UNIVERSITE CATHOLIQUE DE LOUVAIN
INSTITUT DE STATISTIQUE

MODELLING DEPENDENCE IN ACTUARIAL SCIENCE, WITH EMPHASIS ON CREDIBILITY THEORY AND COPULAS

Membres du jury:

Prof. Pierre Ars
Prof. Michel Denuit *(Promoteur)*
Prof. Jan Dhaene (KULeuven)
Prof. Montserrat Guillén (Universitat de Barcelona)
Prof. Philippe Lambert
Prof. Jean-Marie Rolin *(Président du Jury)*
Prof. Ingrid Van Keilegom

Thèse présentée en vue de
l'obtention du grade de
Docteur en Sciences
(orientation statistique) par :

Oana Gabriela Purcaru

Louvain-la-Neuve, August 2005

to the memory of Professor Virgil Craiu

Acknowledgements

As each journey comes to an end, so does the amazingly fulfilling experience of writing a doctoral thesis. The path of this journey was not without obstacles, but they were all overcome by the immense satisfaction of making new enlightening personal discoveries in the area of statistics, some of which have been included between the covers of my thesis.

This milestone in my professional life could not have been reached without the enthusiastic guidance, support and encouragement of my supervisor, Professor Michel Denuit. I am immensely grateful for his sharing with me of his invaluable knowledge and passion for statistics, for his patience and availability and I look forward to continuing our collaboration in the future. I am also indebted to Professor Montserrat Guillen and Professor Ingrid Vankeilegom, with two of the chapters in this thesis being the fruit of our joint research. My special gratitude goes to all the other members of my doctoral committee; their guidance, remarks and suggestions were essential to improving my research.

A particular thank you goes to the informatics team for their precious suggestions in programming, as well as their support and patience, especially when I "blew up" the servers of the Institute with my computations months ago. (Fortunately, the remedy was rapidly found and everything got back to order.). I am also thankful to each member of the Institute of Statistics, for their kindness, great sense of hospitality and friendship, all of which have benefited my experience in the "Louvain-la-Neuve world of statistics".

This entire journey could not have been smoothened without the emotional support and continuous encouragements (and tons of patience!) of some special people. A big "Thank you" to Giovanna, Fausto and Maria, for your kind friendship; to my "neighbours"; to the "female group" of bright statisticians at the Institute; to all other friends I have made here, in Belgium; and, last but not least, to all my dear friends in Bucharest.

Finally, I am especially grateful to my grandmother Male, my parents, my brother Razvan, Adelina and the rest of my family, whose infinite moral support will continue to accompany me along the way.

Contents

Introduction

One basic problem in statistical science is to understand the relationships among multivariate outcomes. In that respect, regression analysis is an important tool because it allows researchers to focus on the effects of explanatory variables. In automobile insurance, for instance, regression models (Poisson regression, and more generally, Generalized Linear Models and nonlinear extensions) allow the actuaries to quantify the effect of observable characteristics of policyholders on the insured peril (and thus to adjust the amount of premium accordingly). Although it remains an important tool and is widely applicable, the regression analysis is limited by the basic setup that requires to identify one dimension of the outcomes as the primary measure of interest (the "dependent" variable) and other dimensions as supporting this variable (the "explanatory" variables).

There are situations where this relationship is not of primary interest. One might be interested in understanding the distribution of several outcomes in a multivariate setting. For instance, when a married couple has bought a joint life insurance or annuity policy, we are concerned with the joint distribution of lifetimes, since the valuation of the policy depends, among other factors, upon the probabilty of survivorship of the couple. Or, in automobile insurance, one might be interested to study the kind of dependence between annual claim numbers, which has an important impact on the premium paid by the policyholder. When such analysis is of interest, appropriate tools are needed in order to evaluate, to compare and to model the strength of dependence existing between different outcomes.

Since modelling random variables is based on probability theory, when comparing risks one might resort to stochastic orderings, which give an idea on how a random variable is

more dangerous than another. For example, given two random variables, we say that the first one is "smaller" (in a stochastic sense defined in the first part of the thesis), than the second one if it has the smallest probability of exceeding a given threshold. Among numerous stochastic orderings existing, we will use, in this thesis, two of them, namely the stochastic dominance and the likelihood ratio order. The latter is actually a sufficient condition to get the former. Besides comparing random variables, these stochastic orderings are also used to define dependence concepts (such as Positive or Negative Quadrant Dependent), which allow to evaluate the strength of dependence existing between different random variables.

An important amount of literature arose from the study of the stochastic orderings in general, as well as their application is actuarial sciences. For more details on this topic we refer e.g. to Lehmann (1955), Barlow & Proschan (1975), Mosler & Scarsini (1993), Shaked & Shanthikumar (1994), Szekeli (1995), Müller & Stoyan (2002). We also refer e.g. to Joe (1997), for more details about dependence concepts.

Now, if one is interested not in comparing random variables, but rather in modelling the dependence existing between them, he might need another statistical tool. Such a tool, is the copula. It allows the construction of multivariate distributions with given marginals. Once one chooses the appropriate marginals and plugs them into a suitable copula, he gets the multivariate distribution. For more details on copulas we refer e.g. to Joe (1997), Nelsen (1999). Modelling dependence between continuous outcomes by means of a particular class of copulas, the archimedean family, is the main topic of the second part of the thesis.

This thesis is devoted to modelling dependence with applications in actuarial sciences and is divided in two parts: the first one concerns dependence in frequency credibility models and the second one deals with dependence between continuous outcomes. As previously mentioned, to this end we will resort to different tools, the stochastic orderings and copulas, respectively.

Part I: Credibility models for claim frequencies

This part is devoted to the study of recently introduced models in frequency credibility theory, which can be seen as models for time series of count data, adapted to actuarial problems. More precisely, we are interested in the number of automobile accidents (also referred to as claim numbers or claim frequencies in this thesis). There is an important

amount of literature devoted to finding an appropriate model for such data, in actuarial sciences. For a review of the existing literature, we refer e.g. to Lemaire (1995) and Denuit (1997).

In this thesis, we will focus on the mixed Poisson model. Now, in this model, of main interest will be the study of the dependence induced among annual claim numbers by the introduction of random effects, representing some important factors which cannot be measured (such that aggressiveness behind the wheel), but which obviously influence the number of accidents. This will be done by means of stochastic orderings and positive dependence structures, which allow the comparison and, respectively, the formalization of the kind of dependence existing between random variables. Intuitively speaking the latter ones capture the fact that large (small) values of one random variable tend to be associated with large (small) values for the others.

Describing the dependence generated by actuarial credibility models by means of stochastic orderings represents one of the important contributions of this thesis, since this aspect, to our knowledge has never been investigated in the literature so far. Let us first start with a brief description of the context which made us to be interested in such topic.

Credibility theory can be seen as the art to combine different collections of data to obtain an accurate overall estimate. In many cases, a compromise estimator is derived from a convex combination of a prior mean and the mean of the current observations. The weight given to the observed mean is called the credibility factor (since it fixes the extent to which the actuary may be confident in the data). An excellent introduction to credibility theory can be found in Herzog (1994).

One of the main tasks of the actuary is to design a tariff structure that will fairly distribute the burden of claims among policyholders. If the risks in the portfolio are not all equal to each other (which means that the associated random financial losses have different distribution functions), it is fair to partition all policies into homogeneous classes with all policyholders belonging to the same class paying the same premium. In automobile third party liability insurance, examples of classification variables encountered in practice include the age, gender and occupation of the policyholders, the type and use of their car, the place where they reside and sometimes even the number of cars in the household, marital status, smoking behavior or the color of the vehicle. It is convenient to achieve *a priori*

classification with the help of generalized linear models; see e.g. Renshaw (1994) and Mc Cullagh & Nelder (1989) or Dobson (1990) for a general overview of the statistical theory.

However, many important factors cannot be taken into account at this stage; think for instance of swiftness of reflexes or aggressiveness behind the wheel in automobile insurance. Consequently, tariff cells are still quite heterogeneous. This residual heterogeneity can be represented by a random effect in a statistical model. The amount of premium charged to all policyholders in a risk class is thus itself an average, so that some policyholders pay too much and subsidize the others. The claims histories can be used to restore fairness in the risk classes, increasing the premium for policyholders reporting claims and decreasing those of good drivers. The allowance for the history of the policyholder in a rating model thus derives from interpretation of serial correlation for longitudinal data resulting from hidden features in the risk distribution.

During the last decade of the 20th century, the world of insurance was confronted with important developments of the *a posteriori* tarification, especially in the field of credibility. This was due to the easing of insurance markets in the European Union, which gave rise to an advanced segmentation. The first important contribution is due to Dionne & Vanasse (1989, 1992), who pointed out the great influence of the *a priori* risk classification on the size of the *a posteriori* corrections. More precisely, the discounts granted, when no claims were reported, were higher for the bad drivers than for the good ones (*bad* and *good* qualifications referring to the *a priori* tariff). Therefore these authors proposed a credibility model which integrates *a priori* and *a posteriori* information on an individual basis. They introduced a regression component in the Poisson counting model in order to use all available information in the estimation of accident frequency. The unexplained heterogeneity was then modeled by the introduction of a latent variable representing the influence of hidden policy characteristics.

Taking this random effect Gamma distributed yields the Negative Binomial model for the claim number. Of course, there is no particular reason to restrict ourselves to Gamma distributed random effects (except perhaps for mathematical convenience). In the biostatistical field, LogNormally distributed random effects are widely used (see also Pinquet (1997) for an application in actuarial science). The use of the Inverse Gaussian distribution has been advocated by Willmot (1987) in conjunction with Poisson mixtures; see also Tremblay

(1992). Another possible choice is Hoffman's distribution; see e.g. Kestemont and Paris (1985). Moreover, semi-parametric approaches (retaining the Poisson assumption for the claim numbers without specifying any distribution for the random effects) are conceivable in the spirit of the Bühlmann-Straub model. See e.g. Walhin & Paris (1999) as well as Denuit & Lambert (2001) for nonparametric maximum likelihood methods.

The vast majority of the papers that appeared in the actuarial literature considered time-independent (or static) heterogeneous models. Noticeable exceptions include the pioneering papers by Gerber & Jones (1975), Sundt (1988) and Pinquet, Guillén & Bolancé (2001, 2003). The allowance for an unknown underlying random parameter that evolves over time is justified since unobservable factors influencing the driving abilities are not constant. One might consider either shocks (induced by events like divorces or nervous breakdown, for instance) or continuous modifications (e.g. due to learning effect). Another reason to allow for random effects that vary with time relates to moral hazard. Indeed, individual efforts to prevent accidents are unobserved and feature temporal dependence. The policyholders may adjust their efforts for loss prevention according to their experience with past claims, the amount of premium and awareness of future consequences of an accident (due to experience rating schemes).

Let us consider that in a portfolio of an insurance company there are n policies during the observation period, each of them being observed during ν_i periods (measured in years for instance). Let $\boldsymbol{N}_i = \{N_{i1}, N_{i2}, \ldots, N_{i\nu_i}\}$ be the sequence of claim numbers reported by policyholder i, $i = 1, 2, \ldots, n$. For the same policyholder, the influence of hidden features (as annual mileage or aggressiveness behind the wheel for instance), i.e. unknown risk characteristics of the policyholder having a significant impact on the occurrence of claims, will be modelled by a vector, $\boldsymbol{\Theta}_i$, of positive random effects with unit mean. Thus the ith policy of the portfolio, $i = 1, 2, \ldots, n$, is represented by a sequence $(\boldsymbol{\Theta}_i, \boldsymbol{N}_i)$. It is essential to understand the philosophy of this classical actuarial construction. Here dependence between annual claim numbers is a consequence of the heterogeneity of the portfolio; the dependence is only apparent. If we had a complete knowledge of policy characteristics then $\boldsymbol{\Theta}_i$ would become deterministic and there would be no more dependence between the N_{ij}'s for fixed i. The unexplained heterogeneity (which has been modeled through the introduction of the vector of random effects $\boldsymbol{\Theta}_i$ for policyholder i) is then revealed by the claims and premiums

histories in a Bayesian way. These histories modify the distribution of $\mathbf{\Theta}_i$ and hence modify the premium.

Once estimated, the heterogenous model can be used to perform prediction on longitudinal data and allows experience rating in casualty insurance. In an empirical Bayesian setting, the prediction is derived from the expectation of a random effect with respect to a posterior distribution taking into account the history of the individual. An excellent introduction to these concepts can be found in Pinquet (2000).

As mentioned at the beginning, in this first part we will study the influence of the random effects, $\mathbf{\Theta}_i$ on the kind of dependence existing between annual claim numbers, the N_{ij}'s. We split the study in two parts, depending on how these random effects are, static (time-independent) or dynamic (time-dependent).

Static random effects

Since the random effects are constant over time, i.e. $\mathbf{\Theta}_i = (\Theta_i, \ldots, \Theta_i)$, the Poisson static credibility model will be based on the following assumptions:

A1 given $\Theta_i = \theta$, the random variables N_{ij}, $j = 1, 2, \ldots$, are independent and conform to the Poisson distribution with mean $\lambda_{ij}\theta$, i.e.

$$\mathbb{P}[N_{ij} = k | \Theta_i = \theta] = \exp(-\lambda_{ij}\theta)\frac{(\lambda_{ij}\theta)^k}{k!},$$

for $k \in \mathbb{N}, j = 1, 2, \ldots$;

A2 at the portfolio level, the sequences $(\mathbf{\Theta}_i, \boldsymbol{N}_i)$ are assumed to be independent.

The very aim of credibility theory is to predict future claims behaviour. In that respect predictive distributions are of prime interest: these are the distributions of claim characteristics for next year, given past observations. As the total claim numbers, $N_{i\bullet} = \sum_{j=1}^{\nu_i} N_{ij}$, is an exhaustive summary of past claims, we show that when $N_{i\bullet}$ increases, the *a posteriori* distribution of Θ_i "increases" in some stochastic sense. This result holds for any Poisson mixture model (i.e. for any chosen distribution for the random effects). We also prove that, in this model, the dependence between annual claim numbers is very strong, namely MPLRD (multivariate positive likelihood ratio dependence).

We then, extend this study to the Generalized Linear Mixed Models (GLMM's). GLMM's are widely used by actuaries, since they form the basis of credibility theory and bonus-malus systems. Building on Lee & Nelder's (1996) work, Nelder & Verrall (1997) showed how credibility theory can be encompassed within the theory of GLMM's. In this context, the variable of interest represents some observation related to the policy, i.e. claim frequency, loss ratio or claim severity (and thus following Poisson, Normal or Gamma law, respectively). The random effects represent hidden features influencing the risk covered by the insurer.

We first prove that equivalent results to those in the Poisson case, still hold in the GLMM framework. Since credibility theory deals with prediction, we were interested in investigating some features of the Bühlmann credibility premium, which gives a prediction of the claims in year ν_i+1, when only ν_i years are observed. For each of the three examples of interest (claim frequency, loss ratio or claim severity) we show that the Bühlmann credibility premium, π_{cred}, is linear in the total past claims. We then prove that increasing the linear credibility premium (i.e. deteriorating the claim record of the policyolder) makes larger the probability to observe more important losses in the future. This result relies on the fact that when the total past claims up to year ν_i are becoming larger, the claim characteristic for the next year, $\nu_i + 1$, is "increasing" in a stochastic sense, property which was also proven.

Dynamic random effects

The main technical interest of letting the random effects evolve over time (i.e. $\mathbf{\Theta}_i = (\Theta_{i1}, \Theta_{i2}, \ldots)$) is to take into account the date of claims. This reflects the fact that the predictive ability of a claim depends on its age: a recent claim is a worse sign to the insurer than a very old one. Contrarily to the static case, the total number of claims reported in the past is no more an exhaustive summary of policyholders' history. Rather, the sequence of annual claim numbers has now to be memorized to determine future premiums.

The Poisson dynamic credibility model relies on the following assumptions:

A1 given $\boldsymbol{\Theta}_i = \boldsymbol{\theta}_i$, the rv's N_{it}, $t = 1, 2, \ldots, \nu_i$, are independent and conform to the Poisson distribution with mean $\lambda_{it}\theta_{it}$, i.e.

$$\begin{aligned} \Pr[\boldsymbol{N}_i = \boldsymbol{k}_i | \boldsymbol{\Theta}_i = \boldsymbol{\theta}_i] &= \prod_{t=1}^{\nu_i} \Pr[N_{it} = k_{it} | \Theta_{it} = \theta_{it}] \\ &= \prod_{t=1}^{\nu_i} \exp(-\lambda_{it}\theta_{it}) \frac{(\lambda_{it}\theta_{it})^{k_{it}}}{k_{it}!}, \quad \boldsymbol{k}_i \in \mathbb{N}^{\nu_i}; \end{aligned}$$

A2 at the portfolio level, the sequences $(\boldsymbol{\Theta}_i, \boldsymbol{N}_i)$, $i = 1, 2, \ldots, n$, are assumed to be independent. Moreover, the Θ_{it}'s are non-negative rv's with unit mean. Defining $\nu_{\max} = \max_i \nu_i$, $\boldsymbol{\Theta}_i$ has the same distribution as the first ν_i components of some random vector $(\Theta_1, \ldots, \Theta_{\nu_{\max}})$.

In addition to the previous assumptions, we consider different structures for the random effects, which match the constraints enumerated in A2. We focus on three models based on the Log-Normal distribution and one model based on copulas (thus allowing to specify another distribution than the Log-Normal). Each of these models is then analyzed separately.

Since the random effects are evolving in time, the dependence existing between them will influence the dependence between annual claim numbers. We first study under which conditions the dependence between the random effects occurs. Then we show that the same dependence structure existing between the random effects is transmitted to the claim numbers. We also prove the influence of increasing claims $\boldsymbol{N}_i$ on the unobservable characteristics $\boldsymbol{\Theta}_i$ as well as on the predictive distribution of Θ_{i,ν_i+1} and the one of N_{i,ν_i+1}. This influence will be seen to be an "increasingness" in a stochastic sense.

Brouhns & Denuit (2003) complemented this work by considering Generalized Additive Mixed Models (GAMM 's), with dynamic random effects following the Multivariate Normal distribution. As mentioned by these authors, some of their results remain valid for other choices of distribution.

This first part of the thesis ends with a numerical illustration. We consider the Log-Normal as mixing distribution in the Poisson mixture model. More precisely, we focus on

three different structures for the random effects: static, autoregressive of order one (AR(1)) on the log-scale and exchangeable. We then fit these models to a large Spanish third part liability automobile data set. We examine the pattern of *a posteriori* corrections generated during a period of 10 years, from three points of vue: the age of claims, the *a priori* characteristics and the model chosen. It will be seen that a recent claim has a great impact on the revised premium than an older one. We will also remark that if one claim is reported during the first year of the 10 years considered, the lower the claim frequency, the higher the relative *a posteriori* correction is. In change, if no claim is reported over 10 years, the higher the claim frequency, the higher the relative *a posteriori* discount is.

Part II: Copula modelling

Whereas the interest in the first part was the analysis of the dependence for longitudinal count data (i.e. dependence between repetead measures for one individual) and its effect on some quantities depending on the data (such as insurance premiums), in this second part we focus on the dependence between different continuous outcomes.

A known dataset from non-life insurance, the loss-ALAE's, is analysed in this second part. It consists in couples of loss and allocated loss adjustment expenses, ALAE's in short, (like lawyers' fees and claims investigation expenses) on a single claim. Now, expensive claims generally need some time to be settled and induce considerable costs for the insurance company. Actuaries therefore expect that large values for losses will tend to be associated with large values for ALAE's. This positive association has some practical implications in the pricing of reinsurance treaties. The reinsurer covers the largest losses (i.e. those exceeding some high threshold and pays that part exceeding this threshold). He also contributes to pay the associated settlement costs on a prorata basis. Since expected reinsurer's payment is a function of loss and ALAE's, its computation depends upon the joint distribution function of these variables. Thus in many cases, neglecting the dependence exhibited by the data leads to serious underestimation of the expected reinsurer's payment. It is therefore crucial for the reinsurer to have an appropriate model for the random couple loss-ALAE at its diposal.

We will thus be interested in studying the dependence existing in this random couple and to this end we will resort to archimedean copulas. More precisely, our main contribution in

this second part consists in a semiparametric modelling strategy, which takes into account the particularity of the data, namely the consorship in the loss variable. We develop an appropriate nonparametric estimator for the joint distribution of loss-ALAE, which will be then used to identify the appropriate archimedean copula fitting the data. Let us first present why we did resort to copulas in our study and briefly describe some previous works on the loss-ALAE dataset.

The Normal distribution has long dominated the study of multivariate distributions. Multivariate Normal distributions are appealing since the marginal distributions are also Normal, and because the association between random outcomes can be fully described knowing only the marginal distributions and one additional parameter, the correlation coefficient. In practice, however, there are many situations where the normality assumption fails (think, for instance to random variables representing lifetimes or long-tailed claims). Thus alternative models are needed for such data.

An extensive literature in statistics deals with non-normal multivariate distributions; we refer, e.g., to Johnson & Kotz (1972), Johnson, Kotz & Balakrishnan (1997, 2000). However, historically many multivariate distributions have been developed as immediate extensions of univarite distributions, examples being the bivariate Pareto, Gamma and so on. The drawbacks of these types of distributions are that (i) a different family is needed for each marginal distribution, (ii) extensions to more than just the bivariate case are not clear and (iii) measures of association often appears in the marginal distribution.

The construction of multivariate distributions based on copulas does not suffer from these drawbacks. With copula construction offered by Sklar's theorem (1959), we select different marginals for each outcome. For instance, if we deal with bivariate outcome associated with the loss and ALAE's, we could use a Log-Normal distribution for expenses and a longer tail distribution, such as Pareto, for the losses. Then, it suffices to plug these marginals into a suitable copula to get the bivariate distribution. The copula construction does not constrain the choice of the marginal distributions.

The copula modelling turns out to be very useful for the analysis of dependence in actuarial science. Applications of copulas to insurance data modelling have been proposed e.g. by Frees, Carrière & Valdez (1996), Frees & Valdez (1998), Klugman & Parsa (1999), Carrière (2000), Valdez (2001) and Embrechts et al. (2002).

A lot of recent research has focused on a subclass of copulas called the archimedean copula class, which indexes the copula by a univariate function (called the generator) and therefore yields more tractable analytical properties. Many well-known systems of bivariate distributions belong to the archimedean class. Frailty models also fall under that general prescription. As illustrated by Genest & Mc Kay (1986a,b), this class of copulas is wide and analytically tractable and its elements have stochastic properties that make them attractive for the statistical treatment of data.

Because copulas characterize the dependence structure of random vectors once the effect of the marginals has been factored out, identifying and fitting a copula to data is not an easy task. In practice, it is often preferable to restrict the search of an appropriate copula to some reasonable family, like the archimedean one. Then, it is extremely useful to have simple graphical procedures to select the best fitting model among some competing alternatives for the data at hand.

Consider a bivariate outcome (X, Y) with continuous marginals F_X and F_Y, and joint distribution function, F, which can be written using a copula representation, as

$$F(x, y) = C\left(F_X(x), F_Y(y)\right)$$

where C is the unique dependence structure (the copula) associated to F (Sklar (1959)). Suppose that of interest is the estimation of C. As described by Genest & Rivest (1993, 2001), the tool of the estimation procedure is a distribution function that can be constructed for any copula and independently of the marginals. This function is the distribution of the bivariate probability integral transformation (BIPIT) of (X, Y), $Z = F(X, Y)$. Let us denote by K this distribution function, i.e

$$K(z) = \Pr[F(X, Y) \leq z] = \mathbb{E}[\mathbb{I}\{F(X, Y) \leq z\}].$$

In contrast to the univariate case, it is not generally true that the distribution function K is uniform on $[0, 1]$, even when F is continuous. Since the distribution function K contains information about the dependence structure, as described by the associated copula C, the estimation of C is based on the estimation of the distribution function, K.

Starting from the assumption that the archimedean dependence structure is appropriate (an assumption that we will retain throughout our work), Genest & Rivest (1993)

constructed an empirical estimate of K, when both X and Y are completely observable (i.e no censoring, nor truncation) and proposed a graphical procedure for selecting the best archimedean copula which fits a given set of data. Broadly speaking their procedure chooses as the best fitting archimedean model the one whose probability integral transformation distribution, K, is the closest to its empirical estimate.

Remaining in the archimedean framework, Wang & Wells (2000b) extended the idea of Genest & Rivest (1993) to right-censored bivariate failure-time data. This kind of censorship is not the one encountered in actuarial problems but, as pointed out by Wang & Wells (2000b), because the censoring issue is handled in the stage of estimating the bivariate distribution function, F, the approach they propose is flexible enough to deal with other censoring mechanisms. The authors then suggested a selection procedure for the best archimedean model based on a goodness-of-fit statistic depending on the nonparametric estimator of the distribution K. The latter one is obtained from the last given expression, by plugging in an appropriate nonparametric estimate for the bivariate distribution, F.

As mentioned in the beginning, in this part we focus on a particular dataset, the loss-ALAE. This data has been examined in parametric settings by Frees & Valdez (1998) (Pareto marginals and Gumbel copula) and Klugman & Parsa (1999) (inverse paralogistic for loss, inverse Burr for ALAE and Frank copula). In Frees & Valdez (1998), techniques developed by Genest & Rivest (1993) for complete data have been applied to loss-ALAE data in order to select the appropriate generator. As pointed out by Frees & Valdez (1998) in their Section 4.2.1, censoring in the loss variable is ignored in the identification process.

The procedure we propose is based on an appropriate nonparametric estimator of the joint distribution of loss-ALAE taking into account the particular censorship present in the data, thus correcting the procedure suggested by Frees & Valdez (1998). We follow the general approach described in Wang & Wells (2000b), but instead of using Dabrowska (1988) estimator for the bivariate distribution, we use the estimator proposed in Akritas (1994), since only the loss variable is subject to censoring. This estimator is an average, over the uncensored variable, of estimates of the conditional distribution function of the censored variable given the uncensored variable. The estimates of the conditional distribution function used are nearest neighbor estimators. Properties of the proposed estimator for the bivariate distribution, such as asymptotic optimality and weak convergence, were

proven in Akritas (1994)

It is of interest to point out that we prove the applicability of Akritas's estimator developed for random right censoring, to loss-ALAE data which is subject to a generalised type I-censoring (i.e. the censoring variable is constant and differs from each individual to another).

As Wang & Wells (2000b), we then estimate the distribution function K and use it in the selection procedure of the best parametric archimedean copula fitting the data.

Plan of the thesis

All the previously mentioned results are gathered in different joint papers, divided in two categories, those devoted to frequency credibility models and to those to copula modellings, as follows:

Part I: Credibility models for claim frequencies

Static random effects

- Purcaru, O., & Denuit, M. (2002), On the Dependence induced by Frequency Credibility Models, *Belgian Actuarial Bulletin*, 2, 1, 73-79.
- Purcaru, O., & Denuit, M. (2002), On the Stochastic Increasingness of Future Claims in the Bühlmann Linear Credibility Premium, *German Actuarial Bulletin*, 25, 4, 781-793.

Dynamic random effects

- Purcaru, O., & Denuit, M. (2003), Dependence in Dynamic Claim Frequency Credibility Models, *ASTIN Bulletin*, 33, 1, 23-40.
- Purcaru, O., Guillén, M. & Denuit, M. (2004), Linear Credibility Models Based on Time Series for Claim Counts, *Belgian Actuarial Bulletin*, 4, 1, 62-74.

Part II: Copula modelling

- Denuit, M., Purcaru, O. & Van Keilegom, I.(2004), Bivariate archimedean copula modelling for loss-ALAE data in non-life insurance, *IS Discussion Papers* **0423**, Institute de statistique, Université catholique de Louvain.

Part I

Credibility models for claim frequencies

1 On the dependence induced by frequency credibility models

Part of the joint research with M. Denuit, published in *Belgian Actuarial Bulletin*, 2, 1, 73-79, (2002).

1. Introduction and Motivation

One of the main tasks of the actuary is to design a tariff structure that will fairly distribute the burden of claims among policyholders. If the risks in the portfolio are not all equal to each other (which means that the associated random financial losses have different distribution functions), it is fair to partition all policies into homogeneous classes with all policyholders belonging to the same class paying the same premium. In third party liability insurance, the classification variables introduced to partition risks commonly include the age, gender and occupation of the policyholders, the type and use of their car, the place where they reside and sometimes even the number of cars in the household, marital status, smoking behavior or the color of the vehicle. It is convenient to achieve *a priori* classification with the help of generalized linear models; see e.g. Renshaw (1994) for applications in actuarial sciences, and McCullagh and Nelder (1989) or Dobson (1990) for a general overview of the statistical theory.

However, many important factors cannot be taken into account at this stage; think for instance of swiftness of reflexes or aggressiveness behind the wheel. Consequently, tariff cells are still quite heterogeneous despite of the use of many *a priori* variables. These hidden features are usually impossible to measure and to incorporate in a price list. But

it is reasonable to believe that these characteristics are revealed by the number of claims reported by the policyholders over the successive insurance periods. Hence the adjustment of the premium based on the individual claims experience in order to restore fairness among policyholders. The allowance for the history of the policyholder in a rating model thus derives from interpretation of serial correlation for longitudinal data resulting from hidden features in the risk distribution.

In seminal papers, Dionne and Vanasse (1989, 1992) proposed a credibility model which integrates *a priori* and *a posteriori* information on an individual basis. These authors introduced a regression component in the Poisson counting model in order to use all available information in the estimation of accident frequency. The unexplained heterogeneity was then modeled by the introduction of a latent variable representing the influence of hidden policy characteristics. Taking this random effect Gamma distributed yields the Negative Binomial model for the claim number. Of course, there is no particular reason to restrict ourselves to Gamma distributed random effects (except perhaps mathematical convenience). In biostatistical circles, LogNormally distributed random effects are widely used (see also Pinquet (1997) for an application in actuarial science). The use of inverse gaussian distribution has been advocated by Willmot (1987) in conjunction with Poisson mixtures; see also Tremblay (1992). Another possible choice is Hoffman's distribution; see e.g. Kestemont and Paris (1985). Moreover, semi-parametric approaches (retaining the Poisson assumption for the claim numbers without specifying any distribution for the random effects) are conceivable in the spirit of the Bühlmann-Straub model. See e.g. Walhin & Paris (1999) as well as Denuit & Lambert (2001) for nonparametric maximum likelihood methods.

Once estimated, the heterogenous model can be used to perform prediction on longitudinal data and allows experience rating in casualty insurance. In an empirical Bayesian setting, the prediction is derived from the expectation of a random effect with respect to a posterior distribution taking into account the history of the individual. An excellent introduction to these concepts can be found in Pinquet (2000).

In this context, the present paper aims to examine the kind of dependence induced among annual claim numbers by the introduction of random effects taking unexplained heterogeneity into account. We will see that this dependence is one of the strongest possible, because of the total positivity of the Poisson kernel. We will also make precise the effect

of reporting claims on the *a posteriori* distribution of the random effect. This will be done by establishing some stochastic monotonicity property of the *a posteriori* distribution with respect to the claims history.

The main interest of this paper is to formalize intuitive ideas with the help of stochastic orderings. Every actuary intuitively feels that the *a posteriori* claim frequency distribution must become more dangerous as claims are reported. We make here precise the meaning of "more dangerous" and we prove that the *a posteriori* premium must increase with the total claim number in the mixed Poisson model.

2. Poisson credibility models incorporating *a priori* classification

During the observation period, n policies were in portfolio, each one observed during ν_i periods. Let N_{ij} be the number of claims reported by policyholder i during the jth period of insurance, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, \nu_i$. Let d_{ij} be the length of this period. Usually, $d_{ij} = 1$, but there are a variety of situations where this is not the case. Indeed, a new period of observation starts as soon as some policy characteristics are modified (think for instance to a moving of the policyholder for a company using postcode as rating factor, policyholder's wedding for a company using marital status, or simply the policyholder buying a new car).

We thus typically face a nested structure: each policyholder generates a sequence $\boldsymbol{N}_i = \{N_{i1}, N_{i2}, \ldots, N_{i\nu_i}\}$ of claim numbers. It is reasonable to assume independence between the series $\boldsymbol{N}_1, \boldsymbol{N}_2, \ldots, \boldsymbol{N}_n$ (at least in third party liability automobile insurance, for instance), but this assumption is very questionable inside the $\boldsymbol{N}_i$'s (in fact, if the components of the $\boldsymbol{N}_i$'s were independent, *a posteriori* ratemaking would be senseless from the purely actuarial point of view, even if these systems remain commercially important because they counteract moral hazard).

Let

$$N_{i\bullet} = \sum_{j=1}^{\nu_i} N_{ij}$$

be the total claim number reported by policyholder i during the ν_i observation periods; the statistic $N_{i\bullet}$ is a convenient summary of past claims history. So, the company has $\sum_{i=1}^{n} \sum_{j=1}^{\nu_i} d_{ij}$ policyholders/year to build its *a priori* ratemaking scheme. It is customize

to assume that N_{ij} is Poisson distributed. Indeed, Poisson distribution is the "law of small numbers" and is well suited for rare events like automobile accidents.

The idea now is to incorporate in the N_{ij}'s exogenous information summarized in the vectors $\boldsymbol{x}_{ij}$. Specifically, $\boldsymbol{x}_{ij}$ contains all the information included in the price list about policyolder i in period j (like age, sex, power of the car, and so on). A linear model for the logarithm of the claim rates is often used in actuarial science. This provides a regression model for count data analogous to the usual normal regression for continuous data (when the counts are small, which is typically the case in automobile insurance, the normal approximation is poor and fails to account for the discreteness of the data). According to standard methodology of generalized linear models, the logarithmic function is also the natural link for the Poisson distribution. Specifically, the retained specification is

$$N_{ij} =_d \text{Poisson}(\lambda_{ij}) \text{ where } \lambda_{ij} = d_{ij} \exp(\boldsymbol{\beta}^t \boldsymbol{x}_{ij}).$$

Of course, inside each risk class, the policies are not identical *stricto sensu*. In order to achieve *a posteriori* ratemaking we recognize the residual heterogeneity of the portfolio by saying that the premium for each risk class is itself an average. Every policy is thus affected by a risk parameter which can be interpreted as an error term in the regression model. For policyholder i, the risk parameter Θ_i represents the influence of hidden features (as annual mileage or aggressiveness behind the wheel for instance), i.e. unknown risk characteristics of the policyholder having a significant impact on the occurrence of claims.

The ith policy of the portfolio, $i = 1, 2, \ldots, n$, is represented by a sequence $(\Theta_i, \boldsymbol{N}_i)$ where Θ_i is a positive random variable with unit mean representing the unexplained heterogeneity. Specifically, the credibility model is based on the following assumptions:

A1 given $\Theta_i = \theta$, the random variables N_{ij}, $j = 1, 2, \ldots$, are independent and conform to the Poisson distribution with mean $\lambda_{ij}\theta$, i.e.

$$\mathbb{P}[N_{ij} = k | \Theta_i = \theta] = \exp(-\lambda_{ij}\theta) \frac{(\lambda_{ij}\theta)^k}{k!},$$

for $k \in \mathbb{N}, j = 1, 2, \ldots$;

A2 at the portfolio level, the sequences $(\Theta_i, \boldsymbol{N}_i)$ are assumed to be independent.

It is essential to understand the philosophy of this classical actuarial construction. Here dependence between annual claim numbers is a consequence of the heterogeneity of the portfolio; the dependence is only apparent. If we had a complete knowledge of policy characteristics then Θ_i would become deterministic and there would be no more dependence between the N_{ij}'s for fixed i. The unexplained heterogeneity (which has been modeled through the introduction of the risk parameter Θ_i for policyholder i) is then revealed by the claims and premiums histories in a Bayesian way. These histories modify the distribution of Θ_i and hence modify the premium.

When Θ_i conforms to a Gamma prior distribution (with mean 1 and variance $1/a$) it is well-known that

$$[\Theta_i|N_{i1}, N_{i2}, \ldots, N_{i\nu_i}] =_d \text{Gamma}\big(a + N_{i\bullet}, a + \lambda_{i\bullet}\big)$$

so that, given the past premiums $\lambda_{i1}, \lambda_{i2}, \ldots, \lambda_{i\nu_i}$, the *a posteriori* distribution of Θ_i increases in the past claims in the likelihood ratio sense (see the next section for the precise definition of this stochastic order relation). This is clearly a very nice property since it expresses the increasing dangerousness inherent to policyholders reporting claims. We would like to investigate whether this important property still holds when other distributions are taken for Θ_i. Our main finding in that direction is that it remains valid in any Poisson mixture model.

Another nice feature of the Poisson-Gamma model is that the theoretical bonus-malus coefficient is given by

$$\mathbb{E}[\Theta_i|N_{i1}, N_{i2}, \ldots, N_{i\nu_i}] = \frac{a + N_{i\bullet}}{a + \lambda_{i\bullet}}$$

which clearly increases in the past claims $N_{i\bullet}$. Again, we show that this holds true whatever the mixture distribution selected by the actuary.

Finally, we could wonder what kind of dependence is induced by the common mixture model A1. In that respect, we show that the dependence existing between the annual claim numbers is very strong (namely, multivariate positive likelihood ratio dependence). This sheds a new light on many properties for the sequence of annual claim frequencies.

In the model A1-A2, we intuitively feel that the following statements are true:

S1 Θ_i "increases" in the past claims $N_{i\bullet}$

S2 N_{i,ν_i+1} "increases" in the past claims $N_{i\bullet}$

S3 N_{i,ν_i+1} and $N_{i\bullet}$ are "positively dependent".

The next section aims to precise the meaning of "increases" in S1 and S2, as well as the nature of the "positive dependence" involved in S3.

3. Statements S1-S3 in the model A1-A2

3.1 Stochastic order relations

In order to formalize the increasingness involved in S1-S2, our study will extensively resort to stochastic orderings. Therefore, we recall in this section the definition of the orderings useful for the analysis of Poisson mixtures.

This section only gives the definitions of the stochastic orderings we will use, as well as some intuitive intepretations. For more details about stochastic orderings, we refer the reader e.g. to Kaas, Van Heerwaarden and Goovaerts (1994) or to Shaked and Shanthikumar (1994).

Let us first recall the definition of the stochastic dominance.

Definition 3.1. Given two random variables X and Y, X is said to be smaller than Y in the stochastic dominance, written as $X \preceq_{st} Y$, if

$$\Pr[X > t] \leq \Pr[Y > t], \quad \text{for all } t \in \mathbb{R}.$$

From Definition 3.1, we see that a ranking in the $\preceq_{st}$ sense translates in probability models the intuitive meaning of "being smaller than": indeed, we compare the probability that both random variables exceed some given threshold t and the smallest one in the $\preceq_{st}$-sense has the smallest probability of exceeding the treshold. Since any non-decreasing function can be obtained as the uniform limit of convex combinations of non-decreasing step functions, $X \preceq_{st} Y \Rightarrow \mathbb{E}[u(X)] \leq \mathbb{E}[u(Y)]$ for all the non-decreasing functions u. In the framework of von Neumann-Morgenstern expected utility theory, $\preceq_{st}$ thus expresses the common preferences of all the profit-seeking decision-makers.

If X and Y are two continuous random variables with respective probability density functions f_X and f_Y, then

$$X \preceq_{st} Y \quad \Leftrightarrow \quad \int_t^{+\infty} f_X(x)dx \leq \int_t^{+\infty} f_Y(y)dy, \quad \forall t \in \mathbb{R}.$$

If M and N are two counting random variables then

$$M \preceq_{st} N \quad \Leftrightarrow \quad \sum_{j=k}^{+\infty} \Pr[M = j] \leq \sum_{j=k}^{+\infty} \Pr[N = j], \quad \forall k \in \mathbb{N}.$$

In parametric models, likelihood ratio order is also very useful (in particular within the exponential family of distributions).

Definition 3.2. Given two random variables X and Y, X is said to be smaller than Y in the likelihood ratio order, written as $X \preceq_{lr} Y$, if

$$\Pr[X \in A] \Pr[Y \in B] \geq \Pr[X \in B] \Pr[Y \in A],$$

for all $A \leq B \subseteq \mathbb{R}$, where $A \leq B$ means that for any $u \in A$ and $v \in B$, $u \leq v$ holds.

Given two subsets A and B of the real line such that A entirely lies on the left of B, $X \preceq_{lr} Y$ means that considering the random vector (X, Y) with independent components, it is more likely that it assumes a value in $A \times B$ than in $B \times A$. In words, this means that Y is more likely to assume the largest values and hence is "larger" than X.

If X and Y are continuous with respective probability density functions f_X and f_Y then

$$X \preceq_{lr} Y \quad \Leftrightarrow \quad \frac{f_X(t)}{f_Y(t)} \quad \text{decreases over} \quad supp(X) \cup supp(Y)$$

where $a/0$ is taken to be equal to $+\infty$ whenever $a > 0$

$$\Leftrightarrow \quad f_X(u) f_Y(v) \geq f_X(v) f_Y(u) \quad \forall u \leq v \in \mathbb{R} \tag{3.1}$$

If M and N are counting random variables, then

$$M \preceq_{lr} N \quad \Leftrightarrow \quad \frac{\Pr[M = j]}{\Pr[N = j]} \quad \text{decreases over} \quad supp(M) \cup supp(N)$$

$$\Leftrightarrow \quad \Pr[M = j] \Pr[N = k] \geq \Pr[M = k] \Pr[N = j]$$
$$\forall j \leq k \in \mathbb{N}. \tag{3.2}$$

The likelihood ratio order is stronger than the stochastic dominance, that is $X \preceq_{lr} Y \Rightarrow X \preceq_{st} Y$. To check the latter assertion, let us consider two rv's X and Y such that $X \preceq_{lr} Y$. Inserting $B = (t, +\infty)$ and $A = (-\infty, t]$ in Definition 3.2 yields

$$\begin{aligned} & \Pr[X \leq t] \Pr[Y > t] \geq \Pr[X > t] \Pr[Y \leq t] \\ \Leftrightarrow \quad & \big(1 - \Pr[X > t]\big) \Pr[Y > t] \geq \Pr[X > t]\big(1 - \Pr[Y > t]\big) \\ \Leftrightarrow \quad & \Pr[Y > t] \geq \Pr[X > t]. \end{aligned}$$

Since the reasoning is valid for any $t \in \mathbb{R}$, we conclude that $X \preceq_{st} Y$ holds.

Now, the following property will be extremely useful in the remainder of our work.

Property 3.3. *Let N_θ obey to the Poisson distribution with mean θ. Then,*

$$\theta \leq \theta' \Rightarrow N_\theta \preceq_{lr} N_{\theta'}.$$

Proof. It suffices to write

$$\frac{\Pr[N_\theta = j]}{\Pr[N_{\theta'} = j]} = \exp(\theta' - \theta) \left(\frac{\theta}{\theta'}\right)^j$$

which clearly decreases over $\mathbb{N}$. The result then follows from (3.2). □

Let $\mathcal{S}$ and $\mathcal{T}$ be two subsets of the real line $\mathbb{R}$. A function $f : \mathcal{S} \times \mathcal{T} \to \mathbb{R}$ is said to be totally positive of order 2 (TP_2, in short) if the inequality

$$f(s_1, t_1) f(s_2, t_2) \geq f(s_1, t_2) f(s_1, t_2) \tag{3.3}$$

holds true for any $s_1 \leq s_2 \in \mathcal{S}$ and $t_1 \leq t_2 \in \mathcal{T}$.

At this stage, it is worth mentioning that Property 3.3 means that the function $f : \mathbb{N} \times \mathbb{R}^+ \to [0, 1]; (k, \theta) \mapsto \Pr[N_\theta = k]$ is TP_2. This simple fact will be extremely useful in the remainder of the work, in conjunction with the result recalled hereafter.

A fundamental property of TP_2 known as the basic composition formula from Karlin (1968) is as follows. Let $\mathcal{S}$, $\mathcal{T}$ and $\mathcal{U}$ be three subsets of the real line. Given some functions $f : \mathcal{S} \times \mathcal{T} \to \mathbb{R}$ and $g : \mathcal{T} \times \mathcal{U} \to \mathbb{R}$, let us define the function h as

$$h(s, u) = \int_{t \in \mathcal{T}} f(s, t) g(t, u) d\sigma(t)$$

where the integral is assumed to converge absolutely and $\sigma(\cdot)$ denotes a sigma-finite measure on $\mathcal{T}$. It can then be shown that if f is TP_2 on $\mathcal{S}\times\mathcal{T}$ and g is TP_2 on $\mathcal{T}\times\mathcal{U}$ then h is TP_2 on $\mathcal{S}\times\mathcal{U}$.

3.2 Statements S1-S2

Let f_Θ be the probability density function of Θ. We then have the following result, inspired from Whitt (1979, Theorem 4) which formalizes statements S1 and S2: the increasingness mentioned there is with respect to $\preceq_{lr}$.

Proposition 3.4. *(i)* $[\Theta_i|N_{i\bullet}=n]\preceq_{lr}[\Theta_i|N_{i\bullet}=n']$ *for* $n\leq n'$*;*

(ii) $[N_{i,\nu_i+1}|N_{i\bullet}=n]\preceq_{lr}[N_{i,\nu_i+1}|N_{i\bullet}=n']$ *for* $n\leq n'$.

Proof. (i) Denote as $f_\Theta(\cdot|n)$ the pdf of $[\Theta_i|N_{i\bullet}=n]$, $n\in\mathbb{N}$. In virtue of (3.1), we have to show that for any $\theta\leq\theta'$ and $n\leq n'$, the inequality

$$\frac{f_\Theta(\theta|n')}{f_\Theta(\theta|n)}\leq\frac{f_\Theta(\theta'|n')}{f_\Theta(\theta'|n)}\Leftrightarrow\frac{f_\Theta(\theta'|n)}{f_\Theta(\theta|n)}\leq\frac{f_\Theta(\theta'|n')}{f_\Theta(\theta|n')}$$

holds true. This result follows from

$$\begin{aligned}\frac{f_\Theta(\theta'|n)}{f_\Theta(\theta|n)} &= \frac{\Pr[N_{i\bullet}=n|\Theta_i=\theta']}{\Pr[N_{i\bullet}=n|\Theta_i=\theta]}\times\frac{f_\Theta(\theta')}{f_\Theta(\theta)}\\ &\leq \frac{\Pr[N_{i\bullet}=n'|\Theta_i=\theta']}{\Pr[N_{i\bullet}=n'|\Theta_i=\theta]}\times\frac{f_\Theta(\theta')}{f_\Theta(\theta)}\\ &= \frac{f_\Theta(\theta'|n')}{f_\Theta(\theta|n')}\end{aligned}$$

where the inequality above follows from Property 3.3.

(ii) In virtue of (3.2), we have to show that for any $k\leq k'$

$$\begin{aligned}&\Pr[N_{i,\nu_i+1}=k|N_{i\bullet}=n]\Pr[N_{i,\nu_i+1}=k'|N_{i\bullet}=n']\\ &\geq\Pr[N_{i,\nu_i+1}=k|N_{i\bullet}=n']\Pr[N_{i,\nu_i+1}=k'|N_{i\bullet}=n]\end{aligned}$$

which is equivalent to

$$\Pr[N_{i,\nu_i+1}=k,N_{i\bullet}=n]\Pr[N_{i,\nu_i+1}=k',N_{i\bullet}=n']$$

$$\geq \Pr[N_{i,\nu_i+1} = k, N_{i\bullet} = n'] \Pr[N_{i,\nu_i+1} = k', N_{i\bullet} = n].$$

From assumption A1, we have that

$$\Pr[N_{i,\nu_i+1} = k, N_{i\bullet} = n]$$

$$= \int_{\theta\in\mathbb{R}^+} \Pr[N_{i,\nu_i+1} = k, N_{i\bullet} = n|\Theta_i = \theta] f_\Theta(\theta) d\theta$$

$$= \int_{\theta\in\mathbb{R}^+} \Pr[N_{i,\nu_i+1} = k|\Theta_i = \theta] \Pr[N_{i\bullet} = n|\Theta_i = \theta] f_\Theta(\theta) d\theta.$$

Now, $[N_{i\bullet} = n|\Theta_i = \theta]$ and $[N_{i,\nu_i+1} = k|\Theta_i = \theta]$ both conform to the Poisson distribution, with respective means $\lambda_{i\bullet}\theta$ and $\lambda_{i,\nu_i+1}\theta$. Therefore, Property 3.3 applies and ensures that the functions $(n, \theta) \mapsto \Pr[N_{i\bullet} = n|\Theta_i = \theta]$ and $(k, \theta) \mapsto \Pr[N_{i,\nu_i+1} = k|\Theta_i = \theta]$ are both TP_2. Hence, invoking Karlin's basic composition formula, $(k, n) \mapsto \Pr[N_{i,\nu_i+1} = k, N_{i\bullet} = n]$ is also TP_2, from which follows the conclusion. □

3.3 Positive dependence notions for random couples

In order to formalize the positive dependence involved in statement S3, we will present several concepts of dependence. The study of concepts of positive dependence for random variables, started in the late 1960's, has yielded numerous useful results in both statistical theory and applications. Applications of these concepts in actuarial science recently received increased interest.

In this section, we recall the definitions of several useful dependence concepts for random couples. In each case, the aim is to formalize the positive dependence existing between the two components of the random couple (i.e. the fact that large values of one component tend to be associated with large values for the other).

Definition 3.5. Let $\boldsymbol{X} = (X_1, X_2)$ be a bivariate random vector.

(i) $\boldsymbol{X}$ is Positive Quadrant Dependent (PQD in short) if

$$\Pr[\boldsymbol{X} > \boldsymbol{x}] \geq \Pr[X_1 > x_1] \Pr[X_2 > x_2] \quad \forall x_1, x_2 \in \mathbb{R}$$
$$\Leftrightarrow \Pr[\boldsymbol{X} \leq \boldsymbol{x}] \geq \Pr[X_1 \leq x_1] \Pr[X_2 \leq x_2] \quad \forall x_1, x_2 \in \mathbb{R};$$

(ii) $\boldsymbol{X}$ is associated (A, in short) if

$$\mathbb{E}[g(\boldsymbol{X})h(\boldsymbol{X})] \geq \mathbb{E}[g(\boldsymbol{X})]\mathbb{E}[h(\boldsymbol{X})]$$

for all $g, h : \mathbb{R}^2 \to \mathbb{R}$ simultaneously non-decreasing (or non-increasing) functions;

(iii) $\boldsymbol{X}$ is positive regression dependent (PRD, in short) if $[X_2|X_1 = x_1] \preceq_{st} [X_2|X_1 = x_1']$ for all $x_1 \leq x_1'$ and $[X_1|X_2 = x_2] \preceq_{st} [X_1|X_2 = x_2']$ for all $x_2 \leq x_2'$;

(iv) $\boldsymbol{X}$ is said to be positively likelihood ratio dependent (PLRD, in short) if $[X_2|X_1 = x_1] \preceq_{lr} [X_2|X_1 = x_1']$ for all $x_1 \leq x_1'$ and $[X_1|X_2 = x_2] \preceq_{lr} [X_1|X_2 = x_2']$ for all $x_2 \leq x_2'$

(v) $\boldsymbol{X}$ is said to be comonotonic (C, in short) if there exists a random variable Z and non-decreasing functions φ_1 and φ_2 such that $\boldsymbol{X} =_d (\varphi_1(Z), \varphi_2(Z))$.

PQD has been introduced by Lehmann (1966). Its intuitive meaning is clear: $\boldsymbol{X}$ is PQD when the probability for the components X_1 and X_2 of $\boldsymbol{X}$ to be simultaneously large (or small) is at least equal as it would be if they were independent. PQD so expresses a higher clustering of data points in the upper right quadrant and lower left quadrant compared to the theoretical situation where X_1 and X_2 are mutually independent.

Association has been introduced by Esary, Proschan and Walkup (1967) and further studied by Esary and Proschan (1972). It is an analytical condition, which is sometimes easy to establish and provides a sufficient condition for PQD. To the knowledge of the authors, there is no intuitive meaning to the inequality used in (ii) to define A, except that it resorts to covariances, a classical tool used to measure the strength of (linear) dependence.

PRD and PLRD impose stochastic increasingness of one component of the random couple in the value assumed by the other component either in the $\preceq_{st}$- or in the $\preceq_{lr}$-sense. These dependence notions are thus rather intuitive since the value assumed by one of the components increases in the value taken by the other, in a stochastic sense.

Note that the couple $\boldsymbol{X}$ of continuous random variables is PLRD if, and only if, for any $x_1 \leq x_1'$ and $x_2 \leq x_2'$, its bivariate density $f_{\boldsymbol{X}}$ satisfies

$$f_{\boldsymbol{X}}(x_1, x_2) f_{\boldsymbol{X}}(x_1', x_2') \geq f_{\boldsymbol{X}}(x_1, x_2') f_{\boldsymbol{X}}(x_1', x_2)$$

that is, according to (3.3), if $f_{\boldsymbol{X}}$ is TP_2. Again, the latter inequality enjoys an intuitive interpretation in terms of likelihood. Assume the values taken by the random couples $\boldsymbol{X}$ and its independent copy $\boldsymbol{X}'$ are x_1, x_2, x_1' and x_2', then it is more likely that one of the two random couples assume the two largest values, and the other one assumes the two smallest values. This expresses the fact that it is more likely that both components of $\boldsymbol{X}$ are simultaneously large or small.

Two counting random variables M_1 and M_2 are PLRD if, and only if,

$$\Pr[M_1 = k_1, M_2 = k_2] \Pr[M_1 = k_1', M_2 = k_2'] \geq$$

$$\geq \Pr[M_1 = k_1, M_2 = k_2'] \Pr[M_1 = k_1', M_2 = k_2]$$

for any $k_1 \leq k_1'$ and $k_2 \leq k_2'$. Coming back to (3.3), M_1 and M_2 are PLRD if their discrete bivariate probability density function is TP_2.

Finally, X_1 and X_2 are comonotonic if they can be represented as non-decreasing function of some underlying random variable Z. The X_i's are thus "common-monotonic" since they both "move together" (increasing Z makes both X_1 and X_2 larger). Comonotonicity is the strongest possible dependence between two random outcomes: it is sometimes referred to as perfect positive dependence since comonotonic random variables X_1 and X_2 are in fact functionally related. This notion allows for many nice applications in actuarial science, as it can be seen from Kaas, Dhaene & Goovaerts (2000) and the references contained in that paper.

Let us now detail the implications between these concepts of dependence (for a proof, see e.g. Lehmann (1955, 1966) and Esary and Proschan (1972)). All the implications are strict:

$$\boldsymbol{X} \text{ C} \Rightarrow \boldsymbol{X} \text{ PLRD} \Rightarrow \boldsymbol{X} \text{ PRD} \Rightarrow \boldsymbol{X} \text{ A} \Rightarrow \boldsymbol{X} \text{ PQD}.$$

3.4 Statement S3

Let us now prove that the total claim number $N_{i\bullet}$ reported in the past periods and the claim frequency N_{i,ν_i+1} for the next coverage period are PLRD. This formalizes statement S3 and is inspired from Fahmy et al. (1982).

Proposition 3.6. *$N_{i\bullet}$ and N_{i,ν_i+1} are PLRD.*

Proof. We have to establish that the function $(n,k) \mapsto \Pr[N_{i\bullet} = n, N_{i,\nu_i+1} = k]$ is TP_2. This is a simple consequence of the basic composition formula applied to

$$\Pr[N_{i\bullet} = n, N_{i,\nu_i+1} = k]$$

$$= \int_{\theta\in\mathbb{R}^+} \Pr[N_{i\bullet} = n|\Theta_i = \theta] \Pr[N_{i,\nu_i+1} = k|\Theta_i = \theta] dF_\Theta(\theta).$$

□

From Section 3.3, we see that Proposition 3.6 provides a host of useful inequalities since PLRD is one of the strongest dependence concepts. In particular, whatever the distribution of Θ_i, the theoretical bonus-malus coefficient $\mathbb{E}[\Theta_i|N_{i\bullet} = n]$ is increasing in n.

4. Dependence between annual claim numbers

In this section, we examine the dependence existing between the components of $\boldsymbol{N}_i$, i.e. between the N_{it}'s, $t = 1, 2, \ldots, \nu_i$. To this end, we need multivariate extensions of the bivariate dependence notions introduced in Definition (3.5).

4.1 Positive dependence notions for random vectors

Let us now extend the notions introduced in Section 3 to the multivariate case. Henceforth, we restrict ourselves to random vectors valued in the positive orthant or in $\mathbb{N}^n$.

Definition 4.1. Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ be a n-dimensional random vector.

(i) $\boldsymbol{X}$ is Positive Orthant Dependent (POD in short) if

$$\Pr[\boldsymbol{X} > \boldsymbol{x}] \geq \prod_{i=1}^{n} \Pr[X_i > x_i] \text{ for all } x_1, x_2, \ldots, x_n \in \mathbb{R},$$

and

$$\Pr[\boldsymbol{X} \leq \boldsymbol{x}] \geq \prod_{i=1}^{n} \Pr[X_i \leq x_i] \text{ for all } x_1, x_2, \ldots, x_n \in \mathbb{R}$$

simultaneously hold;

(ii) $\boldsymbol{X}$ is associated (A, in short) if

$$\mathbb{E}[g(\boldsymbol{X})h(\boldsymbol{X})] \geq \mathbb{E}[g(\boldsymbol{X})]\mathbb{E}[h(\boldsymbol{X})]$$

for all the non-decreasing functions $g, h : \mathbb{R}^n \to \mathbb{R}$;

(iii) $\boldsymbol{X}$ is conditionally increasing (CI, in short) if

$$[X_i | X_j = x_j,\ \ j \in J] \preceq_{st} [X_i | X_j = x'_j,\ \ j \in J]$$

whenever $x_j \leq x'_j$, $j \in J$, $J \subset \{1, 2, \ldots, n\}$ and $i \notin J$.

(iv) $\boldsymbol{X}$ is conditionally increasing in sequence (CIS, in short) if X_i is stochastically increasing in $X_1, \ldots, X_{i-1}$, for $i \in \{2, \ldots, n\}$ i.e.

$$[X_i | X_1 = x_1, \ldots, X_{i-1} = x_{i-1}] \preceq_{st}$$

$$[X_i | X_1 = x'_1, \ldots, X_{i-1} = x'_{i-1}],$$

whenever $x_j \leq x'_j,\ j \in \{1, \ldots, i-1\}$;

(v) $\boldsymbol{X}$ is said to be Multivariate Positive Likelihood Ratio Dependent (MPLRD, in short) if its multivariate probability density function $f_{\boldsymbol{X}}$ is MTP_2, that is if

$$f_{\boldsymbol{X}}(\boldsymbol{x} \vee \boldsymbol{y}) f_{\boldsymbol{X}}(\boldsymbol{x} \wedge \boldsymbol{y}) \geq f_{\boldsymbol{X}}(\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{y})$$

holds true for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, where the lattice operators $\vee$ and $\wedge$ are defined as

$$\boldsymbol{x} \vee \boldsymbol{y} = \big(\max\{x_1, y_1\}, \ldots, \max\{x_n, y_n\}\big)$$

and

$$\boldsymbol{x} \wedge \boldsymbol{y} = \big(\min\{x_1, y_1\}, \ldots, \min\{x_n, y_n\}\big);$$

(vi) $\boldsymbol{X}$ is said to be comonotonic (C, in short) if there exists a random variable Z and non-decreasing functions $\varphi_1, \varphi_2, \ldots, \varphi_n$ such that $\boldsymbol{X} =_d (\varphi_1(Z), \varphi_2(Z), \ldots, \varphi_n(Z))$.

POD is a straightforward generalization of PQD to higher dimensions, obtained by substituting orthants to quadrants. Note however that we have to impose conditions on both $\Pr[\boldsymbol{X} > \boldsymbol{x}]$ and $\Pr[\boldsymbol{X} \leq \boldsymbol{x}]$ whereas it was an equivalence in the bivariate case. The definition of association naturally carries over higher dimensions. The definition of CI is taken from Müller and Scarsini (2001). It is stronger than the classical CIS notion in the sense that it does not depend on the order of the components of $\boldsymbol{X}$. Nevertheless, in our context, CIS is rather natural since time induces a natural order between the components of $\boldsymbol{N}_i$.

Karlin and Rinott (1980) proved that MPLRD expresses strong positive dependence; sometimes $\boldsymbol{X}$ MPLRD is referred to $\boldsymbol{X}$ MTP_2 since the condition on $f_{\boldsymbol{X}}$ imposed for being MPLRD ensures that $f_{\boldsymbol{X}}$ is MTP_2. Kemperman (1977) proved that TP_2 in pairs and MTP_2 are equivalent, a result which relies on the fact that all the random vectors considered in this work are valued in a sub-lattice of $\mathbb{R}^n$.

We have the following relations between the different concepts of positive dependence described above:

$$\boldsymbol{X} \text{ C} \Rightarrow \boldsymbol{X} \text{ MPLRD} \Rightarrow \boldsymbol{X} \text{ CI} \Rightarrow \boldsymbol{X} \text{ CIS} \Rightarrow \boldsymbol{X} \text{ A} \Rightarrow \boldsymbol{X} \text{ POD}.$$

For a proof of this chain of implications, see e.g. Barlow & Proschan (1975), Joe (1997) together with Müller and Scarsini (2001).

4.2 Serial dependence for claim frequencies

Let us now prove the following result.

Proposition 4.2. *$\boldsymbol{N}_i$ is MPLRD.*

Proof. We know from Kemperman (1977) it suffices to show that $\boldsymbol{k} \mapsto \Pr[\boldsymbol{N}_i = \boldsymbol{k}]$ is TP_2 in each pair of arguments when the others are held fixed. We can write :

$$\begin{aligned}
&\Pr[N_{i1} = k_1, N_{i2} = k_2, \ldots, N_{i\nu_i} = k_{\nu_i}] \\
&= \textstyle\int_{\theta \in \mathbb{R}^+} \Pr[N_{i1} = k_1, \ldots, N_{i\nu_i} = k_{\nu_i} | \Theta_i = \theta] dF_\Theta(\theta) \\
&= \textstyle\int_{\theta \in \mathbb{R}^+} \prod_{j=1}^{\nu_i} \Pr[N_{ij} = k_j | \Theta_i = \theta] dF_\Theta(\theta).
\end{aligned}$$

Let us fix $k_3, k_4, \ldots, k_{\nu_i}$. From Property 3.3 we know that the functions $(k_1, \theta) \mapsto \Pr[N_{i1} = k_1 | \Theta_i = \theta]$ and $(k_2, \theta) \mapsto \Pr[N_{i2} = k_2 | \Theta_i = \theta]$ are both TP_2. The basic composition formula then ensures that $(k_1, k_2) \mapsto \Pr[\boldsymbol{N}_i = \boldsymbol{k}]$ is TP_2 in k_1 and k_2. □

5. Conclusion

The present paper aimed to investigate the kind of dependence generated by actuarial credibility models. To the best of the authors' knowledge, this aspect of actuarial modelling has never been investigated in the literature so far. It turns out that the kind of dependence induced by these models is very strong, namely MPLRD. It is thus not surprising that the *a posteriori* corrections computed on the basis of these models are so severe that they are difficultly implemented in commercial practice.

It is worth mentioning that most of the reasonings only use the fact that the Poisson distribution is monotone in its mean in the $\preceq_{\text{lr}}$-sense (as shown in Property 3.3). So the results are readily extended to any other claim frequency distribution possessing this property. See also Shaked and Spizzichino (1998) for similar results involving absolutely continuous conditional distributions.

Acknowledgements

The support of the Belgian Government under contract "Project d'Action de Recherche Concertées" ARC 98/03-217 is gratefully acknowledged.