INSTITUT DE STATISTIQUE

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

DISCUSSION PAPER

0203

ON THE DEPENDENCE INDUCED BY FREQUENCY CREDIBILITY MODELS

II. Dynamic random effects

O. PURCARU & M. DENUIT

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

ON THE DEPENDENCE INDUCED BY FREQUENCY CREDIBILITY MODELS

II. Dynamic 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

January 24, 2002

Abstract

In an effort to incorporate the date of claims in risk prediction, PINQUET, GUILLÉN & BOLANCÉ (2001) specified a stationary process for the random effects. This paper aims to investigate the type of dependence induced by such a dynamic heterogeneity. Contrarily to the static case studied in PURCARU & DENUIT (2001), the correlation structure between latent variables now crucially affects the association between claim numbers. In particular, taking comonotonic random effects gives the static credibility model considered by these authors, whereas taking different copulas yields various dependence structures.

Key words and phrases: Poisson mixture, credibility theory, risk classification, experience rating, copulas.

1 Introduction and Motivation

As explained in Purcaru & Denuit (2001), many important factors cannot be taken into account in the *a priori* ratemaking; think for instance of swiftness of reflexes, aggressiveness behind the wheel or knowledge of the highway code. Consequently, tariff cells are still quite heterogeneous despite the use of many classification variables. 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 vast majority of the papers appeared in the actuarial literature considered time-independent (or static) heterogeneous models. The dependence induced by these models has been studied in Purcaru & Denuit (2001). Noticeable exceptions are the pioneering papers by Gerber & Jones (1975), Sundt (1988) and Pinquet, Guillén & Bolancé (2001), among others. The allowance for an unknown underlying random parameter that develops 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). The effort variable determines the moral hazard and is modeled by a dynamic unobserved factor.

The main technical interest of letting the random effects evolve over time 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 of Purcaru & Denuit (2001), 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.

Let us briefly detail the content of this paper. Section 2 describes the frequency credibility model with dynamic random effects. Section 3 proposes four models for residual heterogeneity. Three of them are taken from Pinquet (2000) and Pinquet, Guillén & Bolancé (2001). The last one proposes an autoregressive structure of order one, based on Sklar's copula construction. In Section 4, after having formalized the concepts of "being larger than" and "being positively dependent" for random vectors, we prove that the correlation structure of random effects is transmitted to claim frequencies. Each of the four heterogeneity models of Section 3 is then carefully examined in this respect. In Section 5, we aim to formalize the idea that policyholders having worse claim records become worse on the random effects modelling unobserved heterogeneity, meaning that they are more dangerous for the insurer. Section 6 focusses on predictive distribution. Specifically, we establish a stochastic monotonicity property of the future random effect given past experience. The final Section 7 concludes.

To end with this introduction, let us say a few words about the notation used throughout this paper. Henceforth, bold symbols are used for multivariate quantities. Further, $\mathcal{N}(\mu, \sigma^2)$ denotes the Normal law with mean μ and variance σ^2, $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the multivariate Normal law with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, $\mathcal{LN}(\mu, \sigma^2)$ stands for the Log-Normal distribution with parameters μ and σ^2, and finally $\mathcal{G}(a, \tau)$ is the Gamma law with mean a/τ and variance a/τ^2. All the operations involving vectors have to be interpreted componentwise.

2 Poisson credibility model with dynamic heterogeneity

At the beginning of each insurance period, the actuary has at his disposal some information about each policyholder, summarized in a vector $\boldsymbol{x}$. Resorting to standard Poisson regression machinery, this information is integrated into the prediction of the annual claim frequency λ. A random effect is superposed to the prediction λ to recognize the residual heterogeneity of the portfolio.

In the classical credibility construction studied in Purcaru & Denuit (2001), the risk parameter relating to each policyholder has been assumed constant over time. This is of course rather unrealistic since driving ability may vary during the driving career (because of the learning effect, or modification in the risk characteristics). In this paper, we will assume that the unknown characteristics relating to policyholder i during year t are represented by a random variable Θ_{it}. The annual numbers of claims $N_{i1}, N_{i2}, N_{i3}, \ldots$ are assumed to be independent given the sequence $\boldsymbol{\Theta}_i = (\Theta_{i1}, \Theta_{i2}, \Theta_{i3}, \ldots)$ of random effects. The latent unobservable process $\boldsymbol{\Theta}_i$ characterizes the correlation structure of the N_{it}'s.

Now, the ith policy of the portfolio, $i = 1, 2, \ldots, n$, is supposed to be observed during ν_i period; it is represented by a double sequence $(\boldsymbol{\Theta}_i, \boldsymbol{N}_i)$ where $\boldsymbol{\Theta}_i$ is a positive random vector with unit mean representing the unexplained heterogeneity. Specifically, the model is based on the following assumptions:

B1 given $\boldsymbol{\Theta}_i = \boldsymbol{\theta}_i$, the random variables N_{it}, $t = 1, 2, \ldots, \nu_i$, are independent and conform to the Poisson distribution with mean $\lambda_{it}\theta_{it}$, i.e.

$$\Pr[N_{it} = k | \Theta_{it} = \theta_{it}] = \exp(-\lambda_{it}\theta_{it}) \frac{(\lambda_{it}\theta_{it})^k}{k!}, \ \ k \in \mathbb{N};$$

B2 at the portfolio level, the $\boldsymbol{\Theta}_i$'s are assumed to be independent and identically distributed; further, 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 random variables with unit mean ($\mathbb{E}[\Theta_{it}] = 1$ for all i, t, which means that the a priori ratemaking is correct on average):

- the distribution of $\boldsymbol{\Theta}_i = (\Theta_{i1}, \Theta_{i2}, \ldots, \Theta_{i\nu_i})$ depends only on ν_i;
- defining $\nu_{\max} = \max_i \nu_i$, $\boldsymbol{\Theta}_i =_d (\Theta_1, \ldots, \Theta_{\nu_i})$, the distribution of $\boldsymbol{\Theta} = (\Theta_1, \ldots, \Theta_{\nu_{\max}})$ is supposed to be stationary.

Henceforth, we denote by $G(\cdot)$ (resp. $g(\cdot)$) the common cumulative distribution function (resp. probability density function) of the Θ_{it}'s, for $t = 1, \ldots, \nu_i$. We suppose also that the squared random effects are integrable.

A fundamental difference between the static model A1-A2 of Purcaru & Denuit (2001) and the dynamic B1-B2 credibility model described above is that the latter incorporates the age of the claims in the risk prediction, whereas the former neglects this information. Since we intuitively feel that the predictive ability of a claim should decrease with its age, dynamic specification seems more acceptable. As pointed out by Pinquet, Guillén & Bolancé (2001), dynamic credibility models agree with economic analysis of multiperiod optimal insurance under moral hazard. In this optic, the stationarity of the $\boldsymbol{\Theta}_i$'s implies that the predictive ability of claims depends solely on the lag between the date of prediction and the date of occurrence (because of time translation invariance of the marginals of the $\boldsymbol{\Theta}_i$'s).

In the model B1-B2, we intuitively feel that the following statements are true: provided that the Θ_{it}'s are "positively dependent",

T1 the N_{it}'s are "positively dependent"

T2 $\boldsymbol{\Theta}_i$ "increases" in the claims $\boldsymbol{N}_i$

T3 N_{i,ν_i+1} "increases" in the past claims $\boldsymbol{N}_i$

T4 Θ_{i,ν_i+1} "increases" in the past claims $\boldsymbol{N}_i$

The purpose of the remainder of this paper will be to formalize these ideas, and to make precise the concepts of positive dependence and increasingness involved in T1-T4.

3 Modelling heterogeneity

Let us now specify some distributions matching the constraints enumerated in B2. Models B3-B5 come from Pinquet et al. (2001) and Pinquet (2000). Model B6 extends these specifications using copulas.

3.1 Model B3

Let $\boldsymbol{W} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ where $\sigma_{st} = \sigma_W^2 \rho_W(|s-t|)$ with $|\rho_W(h)| \leq 1$, $\rho_W(0) = 1$. The correlation function ρ_W plays a central role in the analysis of the predictive ability of past claims, as shown in Pinquet et al. (2001). Now, define

$$\Theta_t = \frac{\exp W_t}{\mathbb{E}[\exp W_t]} \sim \mathcal{LN}(-\frac{\sigma_W^2}{2}, \sigma_W^2), \;\; t = 1, \ldots, \nu_{\max}.$$

Hence, all the $\boldsymbol{\Theta}_i$'s conform to multivariate LogNormal distributions.

Note that if $\rho_W(h) = 1$ for all h, we have in fact $\boldsymbol{\Theta}_i = (\Theta_i, \ldots, \Theta_i)$ with each Θ_i conforming to the LogNormal distribution; we thus find a special case of the model examined in Purcaru & Denuit (2001).

3.2 Model B4

We could further specify model B3 by assuming in addition that $\boldsymbol{W}$ has an autoregressive structure of order one, that is,

$$W_t = \varrho W_{t-1} + \epsilon_t, \ \ t \geq 2,$$

where $\epsilon_t \sim \mathcal{N}(0, \sigma^2(1-\rho^2))$ iid $|\varrho| < 1$, and $W_1 \sim \mathcal{N}(0, \sigma^2)$. In this case, $\rho_W(|s-t|) = \varrho^{|s-t|}$ so that the autocorrelation function decreases exponentially with the lag between observations. This specification is particularly interesting for ratemaking purposes.

3.3 Model B5

This model postulates that there is a static baseline heterogeneity which is perturbated by iid annual effects. Specifically, $\Theta_{it} = R_i S_{it}$ where the S_{it}'s are iid and independent from R_i, $S_{it} \sim \mathcal{LN}(-\frac{\sigma_S^2}{2}, \sigma_S^2)$ and $R_i \sim \mathcal{LN}(-\frac{\sigma_R^2}{2}, \sigma_R^2)$. In this case, $\Theta_{it} \sim \mathcal{LN}(-\frac{\sigma_R^2+\sigma_S^2}{2}, \sigma_R^2 + \sigma_S^2)$. The vector $\boldsymbol{\Theta}_i$ is exchangeable. In this model the autocorrelation function between the random effects is thus constant since

$$\begin{aligned}\mathbb{C}\text{ov}[\Theta_{it}, \Theta_{it+h}] &= \mathbb{E}\Big[\mathbb{C}\text{ov}[\Theta_{it}, \Theta_{it+h}|R_i]\Big] + \mathbb{C}\text{ov}\Big[\mathbb{E}[\Theta_{it}|R_i], \mathbb{E}[\Theta_{it+h}|R_i]\Big] \\ &= \mathbb{C}\text{ov}\Big[R_i\mathbb{E}[S_{it}], R_i\mathbb{E}[S_{it+h}]\Big] = \mathbb{V}\text{ar}[R_i].\end{aligned}$$

3.4 Model B6

Models B3-B5 are based on LogNormal random effects. Now, we would also like to be able to specify other distributions for the random effects, like the Gamma law, for instance (which facilitates the Bayesian analysis of the data). In that respect, the copula construction is of prime interest.

Poisson hidden Markov models are described in MacDonald & Zucchini (1997). Such models generalize the autoregressive models for normal random variables with the help of Markov chains (the classical autoregressive of order one structure is replaced with Markov chain of order one). Provided G is continuous, the idea is to use a copula to describe the joint distribution of $(\Theta_{it}, \Theta_{it+1})$, $t = 1, 2, \ldots, \nu_i - 1$.

Let us recall that a bivariate copula C is the joint distribution for a bivariate distribution with uniform marginals. By virtue of Sklar's theorem, the joint distribution function of Θ_{t-1} and Θ_t can be represented as

$$\Pr[\Theta_{t-1} \leq \theta_{t-1}, \Theta_t \leq \theta_t] = C\big(G(\theta_{t-1}), G(\theta_t)\big)$$

for some copula $C(.,.)$. Provided G is continuous then $C(.,.)$ is unique. For more details about copulas, we refer the interested reader e.g. to Nelsen (1999).

The copula construction is useful for coupling distributions of continuous outcomes. The use of copulas for discrete outcomes (as claims frequencies) is more problematic because of the non-uniqueness of C outside the range of the underlying cdf's. In actuarial science, where the dependence is considered as only apparent (and generated by our ignorance of relevant risk factors), coupling random effects is usually preferred.

Let us denote the conditional distribution of the copula by $C_{2|1}(v|u) = \frac{\partial}{\partial u}C(u,v)$. Then, the transition distribution of the Markov chain can be expressed as

$$H(\theta_t|\theta_{t-1}) = \Pr[\Theta_t \le \theta_t | \Theta_{t-1} = \theta_{t-1}] = C_{2|1}\big(G(\theta_t)|G(\theta_{t-1})\big)$$

Now, if we denote by c the density of the copula C and by h the transition pdf, we get

$$h(\theta_t|\theta_{t-1}) = \frac{\partial}{\partial \theta_t} H(\theta_t|\theta_{t-1}) = c\big(G(\theta_{t-1}), G(\theta_t)\big) g(\theta_t).$$

Taking $C(u,v) = \min\{u,v\}$ (i.e. the Fréchet upper bound copula) yields comonotonic random effects; in this case, we find the static credibility model studied in Purcaru & Denuit (2001).

4 Statement T1

4.1 Stochastic orderings

Most positive dependence concepts aim to formalize the idea that large values of some component of a random vector tend to be associated with large values of the others. Therefore, we need to be able to decide whether a random vector is indeed "larger" than another one. To this end, we resort to stochastic orderings. These probabilistic tools will also be used to formalize the increasingness mentioned in statements T2-T4. In this paper, we use two classical order relations, namely stochastic dominance and likelihood ratio order.

Definition 4.1. Let $\boldsymbol{X} = (X_1, \cdots, X_n)$ and $\boldsymbol{Y} = (Y_1, \cdots, Y_n)$ be two n-dimensional random vectors. Then $\boldsymbol{X}$ is said to be smaller than $\boldsymbol{Y}$ in the stochastic dominance, written as $\boldsymbol{X} \preceq_{st} \boldsymbol{Y}$, if

$$\Pr[\boldsymbol{X} \in U] \le \Pr[\boldsymbol{Y} \in U], \quad \text{for all upper sets} \quad U \quad \text{in} \quad \mathbb{R}^n$$

(a set $U \subseteq \mathbb{R}^n$ is called *upper* if $\boldsymbol{y} \in U$ whenever $\boldsymbol{y} \ge \boldsymbol{x}$ and $\boldsymbol{x} \in U$.)

Intuitively $\preceq_{st}$ means that it is more likely that $\boldsymbol{Y}$ takes on "large" values (i.e. any value in an upper set U, for any upper set) than $\boldsymbol{X}$. Since any non-decreasing function can be obtained as the uniform limit of convex combinations of upper sets indicator functions, $\boldsymbol{X} \preceq_{st} \boldsymbol{Y}$ holds if, and only if, the inequality

$$\mathbb{E}[\phi(\boldsymbol{X})] \le \mathbb{E}[\phi(\boldsymbol{Y})]$$

is valid for all the non-decreasing functions $\phi : \mathbb{R}^n \to \mathbb{R}$, for which the expectations exists.

A univariate upper set is an interval of the form $(t, +\infty)$ so that Definition 4.1 gives $X \preceq_{st} Y \Leftrightarrow \Pr[X > t] \le \Pr[Y > t]$ for all $t \in \mathbb{R}$ for random variables X and Y.

Despite the intuitive nature of $\preceq_{st}$, we will also need a stronger order concept in the remainder of the paper, namely likelihood ratio order. This stochastic ordering is particularly useful in parametric models. Let us now recall the definition of the likelihood ratio order.

Definition 4.2. Let $\boldsymbol{X} = (X_1, \cdots, X_n)$ and $\boldsymbol{Y} = (Y_1, \cdots, Y_n)$ be two n-dimensional random vectors with absolutely continuous (or discrete) distribution functions and let $f_{\boldsymbol{X}}$ and $f_{\boldsymbol{Y}}$ denote their (continuous or discrete) density functions, respectively. Then $\boldsymbol{X}$ is said to be smaller than $\boldsymbol{Y}$ in the multivariate likelihood ratio order, written as $\boldsymbol{X} \preceq_{lr} \boldsymbol{Y}$, if:

$$f_{\boldsymbol{X}}(\boldsymbol{x} \wedge \boldsymbol{y}) f_{\boldsymbol{Y}}(\boldsymbol{x} \vee \boldsymbol{y}) \geq f_{\boldsymbol{X}}(\boldsymbol{x}) f_{\boldsymbol{Y}}(\boldsymbol{y}), \qquad \text{for every } \boldsymbol{x} \text{ and } \boldsymbol{y} \text{ in } \mathbb{R}^n$$

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

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

It is interesting to provide an intuitive explanation for the inequality defining $\preceq_{lr}$: it basically states that it is more likely to observe outcomes whose components are simultaneously all large for $\boldsymbol{Y}$ and small for $\boldsymbol{X}$ than outcomes mixing large and small values. Thus $\boldsymbol{Y}$ is more likely to assume the largest values and $\boldsymbol{X}$ the smallest ones.

If X and Y are random variables then Definition 4.2 gives $X \preceq_{lr} Y \Leftrightarrow f_X(x) f_Y(y) \geq f_X(y) f_Y(x)$ whenever $x \leq y$. Integrating (or summing in the discrete case) both sides of the aforementioned inequality over $x \in (-\infty, t]$ and $y \in (t, +\infty)$ yields

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

This shows that $\preceq_{lr}$ is stronger than $\preceq_{st}$, that is, $X \preceq_{lr} Y \Rightarrow X \preceq_{st} Y$. It can be shown that the multivariate version of $\preceq_{lr}$ is still stronger than $\preceq_{st}$, that is if $\boldsymbol{X}$ and $\boldsymbol{Y}$ are two n-dimensional random vectors, $\boldsymbol{X} \preceq_{lr} \boldsymbol{Y} \Rightarrow \boldsymbol{X} \preceq_{st} \boldsymbol{Y}$; see e.g. Shaked & Shanthikumar (1994).

4.2 Dependence concepts

Let us introduce the definitions of several dependence concepts for random vectors, which will be used in the remainder of this article. For more details about these concepts we refer the reader to Szekli (1995), Joe (1997) or Karlin & Rinott (1980).

Definition 4.3. Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ be a n-dimensional random vector.

(i) $\boldsymbol{X}$ 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 $g, h : \mathbb{R}^n \to \mathbb{R}$ simultaneously non-decreasing (or non-increasing) functions;

(ii) $\boldsymbol{X}$ 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 $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ (note that this amounts to $\boldsymbol{X} \preceq_{lr} \boldsymbol{X}$).

(iii) $\boldsymbol{X}$ 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 $x_j \leq x'_j,\ j \in \{1, \ldots, i-1\}$;

These dependence concepts are linked by the following serie of implications:

$$\boldsymbol{X} \text{ MPLRD} \Rightarrow \boldsymbol{X} \text{ CIS} \Rightarrow \boldsymbol{X} \text{ A}. \tag{4.1}$$

Moreover, they enjoy some functional invariance property, i.e. for all increasing functions $\varphi_1, \ldots, \varphi_n$, we have that:

$$\mathbf{X} \quad \text{MPLRD} \quad (\textit{resp.} \text{ CIS, A}) \;\Rightarrow\; \big(\varphi_1(X_1), \ldots \varphi_n(X_n)\big) \quad \text{MPLRD} \quad (\textit{resp.} \text{ CIS, A}).$$

In the particular case $n = 2$, the MTP_2 concept reduces to the classical TP_2, which is defined as follows: 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

$$f(s_1, t_1) f(s_2, t_2) \geq f(s_1, t_2) f(s_2, t_1) \tag{4.2}$$

holds true for any $s_1 \leq s_2 \in \mathcal{S}$ and $t_1 \leq t_2 \in \mathcal{T}$. The bivariate TP_2 is of great interest for studying MTP_2 since Kemperman (1977) proved that TP_2 in pairs and MTP_2 are equivalent. This result will play an important role in our study, and is thus recalled next.

Property 4.4. *Provided the support of $\boldsymbol{X}$ is a lattice (for instance $\mathbb{R}^n$ or $\mathbb{N}^n$), $\boldsymbol{X}$ is MPLRD if, and only if, its multivariate probability density function $f_{\boldsymbol{X}}$ is TP_2 in pairs, that is for all $i \neq j$*

$$(x_i, x_j) \mapsto f_{\boldsymbol{X}}(\boldsymbol{x}) \text{ is } TP_2 \text{ for fixed } x_k,\ k \neq i, j.$$

The latter property offers an effective procedure for checking the MPLRD nature of some random vector $\boldsymbol{X}$, since it suffices to deal with bivariate functions and to verify the simpler TP_2 condition.

A fundamental property of MTP_2 known as the *basic composition formula* from Karlin (1968) is as follows.

Property 4.5. *Let $\mathcal{S} = \times_{i=1}^n \mathcal{S}_i$, $\mathcal{T} = \times_{i=1}^n \mathcal{T}_i$ and $\mathcal{U} = \times_{i=1}^n \mathcal{U}_i$. Given some functions $h_1 : \mathcal{S} \times \mathcal{T} \to \mathbb{R}$ and $h_2 : \mathcal{T} \times \mathcal{U} \to \mathbb{R}$, let us define the function h_3 as*

$$h_3(\boldsymbol{s}, \boldsymbol{u}) = \int_{\boldsymbol{t} \in \mathcal{T}} h_1(\boldsymbol{s}, \boldsymbol{t}) h_2(\boldsymbol{t}, \boldsymbol{u}) d\sigma(\boldsymbol{t})$$

where the integral is assumed to converge absolutely and $\sigma = \sigma_1 \times \ldots \times \sigma_n$. If h_1 is MTP_2 on $\mathcal{S} \times \mathcal{T}$ and h_2 is MTP_2 on $\mathcal{T} \times \mathcal{U}$ then h_3 is MTP_2 on $\mathcal{S} \times \mathcal{U}$.

4.3 Proof of statement T1

Before formalizing statement T1, we need to establish the following technical poperty.

Property 4.6. *In the model B1-B2, given $k \leq k' \in \mathbb{N}$ and $\theta \leq \theta' \in \mathbb{R}^+$,*

$$\Pr[N_{it} = k|\Theta_{it} = \theta] \Pr[N_{it} = k'|\Theta_{it} = \theta'] \geq \Pr[N_{it} = k|\Theta_{it} = \theta'] \Pr[N_{it} = k'|\Theta_{it} = \theta].$$

Proof. Dividing the right-hand side by the left-hand side of the expression, we obtain :

$$\frac{(\lambda_{it}\theta')^k(\lambda_{it}\theta)^{k'}}{(\lambda_{it}\theta)^k(\lambda_{it}\theta')^{k'}} = \left\{\frac{\theta}{\theta'}\right\}^{k'-k} \leq 1$$

so that the announced result holds true. □

Note that Property 4.6 means that

$$[N_{it}|\Theta_{it} = \theta] \preceq_{lr} [N_{it}|\Theta_{it} = \theta'] \text{ for } \theta \leq \theta' \in \mathbb{R}^+;$$

increasing Θ_{it} makes thus N_{it} "larger" in the $\preceq_{lr}$-sense.

We are now ready to state the main result of this section, namely that positive dependence for $\boldsymbol{\Theta}_i$ is transmitted to $\boldsymbol{N}_i$. This result is closely related to previous studies by Whitt (1979), Fahmy et al. (1982) and Shaked & Spizzichino (1998).

Proposition 4.7. *In the model B1-B2,*

(i) $\boldsymbol{\Theta}_i$ *is MPLRD* $\Rightarrow$ $\boldsymbol{N}_i$ *is MPLRD.*

(ii) $\boldsymbol{\Theta}_i$ *is associated* $\Rightarrow$ $\boldsymbol{N}_i$ *is associated.*

Proof. (i) The joint probability function of $\boldsymbol{N}_i$ can be expressed as

$$\Pr[\boldsymbol{N}_i = \boldsymbol{k}] = \int_{\boldsymbol{\theta}_i \geq \boldsymbol{0}} \left\{\prod_{t=1}^{\nu_i} \Pr[N_{it} = k_t|\Theta_{it} = \theta_{it}]\right\} f_{\boldsymbol{\Theta}_i}(\boldsymbol{\theta}_i) d\boldsymbol{\theta}_i,$$

where $f_{\boldsymbol{\Theta}_i}$ stands for the pdf of $\boldsymbol{\Theta}_i$. Denote as

$$\overline{f}(k_1, \ldots, k_{\nu_i}, \theta_{i1}, \ldots, \theta_{i\nu_i}) \equiv \prod_{t=1}^{\nu_i} \Pr[N_{it} = k_t|\Theta_{it} = \theta_{it}].$$

¿From Property 4.6 it follows that $\overline{f}$ is a product of TP_2 functions and is hence MTP_2 in $(\boldsymbol{k}, \boldsymbol{\theta})$ by virtue of Property 4.4. Now, invoking the basic composition formula of Property 4.5 (with $h_1 = \overline{f}$ and $h_2 = f_{\boldsymbol{\Theta}_i}$, the latter function being MTP_2 since $\boldsymbol{\Theta}_i$ has been assumed MPLRD) we obtain that $\boldsymbol{N}_i$ is MPLRD.
(ii) We have to show that the inequality

$$\mathbb{C}\text{ov}[g(\boldsymbol{N}_i), h(\boldsymbol{N}_i)] \geq 0$$

holds true for all functions $g, h : \mathbb{N}^{\nu_i} \to \mathbb{R}$ simultaneously non-decreasing in each of their arguments. By reformulating the covariance we obtain the expression:

$$\mathbb{C}\mathrm{ov}[g(\boldsymbol{N}_i), h(\boldsymbol{N}_i)] = \mathbb{E}\Big[\mathbb{C}\mathrm{ov}[g(\boldsymbol{N}_i), h(\boldsymbol{N}_i)|\boldsymbol{\Theta}_i]\Big] + \mathbb{C}\mathrm{ov}\Big[\mathbb{E}[g(\boldsymbol{N}_i)|\boldsymbol{\Theta}_i], \mathbb{E}[h(\boldsymbol{N}_i)|\boldsymbol{\Theta}_i]\Big].$$

Given $\boldsymbol{\Theta}_i$, the components of $\boldsymbol{N}_i$ are independent by virtue of B2. ¿From Barlow & Proschan (1975) independent random variables are associated and hence the covariance in the first term is positive and so is the whole first term. Let us denote by $\overline{g}(\boldsymbol{\theta}_i) \equiv \mathbb{E}[g(\boldsymbol{N}_i)|\boldsymbol{\Theta}_i = \boldsymbol{\theta}_i]$ and by $\overline{h}(\boldsymbol{\theta}_i) \equiv \mathbb{E}[h(\boldsymbol{N}_i)|\boldsymbol{\Theta}_i = \boldsymbol{\theta}_i]$. As it was seen in Property 4.6 we have that:

$$[N_{it}|\Theta_{it} = \theta_{it}] \preceq_{lr} [N_{it}|\Theta_{it} = \theta'_{it}], \quad \text{for all} \quad \theta_{it} \leq \theta'_{it}.$$

By Theorem 6.E.3. from Shaked & Shanthikumar (1994), the $\preceq_{lr}$ order is closed under conjunctions, and so we deduce that

$$[\boldsymbol{N}_i|\boldsymbol{\Theta}_i = \boldsymbol{\theta}_i] \preceq_{lr} [\boldsymbol{N}_i|\boldsymbol{\Theta}_i = \boldsymbol{\theta}'_i], \quad \text{for all} \quad \boldsymbol{\theta}_i \leq \boldsymbol{\theta}'_i$$

$$\Rightarrow [\boldsymbol{N}_i|\boldsymbol{\Theta}_i = \boldsymbol{\theta}_i] \preceq_{st} [\boldsymbol{N}_i|\boldsymbol{\Theta}_i = \boldsymbol{\theta}'_i], \quad \text{for all} \quad \boldsymbol{\theta}_i \leq \boldsymbol{\theta}'_i.$$

Since f and g are increasing functions, it follows that $\overline{g}(\cdot)$ and $\overline{h}(\cdot)$ are both increasing functions in $\boldsymbol{\Theta}_i$. In the hypothesis $\boldsymbol{\Theta}_i$ was supposed to be associated, thus the second term in the right is also positive, and hence the conclusion follows. □

4.4 Statement T1 in the models B3-B6

Since the type of dependence existing between the N_{it}'s is induced by the type of dependence existing between the Θ_{it}'s, as presented in Property 4.7, it remains to study the dependence structure of the random effects Θ_{it} in the models mentionned at the begining of the article, i.e. B3-B6.

4.4.1 Model B3

The first model is rather general. Not surprisingly, the type of dependence between the Θ_{it}'s is induced by the form of the covariance matrix $\boldsymbol{\Sigma}$ of $\boldsymbol{W}$.

Property 4.8. *In the model B1-B2-B3, the following properties hold:*

(i) *If the inverse of the covariance matrix $\boldsymbol{\Sigma}$ of $\boldsymbol{W}$ has all the off-diagonal components nonpositive then $\boldsymbol{N}_i$ is MPLRD.*

(ii) *If all the components of the covariance matrix $\boldsymbol{\Sigma}$ of $\boldsymbol{W}$ are nonnegative then $\boldsymbol{N}_i$ is associated.*

Proof. (i) From Tong (1990), we know that $\boldsymbol{W}$ is MPLRD if, and only if, all the off-diagonal components of the inverse of its covariance matrix are non-positive. Since MPLRD is closed under increasing transformation, each $\boldsymbol{\Theta}_i$ is then MPLRD. The announced result then follows from Proposition 4.7(i).
(ii) The reasoning is similar to the one in (i), since we know from Pitt (1982) that $\boldsymbol{W}$ is associated if, and only if, all the elements of its covariance matrix are non-negative. □

4.4.2 Model B4

In this second model, the covariance structure of $\boldsymbol{W}$ is expressed in terms of the autoregressive parameter ϱ. Provided ϱ is non-negative, we get strong positive dependence for the components of $\boldsymbol{N}_i$, as shown in the following result.

Property 4.9. *In the model B1-B2-B4, $\varrho \geq 0 \Rightarrow \boldsymbol{N}_i$ MPLRD.*

Proof. In the model B4, the elements of $\boldsymbol{\Sigma}$ are given by:

$$\sigma_{tt} = \sigma_W^2, \quad \sigma_{st} = \sigma_{ts} = \varrho^{|s-t|}\sigma_W^2 \quad \text{for} \quad |s-t| \geq 1.$$

In this case the off-diagonal elements of the matrix $\boldsymbol{R} = \boldsymbol{\Sigma}^{-1}$ are given by :

$$r_{t,t+1} = r_{t+1,t} = -\frac{\varrho}{\sigma_W^2} \quad \text{for} \quad t = 1, \ldots, \nu_{\max} - 1 \quad \text{and} \quad r_{st} = 0 \quad \text{for} \quad |s-t| \geq 2$$

and are all non positive when $\varrho \geq 0$. Hence $\boldsymbol{W}$ is MPLRD. Since the MPLRD property is functionally invariant, $\boldsymbol{\Theta}_i$ is also MPLRD and hence $\boldsymbol{N}_i$ by Proposition 4.7(i). □

4.4.3 Model B5

Let us now turn to the exchangeable random effects. When the LogNormal specification is retained for R_i and the S_{it}'s, no further conditions are needed for MPLRD, as shown in the following result.

Property 4.10. *In the model B1-B2-B5, $\boldsymbol{N}_i$ is MPLRD.*

Proof. The proof is in the vein of Karlin & Rinott (1980, Proposition 3.9). In the model B5, $\Theta_{it} = R_i S_{it}$, where R_i, S_{it} are Log-Normal random variables, independent from one another. The joint density of $\boldsymbol{\Theta}_i = (\Theta_{i1}, \ldots, \Theta_{i\nu_i})$ is given by

$$f_{\boldsymbol{\Theta}_i}(\theta_{i1}, \ldots, \theta_{i\nu_i}) = \int_{r \geq 0} \left\{ \prod_{t=1}^{\nu_i} f_{S_{it}}\left(\frac{\theta_{it}}{r}\right) \right\} f_{R_i}(r) dr. \tag{4.3}$$

Let us show that $f_{S_{it}}(\frac{\theta_{it}}{r})$ is TP_2 in (θ_{it}, r), i.e. that for all $\theta_1 < \theta_2$ and all $r_1 < r_2$ the inequality

$$f_{S_{it}}\left(\frac{\theta_1}{r_1}\right) f_{S_{it}}\left(\frac{\theta_2}{r_2}\right) \geq f_{S_{it}}\left(\frac{\theta_1}{r_2}\right) f_{S_{it}}\left(\frac{\theta_2}{r_1}\right)$$

holds true. This amounts to prove that

$$\exp\left\{-\frac{1}{2\sigma_S^2}\left(\left(\ln\frac{\theta_1}{r_1} - \frac{\sigma_S^2}{2}\right)^2 + \left(\ln\frac{\theta_2}{r_2} - \frac{\sigma_S^2}{2}\right)^2\right)\right\}$$

$$\geq \exp\left\{-\frac{1}{2\sigma_S^2}\left(\left(\ln\frac{\theta_1}{r_2} - \frac{\sigma_S^2}{2}\right)^2 + \left(\ln\frac{\theta_2}{r_1} - \frac{\sigma_S^2}{2}\right)^2\right)\right\},$$

or, equivalently, that

$$-\frac{1}{2\sigma_S^2}\Big\{(\ln\theta_1-\ln r_1)^2+(\ln\theta_2-\ln r_2)^2-(\ln\theta_1-\ln r_2)^2-(\ln\theta_2-\ln r_1)^2\Big\}\geq 0$$

which reduces to

$$\frac{1}{\sigma_S^2}(\ln\theta_2-\ln\theta_1)(\ln r_2-\ln r_1)\geq 0.$$

The latter relation is obviously true. Thus $f_{S_{it}}(\frac{\theta_{it}}{r})$ is TP_2 in (θ_{it},r). The integrand of (4.3) is then MTP_2 in $(\boldsymbol{\theta}_i,r)$ whence it follows that $f_{\boldsymbol{\Theta}_i}$ is MTP_2. $\square$

4.4.4 Model B6

In the general autoregressive model induced by bivariate copulas C, the dependence structure of $\boldsymbol{N}_i$ is induced by the properties of C, as expected.

Property 4.11. *In the model B1-B2-B6, $\boldsymbol{N}_i$ is MPLRD provided C is TP_2, that is*

$$c(u,v)c(u',v')\geq c(u',v)c(u,v') \text{ whenever } u\leq u',\ \ v\leq v'.$$

Proof. Let us write the joint pdf of $\boldsymbol{N}_i$ as

$$\Pr[\boldsymbol{N}_i=\boldsymbol{k}]=\int_{\boldsymbol{\theta}_i\geq\boldsymbol{0}}\left\{\prod_{t=1}^{\nu_i}\Pr[N_{it}=k_t|\Theta_{it}=\theta_{it}]\right\}f_{\boldsymbol{\Theta}_i}(\boldsymbol{\theta}_i)d\boldsymbol{\theta}_i.$$

Now, exploiting the autoregressive of order 1 structure yields

$$\begin{aligned} f_{\boldsymbol{\Theta}_i}(\boldsymbol{\theta}_i) &= g(\theta_{i1})h(\theta_{i2}|\theta_{i1})\ldots h(\theta_{i\nu_i}|\theta_{i\nu_i-1}) \\ &= \left\{\prod_{t=1}^{\nu_i}g(\theta_{it})\right\}c\big(G(\theta_{i1}),G(\theta_{i2})\big)\ldots c\big(G(\theta_{i\nu_i-1}),G(\theta_{i\nu_i})\big). \end{aligned}$$

If $(u,v)\mapsto c(u,v)$ is TP_2 then $(\theta_{it},\theta_{it+1})\mapsto c\big(G(\theta_{it}),G(\theta_{it+1})\big)$ is also TP_2 so that $f_{\boldsymbol{\Theta}_i}$ is MTP_2 by Property 4.4. Then, invoking Property 4.5 shows that $\boldsymbol{k}\mapsto\Pr[\boldsymbol{N}_i=\boldsymbol{k}]$ is MTP_2, which ends the proof. $\square$

Let us give now some examples of copulas, which enjoy the TP_2 property for some values of their parameters:

(i) *Frank*: For $\alpha\in\mathbb{R}$ the Frank copula is given by:

$$C(u,v;\alpha)=-\alpha^{-1}\log([\eta-(1-e^{-\alpha u})(1-e^{-\alpha v})]/\eta),\qquad\text{where}\quad\eta=1-e^{-\alpha}$$

and the density:

$$c(u,v;\alpha)=\frac{\alpha\eta e^{-\alpha(u+v)}}{[\eta-(1-e^{-\alpha u})(1-e^{-\alpha v})]^2}$$

The TP_2 property of the copula density is verified for $\alpha\geq 0$.

(ii) *Normal*: For $\alpha \in [-1, 1]$, the normal copula is given by

$$C_\alpha(u_1, u_2) = H_\alpha\left(\Phi^{-1}(u_1), \Phi^{-1}(u_2)\right),$$

where Φ is the $\mathcal{N}(0,1)$ cdf and H_α is the bivariate standard normal cdf with correlation coefficient α, i.e.

$$C_\alpha(u_1, u_2) = \frac{1}{2\pi\sqrt{1-\alpha^2}} \int_{\xi_1=-\infty}^{\Phi^{-1}(u_1)} \int_{\xi_2=-\infty}^{\Phi^{-1}(u_2)} \exp\left\{\frac{-(\xi_1^2 - 2\alpha\xi_1\xi_2 + \xi_2^2)}{2(1-\alpha^2)}\right\} d\xi_1 d\xi_2$$

and the density

$$c_\alpha(u_1, u_2) = \frac{1}{\sqrt{1-\alpha^2}} \exp\left\{\frac{-(\zeta_1^2 - 2\alpha\zeta_1\zeta_2 + \zeta_2^2)}{2(1-\alpha^2)}\right\} \exp\left\{\frac{\zeta_1^2 + \zeta_2^2}{2}\right\}$$

where $\zeta_i = \Phi^{-1}(u_i)$, $i = 1, 2$. If $\alpha \geq 0$, C_α expresses strong positive dependence since any random couple with copula C_α is TP_2.

In the two following copulas the TP_2 property holds without restriction over the domain of the dependence parameter.

(iii) *Gumbel*: Let us denote by $\tilde{u} = -\ln u$ and $\tilde{v} = -\ln v$. Then for $\alpha \geq 1$ the Gumbel copula is given by

$$C(u, v; \alpha) = \exp\left\{-\left(\tilde{u}^\alpha + \tilde{v}^\alpha\right)^{-1/\alpha}\right\}$$

and the density is

$$c(u, v; \alpha) = C(u, v; \alpha)\frac{(\tilde{u}\tilde{v})^{\alpha-1}}{uv} \cdot \frac{(\tilde{u}^\alpha + \tilde{v}^\alpha)^{1/\alpha} + \alpha - 1}{(\tilde{u}^\alpha + \tilde{v}^\alpha)^{2-1/\alpha}}$$

(iv) *Clayton*: For $\alpha \geq 0$ the Clayton copula is given by

$$C(u, v; \alpha) = (u^{-\alpha} + v^{-\alpha} - 1)^{-1/\alpha}$$

and the density is

$$c(u, v; \alpha) = \frac{\alpha + 1}{(uv)^{\alpha+1}[u^{-\alpha} + v^{-\alpha} - 1]^{2+1/\alpha}}$$

5 Statement T2

Let us now establish the following result inspired from Theorems 3-4 in Fahmy et al. (1982). We purpose to prove that policyholders reporting more claims in the past become more dangerous on unobservable characteristics than those who report less claims.

Proposition 5.1. *In the model B1-B2,*

$$[\boldsymbol{\Theta}_i | \boldsymbol{N}_i = \boldsymbol{k}] \preceq_{lr} [\boldsymbol{\Theta}_i | \boldsymbol{N}_i = \boldsymbol{k}'], \quad \textit{whenever} \quad \boldsymbol{k} \leq \boldsymbol{k}'$$

provided $\boldsymbol{\Theta}_i$ *is MPLRD.*

Proof. The conditional pdf of $\boldsymbol{\Theta}_i$ given $[\boldsymbol{N}_i = \boldsymbol{k}]$ can be expressed as

$$f_{\boldsymbol{\Theta}_i}(\boldsymbol{\theta}|\boldsymbol{k}) = c(\boldsymbol{k}) \Pr[\boldsymbol{N}_i = \boldsymbol{k}|\boldsymbol{\Theta}_i = \boldsymbol{\theta}] f_{\boldsymbol{\Theta}_i}(\boldsymbol{\theta})$$

where $c(\boldsymbol{k})$ is a normalizing constant given by

$$[c(\boldsymbol{k})]^{-1} = \Pr[\boldsymbol{N}_i = \boldsymbol{k}] = \int_{\boldsymbol{\lambda}\geq\boldsymbol{0}} \Pr[\boldsymbol{N}_i = \boldsymbol{k}|\boldsymbol{\Theta}_i = \boldsymbol{\lambda}] f_{\boldsymbol{\Theta}_i}(\boldsymbol{\lambda}) d\boldsymbol{\lambda}.$$

Let us now consider two claim histories $\boldsymbol{k} \leq \boldsymbol{k}'$ as well as two possible values for the random effects $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$. Starting from

$$\begin{aligned}
f_{\boldsymbol{\Theta}_i}(\boldsymbol{\theta}|\boldsymbol{k}) f_{\boldsymbol{\Theta}_i}(\boldsymbol{\theta}'|\boldsymbol{k}') &= c(\boldsymbol{k})c(\boldsymbol{k}') \Pr[\boldsymbol{N}_i = \boldsymbol{k}|\boldsymbol{\Theta}_i = \boldsymbol{\theta}] \Pr[\boldsymbol{N}_i = \boldsymbol{k}'|\boldsymbol{\Theta}_i = \boldsymbol{\theta}'] f_{\boldsymbol{\Theta}_i}(\boldsymbol{\theta}) f_{\boldsymbol{\Theta}_i}(\boldsymbol{\theta}') \\
&= c(\boldsymbol{k})c(\boldsymbol{k}') \left\{ \prod_{t=1}^{\nu_i} \Pr[N_{it} = k_t|\Theta_{it} = \theta_{it}] \right\} \\
&\qquad \left\{ \prod_{t=1}^{\nu_i} \Pr[N_{it} = k'_t|\Theta_{it} = \theta'_{it}] \right\} f_{\boldsymbol{\Theta}_i}(\boldsymbol{\theta}) f_{\boldsymbol{\Theta}_i}(\boldsymbol{\theta}')
\end{aligned}$$

we get from the MPLRD nature of $\boldsymbol{\Theta}_i$ that

$$\begin{aligned}
f_{\boldsymbol{\Theta}_i}(\boldsymbol{\theta}|\boldsymbol{k}) f_{\boldsymbol{\Theta}_i}(\boldsymbol{\theta}'|\boldsymbol{k}') &\leq c(\boldsymbol{k})c(\boldsymbol{k}') \left\{ \prod_{t=1}^{\nu_i} \Pr[N_{it} = k_t \wedge k'_t|\Theta_{it} = \theta_{it} \wedge \theta'_{it}] \right\} \\
&\qquad \left\{ \prod_{t=1}^{\nu_i} \Pr[N_{it} = k_t \vee k'_t|\Theta_{it} = \theta_{it} \vee \theta'_{it}] \right\} f_{\boldsymbol{\Theta}_i}(\boldsymbol{\theta}\wedge\boldsymbol{\theta}') f_{\boldsymbol{\Theta}_i}(\boldsymbol{\theta}\vee\boldsymbol{\theta}') \\
&= c(\boldsymbol{k})c(\boldsymbol{k}') \Pr[\boldsymbol{N}_i = \boldsymbol{k}\wedge\boldsymbol{k}'|\boldsymbol{\Theta}_i = \boldsymbol{\theta}\wedge\boldsymbol{\theta}'] \Pr[\boldsymbol{N}_i = \boldsymbol{k}\vee\boldsymbol{k}'|\boldsymbol{\Theta}_i = \boldsymbol{\theta}\vee\boldsymbol{\theta}'] \\
&\qquad f_{\boldsymbol{\Theta}_i}(\boldsymbol{\theta}\wedge\boldsymbol{\theta}') f_{\boldsymbol{\Theta}_i}(\boldsymbol{\theta}\vee\boldsymbol{\theta}') \\
&= \Big\{ c(\boldsymbol{k}) \Pr[\boldsymbol{N}_i = \boldsymbol{k}|\boldsymbol{\Theta}_i = \boldsymbol{\theta}\wedge\boldsymbol{\theta}'] f_{\boldsymbol{\Theta}_i}(\boldsymbol{\theta}\wedge\boldsymbol{\theta}') \Big\} \\
&\qquad \Big\{ c(\boldsymbol{k}') \Pr[\boldsymbol{N}_i = \boldsymbol{k}'|\boldsymbol{\Theta}_i = \boldsymbol{\theta}\vee\boldsymbol{\theta}'] f_{\boldsymbol{\Theta}_i}(\boldsymbol{\theta}\vee\boldsymbol{\theta}') \Big\} \\
&= f_{\boldsymbol{\Theta}_i}(\boldsymbol{\theta}\wedge\boldsymbol{\theta}'|\boldsymbol{k}) f_{\boldsymbol{\Theta}_i}(\boldsymbol{\theta}\vee\boldsymbol{\theta}'|\boldsymbol{k}')
\end{aligned}$$

whence the announced result follows. □

The conditions under which $\boldsymbol{\Theta}_i$ is MPLRD have been studied in Section 4.4 for the models B3-B6. It is interesting to contrast Proposition 5.1 to the corresponding result for the static credibility model (Proposition 3.4 in Purcaru & Denuit (2001)). In the static case, the condition involves $N_{i\bullet}$, the total number of claims reported by policyholder i in the past: increasing the value of $N_{i\bullet}$ makes the unique random effect Θ_i larger in the univariate $\preceq_{lr}$-sense. In that framework, the distribution of the claims during the driving career of the policyholder does not matter. In the dynamic credibility model, the conditioning is now on the whole vector $\boldsymbol{N}_i$ recording annual claim numbers for policyholder i. To make the vector $\boldsymbol{\Theta}_i$ of random effects larger in the multivariate $\preceq_{lr}$-sense, we have to increase each and every annual claim number, which drastically restricts the generality of the result.

6 Statements T3 and T4 in models B4-B6

The whole credibility theory aims to predict as accurately as possible the future claims given the past. This section focusses on predictive distributions, that is the distribution of the future number of claims $N_{i\nu_i+1}$ given past claims $\boldsymbol{N}_i$. Model B3 is hardly used to perform prediction on longitudinal basis. This is due to the fact that the correlation function ρ_W has to be continued beyond the observation period before evaluating experience premiums (see Pinquet et al. (2001) for a complete treatment of this subject). In this section, we concentrate on models B4-B6 and assume that the conditions of Properties 4.9, 4.10 and 4.11 are fulfilled. This ensures that the vector $(\Theta_{i1}, \ldots, \Theta_{i\nu_i}, \Theta_{i\nu_i+1})$ is also MPLRD.

Property 6.1. *In the model B1-B2 completed by either B4, B5 or B6,* $[\Theta_{i\nu_i+1}|\boldsymbol{N}_i = \boldsymbol{k}] \preceq_{lr} [\Theta_{i\nu_i+1}|\boldsymbol{N}_i = \boldsymbol{k}']$ *for all* $\boldsymbol{k} \leq \boldsymbol{k}'$ *under the conditions of Property 4.9, 4.10 or 4.11.*

Proof. The aim is to prove that the inequality

$$f(\theta_{i\nu_i+1}|\boldsymbol{k})f(\theta'_{i\nu_i+1}|\boldsymbol{k}') \geq f(\theta'_{i\nu_i+1}|\boldsymbol{k})f(\theta_{i\nu_i+1}|\boldsymbol{k}') \tag{6.1}$$

holds true for all $\boldsymbol{k} \leq \boldsymbol{k}'$ and $\theta_{i\nu_i+1} \leq \theta'_{i\nu_i+1}$. Bayes formula gives

$$f(\theta_{i\nu_i+1}|\boldsymbol{k}) = \frac{\Pr[\boldsymbol{N}_i = \boldsymbol{k}|\Theta_{i\nu_i+1} = \theta_{i\nu_i+1}]g(\theta_{i\nu_i+1})}{\int_{\lambda\geq 0}\Pr[\boldsymbol{N}_i = \boldsymbol{k}|\Theta_{i\nu_i+1} = \lambda]g(\lambda)d\lambda}$$

and thus showing (6.1) is equivalent to verify that the inequality

$$\begin{aligned}&\Pr[\boldsymbol{N}_i = \boldsymbol{k}|\Theta_{i\nu_i+1} = \theta_{i\nu_i+1}]\Pr[\boldsymbol{N}_i = \boldsymbol{k}'|\Theta_{i\nu_i+1} = \theta'_{i\nu_i+1}]\\ &\geq \Pr[\boldsymbol{N}_i = \boldsymbol{k}'|\Theta_{i\nu_i+1} = \theta_{i\nu_i+1}]\Pr[\boldsymbol{N}_i = \boldsymbol{k}|\Theta_{i\nu_i+1} = \theta'_{i\nu_i+1}],\end{aligned}$$

holds for all $\boldsymbol{k} \leq \boldsymbol{k}'$ and $\theta_{i\nu_i+1} \leq \theta'_{i\nu_i+1}$. Let us denote as $g(\boldsymbol{k}, \theta)$ the joint density of $(\boldsymbol{N}_i, \Theta_{i\nu_i+1})$ evaluated at $(\boldsymbol{k}, \theta)$. It then remains to prove that

$$g(\boldsymbol{k}, \theta_{i\nu_i+1})g(\boldsymbol{k}', \theta'_{i\nu_i+1}) \geq g(\boldsymbol{k}', \theta_{i\nu_i+1})g(\boldsymbol{k}, \theta'_{i\nu_i+1}).$$

To establish the latter inequality, we use an interesting property of the MTP_2 concept, which stipulates that if the joint distribution of a random vector is MTP_2, then the same property still holds for any of the marginals. In our case,

$$g(k_1, \ldots, k_{\nu_i}, \theta_{i\nu_i+1}) = \sum_{k_{\nu_i+1}} \int \ldots \int g(k_1, \ldots, k_{\nu_i}, k_{\nu_i+1}, \theta_{i1}, \ldots, \theta_{i\nu_i}, \theta_{i\nu_i+1})\ d\theta_{i1}\ldots d\theta_{i\nu_i}$$

in obvious notation. The joint density of $(N_{i1}, \ldots, N_{i\nu_i+1}, \Theta_{i1}, \ldots, \Theta_{i\nu_i+1})$ given by

$$g(k_1, \ldots, k_{\nu_i+1}, \theta_{i1}, \ldots, \theta_{i\nu_i+1}) \equiv \left\{\prod_{t=1}^{\nu_i+1}\Pr[N_{it} = k_t|\Theta_{it} = \theta_{it}]\right\} f_{\boldsymbol{\Theta}_i}(\theta_{i1}, \ldots, \theta_{i\nu_i+1})$$

is obviously MTP_2: indeed, the product is TP_2 in pair and hence it is MTP_2 by Property 4.4 while from hypothesis $(\Theta_{i1}, \ldots, \Theta_{i\nu_i}, \Theta_{i\nu_i+1})$ is MPLRD and therefore $f_{\boldsymbol{\Theta}_i}(\cdot)$ is a MTP_2 function. Hence the result. □

Now that we have shown that a policyholder having reported more claims in the past will be more dangerous on the unobservable characteristics relating to the year $\nu_i + 1$, we would like to establish the same result for the future number of claims. This is precisely done next.

Property 6.2. *Under the assumptions of Property 6.1,* $[N_{i\nu_i+1}|\boldsymbol{N}_i = \boldsymbol{k}] \preceq_{lr} [N_{i\nu_i+1}|\boldsymbol{N}_i = \boldsymbol{k}']$*, for all* $\boldsymbol{k} \leq \boldsymbol{k}'$.

Proof. We have to show that for all $n \leq n'$ and $\boldsymbol{k} \leq \boldsymbol{k}'$ the following inequality

$$\Pr[N_{i\nu_i+1} = n|\boldsymbol{N}_i = \boldsymbol{k}] \Pr[N_{i\nu_i+1} = n'|\boldsymbol{N}_i = \boldsymbol{k}']$$

$$\geq \Pr[N_{i\nu_i+1} = n|\boldsymbol{N}_i = \boldsymbol{k}'] \Pr[N_{i\nu_i+1} = n'|\boldsymbol{N}_i = \boldsymbol{k}]$$

holds true. The conditionnal probability could also be expressed in the form :

$$\Pr[N_{i\nu_i+1} = n|\boldsymbol{N}_i = \boldsymbol{k}] = \int_{\boldsymbol{\theta} \geq \boldsymbol{0}} \Pr[N_{i\nu_i+1} = n|\Theta_{i\nu_i+1} = \theta] \ g(\theta|\boldsymbol{k}) \ d\theta$$

where $g(\cdot|\boldsymbol{k})$ denotes the conditional density of $\Theta_{i\nu_i+1}$ given $\boldsymbol{N}_i = \boldsymbol{k}$. The first factor of the integrand is TP_2 in (n, θ), based on Property 4.6. Moreover, we have seen in Property 6.1 that $g_\Theta(\theta|\boldsymbol{k})$ is MTP_2 in $(k_1, \ldots, k_{\nu_i}, \theta)$. By applying the basic composition formula for

$$\overline{f}(n, \theta) \equiv \Pr[N_{i\nu_i+1} = n|\Theta_{i\nu_i+1} = \theta] \quad \text{and} \quad \overline{g}(\theta, \boldsymbol{k}) \equiv g_\Theta(\theta|\boldsymbol{k})$$

it follows that $(n, \boldsymbol{k}) \mapsto \Pr[N_{i\nu_i+1} = n|\boldsymbol{N}_i = \boldsymbol{k}]$ is MTP_2, which ends our proof. $\square$

7 Conclusions

This paper aimed to examine the kind of dependence induced by dynamic credibility models. Such models will certainly be widely used in the future, in order to increase the accuracy of the ratemaking. As expected, the structure of dependence between the observable claim numbers is inherited from the type of correlation existing between the unobservable random effects modelling residual heterogeneity.

The results of Sections 5 and 6 certainly need to be refined in the future. Indeed, the conditioning events $[\boldsymbol{N}_i = \boldsymbol{k}]$ and $[\boldsymbol{N}_i = \boldsymbol{k}']$ with $\boldsymbol{k} \leq \boldsymbol{k}'$ are rather restrictive (but seem necessary to get a ranking in the $\preceq_{lr}$-sense). In the future, we purpose to examine other claim histories $\boldsymbol{k}$ and $\boldsymbol{k}'$, satisfying for instance

$$\sum_{t=1}^{j} k_t \leq \sum_{t=1}^{j} k'_t \text{ for } j = 1, \ldots, \nu_i$$

together with $\sum_{t=1}^{\nu_i} k_t = \sum_{t=1}^{\nu_i} k'_t$. Another approach consists in comparing $\boldsymbol{k}$ and $\boldsymbol{k}'$ through majorization. Of course, $\preceq_{lr}$ should then be replaced with appropriate stochastic orderings.

Acknowledgements

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

References

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

[2] Fahmy, S., Pereira, C.A., Proschan, F., and Shaked, M. (1982). The influence on the sample on the posterior distribution. *Communications in Statistics – Theory & Methods* **11**, 1757-1768.

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

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

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

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

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

[8] MacDonald, I.L., and Zucchini, W. (1997) *Hidden Markov and Other Models for Discrete-valued Times Series.* Chapman & Hall, New York.

[9] Nelsen, R.B. (1999). *An introduction to copulas.* Springer-Verlag, New York.

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

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

[12] Pitt, L.D. (1982). Positively correlated normal variables are associated. *The Annals of Probability* **10**, 496-499.

[13] Purcaru, O., & Denuit, M. (2001). On the dependence induced by frequency credibility models. I. Time-invariant random effects. Discussion Paper 01-34, Institut de Statistique, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

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

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

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

[17] Szekli R. (1995). *Stochastic Ordering and Dependence in Applied Probability.* Lecture Notes in Statistics, Volume **97**. Springer, New York.

[18] Tong, Y.L. (1990). *The Multivariate Normal Distribution*. Springer-Verlag, New York.

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