I N S T I T U T D E

S T A T I S T I Q U E

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

D I S C U S S I O N

P A P E R

0134

ON THE DEPENDENCE INDUCED BY FREQUENCY CREDIBILITY MODELS

I. Time-invariant random effects

O. PURCARU and M. DENUIT

http://www.stat.ucl.ac.be

ON THE DEPENDENCE INDUCED BY FREQUENCY CREDIBILITY MODELS

I. Time-invariant random effects

OANA PURCARU & MICHEL DENUIT
Institut de Statistique
Université Catholique de Louvain
Voie du Roman Pays, 20
B-1348 Louvain-la-Neuve
Belgium
purcaru@stat.ucl.ac.be & denuit@stat.ucl.ac.be

September 5, 2001

Abstract

In casualty insurance, actuaries usually resort to random effects to take unexplained heterogeneity into account in the spirit of the Bühlmann-Straub model. This paper aims to study the kind of dependence induced by the introduction of common latent variables in the annual numbers of claims reported by policyholders. It will be seen that this classical construction entails strong positive dependence between the underlying random variables.

Key words and phrases: Poisson mixture, credibility theory, explanatory variables, experience rating, risk classification

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 Mc Cullagh 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 (2000) 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.

$$\mathbb{P}[N_{ij} = k | \Theta_i = \theta] = \exp(-\lambda_{ij}\theta) \frac{(\lambda_{ij}\theta)^k}{k!}, \quad k \in \mathbb{N}, j = 1, 2, \ldots;$$

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 the Gamma 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 is the kind of dependence 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

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

¿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

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

If M and N are two counting random variables then

$$M \preceq_{st} N \Leftrightarrow \sum_{j=k}^{+\infty} \Pr[M=j] \leq \sum_{j=k}^{+\infty} \Pr[N=j] \text{ for all } k \in \mathbb{N}.$$

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

$$\Pr[X \in A]\Pr[Y \in B] \geq \Pr[X \in B]\Pr[Y \in A] \text{ for all } A \leq B \subseteq \mathbb{R},$$

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

$$\begin{aligned} X \preceq_{lr} Y \quad &\Leftrightarrow \quad \frac{f_X(t)}{f_Y(t)} \text{ decreases over the union of the supports of } X \text{ and } Y \\ &\qquad \text{where } a/0 \text{ is taken to be equal to } +\infty \text{ whenever } a>0 \\ &\Leftrightarrow \quad f_X(u)f_Y(v) \geq f_X(v)f_Y(u) \text{ for all } u \leq v \in \mathbb{R} \end{aligned} \tag{3.1}$$

If M and N are counting random variables, then

$$\begin{aligned} M \preceq_{lr} N \quad &\Leftrightarrow \quad \frac{\Pr[M=j]}{\Pr[N=j]} \text{ decreases over the union of the supports of } M \text{ and } N \\ &\Leftrightarrow \quad \Pr[M=j]\Pr[N=k] \geq \Pr[M=k]\Pr[N=j] \text{ for all } j \leq k \in \mathbb{N}. \end{aligned} \tag{3.2}$$

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). $\square$

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 $k \leq k'$

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

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

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

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

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

(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, its bivariate density $f_{\boldsymbol{X}}$ satisfies

$$f_{\boldsymbol{X}}(x_1, x_2) f_{\boldsymbol{X}}(x_1', x_2') \geq f_{\boldsymbol{X}}(x_1, x_2') f_{\boldsymbol{X}}(x_1', x_2) \text{ for any } x_1 \leq x_1' \text{ and } x_2 \leq x_2',$$

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,

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

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. (1989).

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 :

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

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 $\preceq_{\mathrm{lr}}$-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.

In the second part of our study (Purcaru and Denuit (2001)), we consider alternative credibility models, generating weaker dependence than the classical ones. More precisely, we focus on time-dependent random effects. Whereas classical actuarial theory assumes time-invariance of the latent variables, Pinquet, Guillén and Bolancé (2000) have recently generalized the Dionne-Vanasse model by allowing the random effects to vary with time, in the spirit of Gerber and Jones (1975) and Sundt (1988). Specifically, random effects are now described by a stationary process; the dependence between annual claim numbers induced by this model is also examined. Of course, this dependence is weaker (compared to the Dionne-Vanasse model, which appears as an extreme case) and is determined by the kind of dependence expressed by the stochastic process governing the random effects. Further, we also consider copula models to describe serial correlation and study the properties of the associated bonus-malus coefficient.

References

[1] Barlow, R.E., Proschan, F. (1975). *Statistical Theory of Reliability and Life Testing – Probability Models.* Holt, Rinehart and Winston.

[2] Denuit, M., & Lambert, Ph. (2001). Smoothed NPML estimation of the risk distribution underlying Bonus-Malus systems. *Proceedings of the Casualty Actuarial Society*, to appear.

[3] Dionne, G., and Vanasse, C. (1989). A generalization of actuarial automobile insurance rating models: the Negative Binomial distribution with a regression component. *ASTIN Bulletin* **19**, 199-212.

[4] Dionne, G., and Vanasse, C. (1992). Automobile insurance ratemaking in the presence of asymmetrical information. *Journal of Applied Econometrics* **7**, 149-165.

[5] Dobson, A.J. (1990). *An Introduction to Generalized Linear Models.* Chapman & Hall/CRC, London.

[6] Esary, J.D., and Proschan, F. (1972). Relationships among some notions of bivariate dependence. *Annals of Mathematical Statistics* **43**, 651-655.

[7] Esary, J.D.,Proschan, F. and Walkup, D.W. (1967). Association of random variables, with applications. *Annals of Mathematical Statistics* **38**, 1466-1474.

[8] Fahmy, S., Pereira, C.A., Proschan, F., and Shaked, M. (1989). The influence on the sample on the posterior distribution. *Communications in Statistics – Theory & Methods* **11**, 1157-1168.

[9] Gerber, H.U., and Jones, D. (1975). Credibility formulas of the updating type. *Transactions of the Society of Actuaries* **27**, 31-52.

[10] Joe, H. (1997). *Multivariate Models and Dependence Concepts.* Chapman & Hall, London.

[11] Kaas, R., Dhaene, J., & Goovaerts, M.J. (2000). Upper and lower bounds for sums of random variables. *Insurance: Mathematics & Economics* **27**, 151-168 IME

[12] Kaas, R., Van Heerwaarden, A.E. and Goovaerts, M.J. (1994). *Ordering of Actuarial Risks.* CAIRE. Brussels.

[13] Karlin, S. (1968). *Total Positivity.* Stanford University Press, Stanford, California.

[14] Karlin, S., and Rinott, Y. (1980). Classes of Orderings of Measures and Related Correaltion Inequalities. I. Multivariate Totally Positive Distributions. *Journal of Multivariate Analysis* **10**, 467-498.

[15] Kemperman, J.H.B (1977). On the FKG-inequality for measures on a partially ordered space. *Indagationes Mathematicae* **39**, 313-331.

[16] Kestemont, R.-M., and Paris, J. (1985). Sur l'ajustement du nombre de sinistres. *Bulletin of the Swiss Association of Actuaries*, 157-163.

[17] Lehmann, E.L. (1955). Ordered families of distributions. *Annals of Mathematical Statistics* **26**, 399-419.

[18] Lehmann, E.L. (1966). Some concepts of dependence. *Annals of Mathematical Statistics* **37**, 1137-1153.

[19] Mc Cullagh, P., and Nelder, J.A. (1989). *Generalized Linear Models.* Chapman & Hall, New York.

[20] Müller, A., and Scarsini, M. (2001) Stochastic comparison of random vectors with a common copula. *Mathematics of Operations Research*, to appear.

[21] Pinquet, J. (1997). Allowance for cost of claims in Bonus-Malus systems. *ASTIN Bulletin* **27**, 33-57.

[22] Pinquet, J. (2000). Experience rating through heterogeneous models. In *Handbook of Insurance*, edited by G. Dionne. Kluwer Academic Publishers.

[23] Pinquet, J., Guillen, M., & Bolancé, C. (2001). Allowance for the age of claims in bonus-malus systems. *ASTIN Bulletin*, to appear.

[24] Purcaru, O., & Denuit, M. (2001). On the dependence induced by the frequency credibility models. II. Time-evolutionary random effects and copula models. In preparation.

[25] Renshaw, A.E. (1994). Modelling the claim process in the presence of covariates. *ASTIN Bulletin* **24**, 265-285.

[26] Shaked, M., and Shantikumar, J.G. (1994). *Stochastic Orders and Their Applications.* Academic Press, New York.

[27] Shaked, M., and Spizzichino, F. (1998). Positive dependence properties of conditionally independent random lifetimes. *Mathematics of Operations Research* **23**, 944-959.

[28] Sundt, B. (1988). Credibility estimators with geometric weights. *Insurance: Mathematics and Economics* **7**, 113-122.

[29] Tremblay, L. (1992). Using the Poisson-Inverse Gaussian distribution in Bonus-Malus Systems. *ASTIN Bulletin* **22**, 97-106.

[30] Walhin, J.-F., and Paris, J. (1999). Using mixed Poisson distributions in connection with Bonus-Malus systems. *ASTIN Bulletin* **29**, 81-99.

[31] Whitt, W. (1979). A note on the influence of the sample on the posterior distribution. *Journal of the American Statistical Association* **74**, 424-426.

[32] Willmot, G.E. (1987). The Poisson-Inverse Gaussian distribution as an alternative to the Negative Binomial. *Scandinavian Actuarial Journal*, 113-127.