonus-Malus systems distinguishing between different types of claims. *Risks*, 6(2):34.

Gómez-Déniz, E., Hernández, A., Pérez, J. M., and Vázquez-Polo, F. J. (2002). Measuring sensitivity in a Bonus-Malus system. *Insurance: Mathematics and Economics*, 31(1):105–113.

Gómez-Déniz, E., Hernández-Bastida, A., and Fernández-Sánchez, M. P. (2014). Computing credibility Bonus-Malus premiums using the total claim amount distribution. *Hacettepe Journal of Mathematics and Statistics*.

Gómez-Déniz, E., Sarabia, J. M., and Calderín-Ojeda, E. (2008). Univariate and multivariate versions of the Negative Binomial-Inverse Gaussian distributions with applications. *Insurance: Mathematics and Economics*, 42(1):39–49.

Gómez-Déniz, E. and Vázquez-Polo, F. (2003). Robustness in Bayesian models for Bonus-Malus systems. In *Intelligent And Other Computational Techniques In Insurance: Theory and Applications*, pages 435–463. World Scientific.

Gómez-Déniz, E., Vázquez Polo, F. J., and Bastida, A. H. (2000). Robust Bayesian premium principles in actuarial science. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 49(2):241–252.

Grothe, O. (2013). Jump tail dependence in Lévy copula models. *Extremes*, 16(3):303–324.

Grothe, O. and Nicklas, S. (2013). Vine constructions of Lévy copulas. *Journal of Multivariate Analysis*, 119:1–15.

Gumbel, E. J. (1958). Distributions à plusieurs variables dont les marges sont données. *Comptes Rendus Hebdomadaires Des Seances De L Academie Des Sciences*, 246(19):2717–2719.

Gurland, J. (1957). Some interrelations among compound and generalized distributions. *Biometrika*, 44(1/2):265–268.

Gurmu, S. and Elder, J. (2000). Generalized bivariate count data regression models. *Economics Letters*, 68(1):31–36.

Hanbali, H. and Linders, D. (2019). American-type basket option pricing: a simple two-dimensional partial differential equation. *Quantitative Finance*, 19(10):1689–1704.

Heilmann, W.-R. (1989). Decision theoretic foundations of credibility theory. *Insurance: Mathematics and Economics*, 8(1):77–95.

Hirsa, A. and Madan, D. B. (2004). Pricing American options under Variance Gamma. *Journal of Computational Finance*, 7(2):63–80.

Ho, L. L. and Singer, J. d. M. (2001). Generalized least squares methods for bivariate Poisson regression. *Communications in Statistics - Theory and Methods*, 30(2):263–277.

Hutchinson, T. P. (1990). Continuous bivariate distributions emphasising applications. Technical report.

Insurance Europe (2015). European motor insurance markets, addendum. *Accessed*, 15:2017.

Insurance Europe (2019). Insurance fraud: not a victimless crime. *INSURANCE EUROPE [cit. 2020-04-06]. Dostupné z: https://www. insuranceeurope. eu/sites/default-/files/attachments/Insurance% 20fraud.*

Jinghui, Z. (2009). *Multi-asset option pricing with Lévy process.* PhD thesis, National University of Singapore.

Joe, H. (1997). *Multivariate Models and Dependence Concepts.* Chapman & Hall, London.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1997). *Discrete multivariate distributions*, volume 165. Wiley New York.

Jones, M. (2002). A dependent bivariate t distribution with marginals on different degrees of freedom. *Statistics & probability letters*, 56(2):163–170.

Jørgensen, B. (1982). Statistical properties of the Generalized Inverse Gaussian distribution. *Lecture Notes in Statistics*, 9.

Jung, R. C. and Winkelmann, R. (1993). Two aspects of labor mobility: a bivariate Poisson regression approach. *Empirical Economics*, 18(3):543–556.

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

Karlis, D. (2001). A general EM approach for maximum likelihood estimation in mixed Poisson regression models. *Statistical Modelling*, 1(4):305–318.

Karlis, D. (2005). EM algorithm for mixed Poisson and other discrete distributions. *ASTIN Bulletin: The Journal of the IAA*, 35(1):3–24.

Karlis, D. and Meligkotsidou, L. (2005). Multivariate Poisson regression with covariance structure. *Statistics and Computing*, 15(4):255–265.

Karlis, D., Tzougas, G., and Frangos, N. (2018). Confidence intervals of the premiums of optimal Bonus Malus systems. *Scandinavian Actuarial Journal*, 2018(2):129–144.

Katti, S. (1966). Interrelations among generalized distributions and their components. *Biometrics*, 22(1):44–52.

Kemp, C. and Papageorgiou, H. (1982). Bivariate Hermite distributions. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 269–280.

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

Kocherlakota, S. (1988). On the compounded bivariate Poisson distribution: A unified treatment. *Annals of the Institute of Statistical Mathematics*, 40(1):61–76.

Kocherlakota, S. and Kocherlakota, K. (2001). Regression in the bivariate Poisson distribution. *Communications in Statistics - Theory and Methods*, 30(5):815–825.

Kocherlakota, S. and Kocherlakota, K. (2004). Bivariate discrete distributions. *Encyclopedia of Statistical Sciences.*

Kotz, S., Balakrishnan, N., and Johnson, N. L. (2004). *Continuous multivariate distributions, Volume 1: Models and Applications.* John Wiley & Sons.

Kotz, S. and Nadarajah, S. (2004). *Multivariate t-distributions and their applications.* Cambridge University Press.

Krummenauer, F. (1998). Limit theorems for multivariate discrete distributions. *Metrika*, 47(1):47–69.

Kyprianou, A. E. (2014). *Fluctuations of Lévy processes with applications: Introductory Lectures.* Springer Science & Business Media.

Laeven, R. (2009). Non-parametric estimation of the Lévy copula.

Lai, C. (2004). Constructions of continuous bivariate distributions. *Journal of the Indian Society for Probability and Statistics*, 8:21–43.

Lai, C.-D. (2006). Constructions of discrete bivariate distributions. In *Advances in Distribution Theory, Order Statistics, and Inference*, pages 29–58. Springer.

Lee, A. (1999). Applications: Modelling rugby league data via bivariate Negative Binomial regression. *Australian & New Zealand Journal of Statistics*, 41(2):141–152.

Lemaire, J. (1995). Bonus-Malus systems in automobile insurance. *Kluwer Academic Publishers.*

Linders, D. and Schoutens, W. (2014). Basket option pricing and implied correlation in a Lévy copula model. *FEB Research Report AFI_1494*, pages 1–36.

Linders, D. and Stassen, B. (2016). The multivariate Variance Gamma model: basket option pricing and calibration. *Quantitative Finance*, 16(4):555–572.

Lindsay, B. G. and Roeder, K. (1993). Uniqueness of estimation and identifiability in mixture models. *Canadian Journal of Statistics*, 21(2):139–147.

Longstaff, F. A. and Schwartz, E. S. (2001). Valuing American options by simulation: a simple least-squares approach. *The Review of Financial Studies*, 14(1):113–147.

Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 44(2):226–233.

Luciano, E. and Schoutens, W. (2006). A multivariate jump-driven financial asset model. *Quantitative Finance*, 6(5):385–402.

Lynch, J. (1988). Mixtures, generalized convexity and balayages. *Scandinavian Journal of Statistics*, pages 203–210.

Ma, J., Kockelman, K. M., and Damien, P. (2008). A multivariate Poisson-Lognormal regression model for prediction of crash counts by severity, using Bayesian methods. *Accident Analysis & Prevention*, 40(3):964–975.

Madan, D. B., Carr, P. P., and Chang, E. C. (1998). The Variance Gamma process and option pricing. *Review of Finance*, 2(1):79–105.

Madan, D. B. and Seneta, E. (1990). The Variance Gamma (VG) model for share market returns. *Journal of Business*, pages 511–524.

Marshall, A. W. and Olkin, I. (1967). A multivariate Exponential distribution. *Journal of the American Statistical Association*, 62(317):30–44.

Marshall, A. W. and Olkin, I. (2007). *Life distributions*, volume 13. Springer.

McLachlan, G. J. and Krishnan, T. (2007). *The EM algorithm and extensions*, volume 382. John Wiley & Sons.

McNeil, A. J., Nešlehová, J., et al. (2009). Multivariate archimedean copulas, d-monotone functions and l_1-norm symmetric distributions. *The Annals of Statistics*, 37(5B):3059–3097.

Mert, M. and Saykan, Y. (2005). On a Bonus-Malus system where the claim frequency distribution is Geometric and the claim severity distribution is Pareto. *Hacettepe Journal of Mathematics and Statistics*, 34:75–81.

Mikosch, T. (2006). Copulas: Tales and facts. *Extremes*, 9(1):3–20.

M'Kendrick, A. (1925). Applications of mathematics to medical problems. *Proceedings of the Edinburgh Mathematical Society*, 44:98–130.

Munkin, M. K. and Trivedi, P. K. (1999). Simulated maximum likelihood estimation of multivariate mixed-Poisson regression models, with application. *The Econometrics Journal*, 2(1):29–48.

Nelsen, R. B. (2007). *An Introduction to Copulas*. Springer Science & Business Media.

Ni, W., Constantinescu, C., and Pantelous, A. A. (2014a). Bonus-Malus systems with Weibull distributed claim severities. *Annals of Actuarial Science*, 8(2):217–233.

Ni, W., Li, B., Constantinescu, C., and Pantelous, A. A. (2014b). Bonus-Malus systems with hybrid claim severity distributions. In *Vulnerability, Uncertainty, and Risk: Quantification, Mitigation, and Management*, pages 1234–1244.

Nikoloulopoulos, A. K. (2013a). Copula-based models for multivariate discrete response data. In *Copulae in Mathematical and Quantitative Finance*, pages 231–249. Springer.

Nikoloulopoulos, A. K. (2013b). On the estimation of Normal copula discrete regression models using the continuous extension and simulated likelihood. *Journal of Statistical Planning and Inference*, 143(11):1923–1937.

Nikoloulopoulos, A. K. (2016). Efficient estimation of high-dimensional multivariate Normal copula models with discrete spatial responses. *Stochastic Environmental Research and Risk Assessment*, 30(2):493–505.

Nikoloulopoulos, A. K., Joe, H., and Chaganty, N. R. (2011). Weighted scores method for regression models with dependent data. *Biostatistics*, 12(4):653–665.

Nikoloulopoulos, A. K. and Karlis, D. (2009a). Finite normal mixture copulas for multivariate discrete data modeling. *Journal of Statistical Planning and Inference*, 139(11):3878–3890.

Nikoloulopoulos, A. K. and Karlis, D. (2009b). Modeling multivariate count data using copulas. *Communications in Statistics - Simulation and Computation*, 39(1):172–187.

Nikoloulopoulos, A. K. and Karlis, D. (2010). Regression in a copula model for bivariate count data. *Journal of Applied Statistics*, 37(9):1555–1568.

Olkin, I. and Liu, R. (2003). A bivariate Beta distribution. *Statistics & Probability Letters*, 62(4):407–412.

Ord, J. (1972). Families of Frequency Distributions. *London. zbMATH.*

Palmes, C., Funke, B., and Hosseini, B. S. (2018). Nonparametric low-frequency Lévy copula estimation in a general framework. *Journal of Nonparametric Statistics*, 30(3):523–555.

Park, E. S. and Lord, D. (2007). Multivariate Poisson-Lognormal models for jointly modeling crash frequency by severity. *Transportation Research Record*, 2019(1):1–6.

Pechon, F., Denuit, M., and Trufin, J. (2019a). Home and motor insurance joined at a household level using multivariate credibility. *Annals of Actuarial Science*, pages 1–33.

Pechon, F., Denuit, M., and Trufin, J. (2019b). Multivariate modelling of multiple guarantees in motor insurance of a household. *European Actuarial Journal*, 9(2):575–602.

Pechon, F., Trufin, J., Denuit, M., et al. (2018). Multivariate modelling of household claim frequencies in motor third-party liability insurance. *ASTIN Bulletin: The Journal of the IAA*, 48(3):969–993.

Picech, L. (1994). The merit-rating factor in a multiplicating rate-making model. *ASTIN Colloquium, Cannes.*

Pinquet, J. (1997). Allowance for cost of claims in Bonus-Malus systems. *ASTIN Bulletin: The Journal of the IAA*, 27(1):33–57.

Pinquet, J. (1998). Designing optimal Bonus-Malus systems from different types of claims. *ASTIN Bulletin: The Journal of the IAA*, 28(2):205–220.

Rachev, S. T., Sun, W., and Stein, M. (2009). Copula Concepts in Financial Markets. *Portfolio Institutionell (Ed.)*, 4:12–15.

Raftery, A. E. (1984). A continuous multivariate Exponential distribution. *Communications in Statistics - Theory and Methods*, 13(8):947–965.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, pages 111–163.

Rigby, R., Stasinopoulos, D., and Akantziliotou, C. (2008). A framework for modelling overdispersed count data, including the Poisson-shifted Generalized Inverse Gaussian distribution. *Computational Statistics & Data Analysis*, 53(2):381–393.

Rigby, R. A. and Stasinopoulos, D. (1996a). A semi-parametric additive model for variance heterogeneity. *Statistics and Computing*, 6(1):57–65.

Rigby, R. A. and Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(3):507–554.

Rigby, R. A. and Stasinopoulos, M. D. (1996b). Mean and dispersion additive models. In *Statistical Theory and Computational Aspects of Smoothing*, pages 215–230. Springer.

Salinetti, G., Dall'Aglio, G., and Kotz, S. (1991). *Advances in probability distributions with given marginals: beyond the copulas*. Kluwer.

Santi, D., Purnaba, I., and Mangku, I. (2016). Bonus-Malus system with the claim frequency distribution is Geometric and the severity distribution is truncated Weibull. In *IOP Conference Series: Earth and Environmental Science*, volume 31. IOP Publishing.

Sapatinas, T. (1995). Identifiability of mixtures of power-series distributions and related characterizations. *Annals of the Institute of Statistical Mathematics*, 47(3):447–459.

Sarabia, J. M. and Castillo, E. (2006). Bivariate distributions based on the generalized three-parameter Beta distribution. In *Advances in Distribution Theory, Order Statistics, and Inference*, pages 85–110. Springer.

Sarabia, J. M. and Gómez Déniz, E. (2008). Construction of multivariate distributions: a review of some recent results. *SORT*, 32(1):3–36.

Sarabia, J. M. and Gómez-Déniz, E. (2011). Multivariate poisson-beta distributions with applications. *Communications in Statistics - Theory and Methods*, 40(6):1093–1108.

Sarhan, A. M. and Balakrishnan, N. (2007). A new class of bivariate distributions and its mixture. *Journal of Multivariate Analysis*, 98(7):1508–1527.

Sarmanov, O. V. (1966). Generalized Normal correlation and two-dimensional Fréchet classes. In *Doklady Akademii Nauk*, volume 168, pages 32–35. Russian Academy of Sciences.

Sato, K.-I. (1999). *Lévy processes and infinitely divisible distributions*. Cambridge University Press.

Schoutens, W. (2003). *Lévy Processes in Finance: Pricing Financial Derivatives*. Wiley Series in Probability and Statistics.

Schweizer, B. and Sklar, A. (1983). Probabilistic Metric Spaces, North-Holland Series in Probability and Applied Mathematics.

Seneta, E. (2004). Fitting the Variance-Gamma model to financial data. *Journal of Applied Probability*, pages 177–187.

Seneta, E. (2007). The early years of the Variance-Gamma process. In *Advances in Mathematical Finance*, pages 3–19. Springer.

Seshadri, V. (1993). *The Inverse Gaussian Distribution: Statistical Theory and Applications*. Oxford Science Publications.

Shared, M. (1980). On mixtures from Exponential families. *Journal of the Royal Statistical Society: Series B (Methodological)*, 42(2):192–198.

Shaw, W. and Lee, K. (2008). Bivariate student t distributions with variable marginal degrees of freedom and independence. *Journal of Multivariate Analysis*, 99(6):1276–1287.

Shengwang, M., Wei, Y., and Whitmore, G. (1999). Accounting for individual over-dispersion in a Bonus-Malus automobile insurance system. *ASTIN Bulletin: The Journal of the IAA*, 29(2):327–337.

Shi, P. and Valdez, E. A. (2014). Multivariate Negative Binomial models for insurance claim counts. *Insurance: Mathematics and Economics*, 55:18–29.

Sichel, H. (1971). On a family of discrete distributions particularly suited to represent long-tailed frequency data. In *Proceedings of the third symposium on mathematical statistics*. SACSIR.

Silva, A., Rothstein, S. J., McNicholas, P. D., and Subedi, S. (2017). A multivariate Poisson-Lognormal mixture model for clustering transcriptome sequencing data. *arXiv preprint arXiv:1711.11190*.

Sklar, M. (1959). Fonctions de repartition an dimensions et leurs marges. *Publ. inst. statist. univ. Paris*, 8:229–231.

Stasinopoulos, M., Rigby, B., and Akantziliotou, C. (2008). Instructions on how to use the GAMLSS package in R. Second Edition.

Stein, G. Z. and Juritz, J. M. (1987). Bivariate compound Poisson distributions. *Communications in Statistics - Theory and Methods*, 16(12):3591–3607.

Stein, G. Z., Zucchini, W., and Juritz, J. M. (1987). Parameter estimation for the Sichel distribution and its multivariate extension. *Journal of the American Statistical Association*, 82(399):938–944.

Tallis, G. (1969). The identifiability of mixtures of distributions. *Journal of Applied Probability*, 6(2):389–398.

Tankov, P. (2003a). Dependence structure of spectrally positive multidimensional Lévy processes. *Unpublished manuscript (cf. www. cmap. polytechnique. fr/ tankov).*

Tankov, P. (2003b). *Financial Modelling with Jump Processes*. CRC press.

Tankov, P. (2004). *Lévy Processes in Finance: Inverse Problems and Dependence Modelling*. PhD thesis, Ecole Polytechnique X.

Tankov, P. (2006). Simulation and option pricing in Lévy copula models. *Mathematical Modelling of Financial Derivatives, IMA Volumes in Mathematics and Applications, Springer*.

Tankov, P. (2016). Lévy copulas: review of recent results. In *The Fascination of Probability, Statistics and their Applications*, pages 127–151. Springer.

Teicher, H. (1963). Identifiability of finite mixtures. *The Annals of Mathematical Statistics*, pages 1265–1269.

Ting Lee, M.-L. (1996). Properties and applications of the Sarmanov family of bivariate distributions. *Communications in Statistics-Theory and Methods*, 25(6):1207–1222.

Tonda, T. et al. (2005). A class of multivariate discrete distributions based on an approximate density in GLMM. *Hiroshima Mathematical Journal*, 35(2):327–349.

Tremblay, L. (1992). Using the Poisson Inverse Gaussian in Bonus-Malus systems. *ASTIN Bulletin: The Journal of the IAA*, 22(1):97–106.

Tseung, S. C., Badescu, A., Fung, T. C., and Lin, X. S. (2020). LRMoE: An R Package for Flexible Actuarial Loss Modelling Using Mixture of Experts Regression Model. *Available at SSRN 3740215*.

Tzougas, G. (2020). EM estimation for the Poisson-Inverse Gamma regression model with varying dispersion: an application to insurance ratemaking. *Risks*, 8(3):97.

Tzougas, G. and Frangos, N. (2014). The design of an optimal Bonus-Malus system based on the Sichel distribution. In *Modern Problems in Insurance Mathematics*, pages 239–260. Springer.

Tzougas, G., Hoon, W., and Lim, J. (2019). The Negative Binomial-Inverse Gaussian regression model with an application to insurance ratemaking. *European Actuarial Journal*, 9(1):323–344.

Tzougas, G. and Karlis, D. (2020). An EM algorithm for fitting a new class of mixed Exponential regression models with varying dispersion. *ASTIN Bulletin: The Journal of the IAA*.

Tzougas, G., Vrontos, S., and Frangos, N. (2018). Bonus-Malus systems with two-component mixture models arising from different parametric families. *North American Actuarial Journal*, 22(1):55–91.

Tzougas, G., Yik, W. H., and Mustaqeem, M. W. (2020). Insurance ratemaking using the Exponential-Lognormal regression model. *Annals of Actuarial Science*, 14(1):42–71.

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

Vernic, R. (1997). On the bivariate Generalized Poisson distribution. *ASTIN Bulletin: The Journal of the IAA*, 27(1):23–32.

Vernic, R. (2000). A multivariate generalization of the Generalized Poisson distribution. *ASTIN Bulletin: The Journal of the IAA*, 30(1):57–67.

Wang, P. (2003). A bivariate zero-inflated Negative Binomial regression model for count data with excess zeros. *Economics Letters*, 78(3):373–378.

Willmot, G. E. (1990). Asymptotic tail behaviour of Poisson mixtures by applications. *Advances in Applied Probability*, 22(1):147–159.

Winkelmann, R. (2008). *Econometric analysis of count data*. Springer Science & Business Media.

Wüthrich, M. V. (2019). From generalized linear models to neural networks, and back. *Available at SSRN 3491790*.

Xekalaki, E. (1981). Chance mechanisms for the univariate Generalized Waring distribution and related characterizations. In *Statistical Distributions in Scientific Work*, pages 157–171. Springer.

Xekalaki, E. and Panaretos, J. (1983). Identifiability of compound Poisson distributions. *Scandinavian Actuarial Journal*, 1983(1):39–45.

Yakowitz, S. J. and Spragins, J. D. (1968). On the identifiability of finite mixtures. *The Annals of Mathematical Statistics*, 39(1):209–214.

Zamani, H. and Ismail, N. (2010). Negative Binomial-Lindley distribution and its application. *Journal of Mathematics and Statistics*, 6(1):4–9.

Zhan, X., Aziz, H. A., and Ukkusuri, S. V. (2015). An efficient parallel sampling technique for multivariate Poisson-Lognormal model: Analysis with two crash count datasets. *Analytic Methods in Accident Research*, 8:45–60.

Zimmer, D. M. and Trivedi, P. K. (2006). Using trivariate copulas to model sample selection and treatment effects: application to family health care demand. *Journal of Business & Economic Statistics*, 24(1):63–76.