INSTITUT DE STATISTIQUE

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

DISCUSSION PAPER

0220

ACTUARIAL MODELLING OF LONGITUDINAL CLAIMS DATA THROUGH GAMM'S: SOME METHODOLOGICAL RESULTS

N. BROUHNS & M. DENUIT

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

ACTUARIAL MODELLING OF LONGITUDINAL CLAIMS DATA THROUGH GAMM'S: SOME METHODOLOGICAL RESULTS

Natacha Brouhns[*] & Michel Denuit[*‡]
[*]Institut de Statistique
Université Catholique de Louvain
Louvain-la-Neuve, Belgium
[‡]Institut de Sciences Actuarielles
Université Catholique de Louvain
Louvain-la-Neuve, Belgium

August 27, 2002

Abstract

This paper discusses the type of dependence induced by the Generalized Additive Mixed Model (GAMM) approach to regression analysis with correlated data. In this framework, random effects are added on the same scale as the fixed effects. Dependence between outcomes is thus generated by their sharing of common/correlated latent variables. In many cases, this results in strong positive association.

Key words and phrases: GAM, random effects, stochastic dominance, likelihood ratio order, supermodular order, positive orthant dependence, association, conditional increasingness, MTP_2, comonotonicity, credibility theory.

1 Introduction and Motivation

1.1 GAM's

Generalized Additive Models (GAM's) unify regression methods for a variety of discrete and continuous outcomes; a standard reference for GAM's is Hastie & Tibshirani (1990). The basis of GAM's is the assumption that the data are sampled from a one-parameter exponential family of distributions. This happens for instance with data conforming to binomial law, Poisson law, Normal law with a known variance or Gamma law with a known dispersion parameter. Specifically, consider a single observation y; GAM's admit a log-likelihood of the form

$$\ell(\theta, \phi; y) = \frac{y\theta - b(\theta)}{\phi} + c(y, \phi) \tag{1.1}$$

where θ denotes the canonical parameter and $\phi > 0$ is the dispersion parameter (assumed known). It is easily seen that

$$\mu = \mathbb{E}[Y] = b'(\theta) \text{ and } \mathbb{V}\mathrm{ar}[Y] = b''(\theta)\phi = v(\mu)\phi \tag{1.2}$$

where $v(\cdot)$ is called the variance function. A link function relating the mean μ to the linear predictor is then specified. More precisely, given a vector $(\boldsymbol{x}, \boldsymbol{w})$ of explanatory variables, where the components of $\boldsymbol{x} = (x_1, \ldots, x_p)^t$ are continuous and those of $\boldsymbol{w}$ are further binary covariates (coding categorical explanatory variables) and a vector $\boldsymbol{\beta}$ of regression coefficients, the additive predictor takes the form

$$\eta = \sum_{j=1}^{p} f_j(x_j) + \boldsymbol{w}^t\boldsymbol{\beta} \tag{1.3}$$

where $f_1(\cdot), \ldots, f_p(\cdot)$ are unknown smooth functions of the covariates. The score η is then related to the mean by $\eta = a(\mu)$. The function $a(\cdot)$ is called the link function and the special case $a(\mu) = \theta$ is called the canonical link function.

1.2 GAMM's

Several extensions of GAM's involve models with random terms in the linear predictor. Such Generalized Additive Mixed Models (GAMM's) are useful e.g. for accomodating the overdispersion for count data and for modeling the dependence in longitudinal studies or the correlation arising from covariates that are omitted or inadequately measured.

Given an unobserved vector of random effects ($\boldsymbol{\Lambda}_i = (\Lambda_{i1}, \Lambda_{i2} \ldots \Lambda_{in_i})$, say), the n_i observations $Y_{i1}, Y_{i2}, \ldots, Y_{in_i}$ relating to the ith subject are assumed to be conditionally independent with means that depend on the linear predictor through a specified link function and conditional variances that are specified by a variance function, and a scale factor, in the spirit of (1.1)-(1.2)-(1.3).

Assume now that m random vectors $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_m$ of respective dimensions $n_1, n_2, \ldots, n_m$ are observed. The $\boldsymbol{Y}_i$'s are mutually independent, but their components may be correlated (for instance because of repeated measures on the same individuals). GAMM's rely on the two following assumptions:

A1 Given $\boldsymbol{\Lambda}_i = \boldsymbol{\lambda}_i$ the responses Y_{it}, $t = 1, \ldots, n_i$, are mutually independent and admit a log-likelihood $\ell(\theta_{it}, \phi; y_{it})$ of the form (1.1) where

$$\mu_{it}(\lambda_{it}) = \mathbb{E}[Y_{it}|\Lambda_{it} = \lambda_{it}] \text{ and } v_{it} = \mathbb{V}\text{ar}[Y_{it}|\Lambda_{it} = \lambda_{it}]$$

satisfy

$$a(\mu_{it}(\lambda_{it})) = \eta_{it} + \lambda_{it} \text{ and } v_{it} = v(\mu_{it})\phi$$

where the score η_{it} is as in (1.3), $a(\cdot)$ and $v(\cdot)$ are known link and variance functions, respectively.

Henceforth, we restrict our study to the special case of the canonical link function, that is, $a(\mu_{it}(\lambda_{it})) = \theta_{it}$. This does not really restrict the generality of our results (most of them relying on the monotonicity of the function $a(\cdot)$) but greatly facilitates the exposition. Moreover, we treat η_{it} as a constant.

A2 The vectors of random effects $\boldsymbol{\Lambda}_1, \ldots, \boldsymbol{\Lambda}_m$ are mutually independent with a common underlying multivariate distribution.

Often, the $\boldsymbol{\Lambda}_i$'s are assumed to conform to the multivariate normal distribution. We restrict our attention to this important case in the present paper. Most results derived in this paper remain nevertheless valid for other choices of distributions.

1.3 GAMM's in actuarial science

There are numerous applications of the GAMM's in risk theory. Those techniques encompass most of the classical actuarial procedures. We will here describe a few typical examples, that we will use throughout the remainder of the paper.

Example 1.1. (Credibility with static random effects)

GAMM's are widely used by actuaries, since they form the basis of credibility theory, as shown by Nelder & Verrall (1997). In this context, Y_{it} represents the number or amount of claims reported to the insurance company by policyholder i in year t. Random effects $\boldsymbol{\Lambda}_i$ represent hidden features influencing the risk covered by the insurer.

The classical credibility model assumes the time-invariance of the random effects. More precisely, all the components of $\boldsymbol{\Lambda}_i$ are equal to some random variable Λ_i, that is, $\boldsymbol{\Lambda}_i = (\Lambda_i, \ldots, \Lambda_i)$. So, $\boldsymbol{\Lambda}_i$ has perfectly dependent components (i.e. $\boldsymbol{\Lambda}_i$ is comonotonic) and one could expect strong dependence between the Y_{it}'s for fixed i.

Example 1.2. (Credibility with dynamic random effects)

Recently, Pinquet, Guillén & Bolancé (2001), among others, allow for random effects that develop over time. This is justified since unobservable factors influencing the risk are not constant and 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 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.

In this paper, $\mathbf{\Lambda}_i$ will be taken multivariate normally distributed $\mathcal{N}(\boldsymbol{\mu}, \mathbf{\Sigma})$ with covariance matrix $\mathbf{\Sigma}$ featuring the temporal dependence. Classical ARMA models are often used for $\mathbf{\Lambda}_i$.

Example 1.3. (Credibility with AR1 random effects)
A particularly simple and efficient dynamic credibility model is obtained by assuming that $\mathbf{\Lambda}_i$ has an autoregressive structure of order one, that is,

$$\Lambda_{it} = \varrho\Lambda_{i,t-1} + \epsilon_{it}, \;\; t \geq 2,$$

where $\epsilon_{it} \sim \mathcal{N}(0, \sigma^2(1-\rho^2))$ are iid, $|\varrho| < 1$,, and $\epsilon_{i1} \sim \mathcal{N}(0, \sigma^2)$. In this model, the heterogeneity Λ_{it} for period t is influenced by the preceding period $\Lambda_{i,t-1}$ but has also its own characteristics ϵ_{it}.

Example 1.4. (Credibility with exchangeable random effects)
Another model (see e.g. Pinquet (2000)) postulates that there is a static baseline heterogeneity R_i for policyholder i which is perturbated by iid annual effects $S_{i1}, S_{i2}, \ldots$. Specifically, $\Lambda_{it} = R_i + S_{it}$ where the S_{it}'s are iid and independent from R_i, and they all conform to the Normal law. The vector $\mathbf{\Lambda}_i$ is exchangeable.

1.4 Aim of the work

This paper purposes to examine the kind of dependence generated by GAMM's. We believe that this will be particularly useful for applied modelling. Indeed, a better understanding of the features of GAMM's will enable actuaries to draw more appropriate conclusions from the model fit. This is especially true when predictions are made from conditional expectations (given past observations). The type of dependence between the components of $\boldsymbol{Y}_i$ will strongly affect the forecast.

In GAMM's, correlation among observations $Y_{i1}, Y_{i2}, \ldots, Y_{in_i}$ relating to subject i arises from their sharing unobservable correlated latent variables $\Lambda_{i1}, \Lambda_{i2}, \ldots, \Lambda_{in_i}$. It will be seen that the classical GAMM construction may entail strong positive dependence between the observations provided the outcomes are "increasing" in the random effects. This intuitive idea is formalized in a number of stochastic inequalities, each one enjoying a nice actuarial interpretation.

The theoretical results will be applied to the four examples listed in the preceding subsection; this will enhance their relevance for actuarial modelling.

1.5 Technical aspects

Important concepts in the analysis of dependence induced by GAMM's include likelihood ratio order together with multivariate total positivity, as well as stochastic dominance together with association. The main technical argument is the stochastic monotonicity of GAM's in their canonical parameter with respect to the likelihood ratio order. Intuitively speaking, this ensures that the outcome tends to become larger (in some sense) when the canonical parameter increases.

The present study expands on several premvious works. SHAKED & SPIZZICHINO (1998) established positive dependence properties for mixed models built from stochastically increasing parametric families of continuous distributions. PURCARU & DENUIT (2001, 2002) studied mixed Poisson models in an actuarial context and derived condition on the mixing parameter to generate positive dependence between Poisson counting outcomes. We show here that similar results apply to the whole GAMM family, because of stochastic monotonicity properties shared by all GAM's.

The major innovation of the paper is to apply the results to the credibility framework and to draw conclusions useful for the understanding of the models commonly used by actuaries. To the best of the authors' knowledge, credibility models have never been investigated from that point of view so far in the literature.

1.6 Agenda

In Section 2, we introduce various stochastic order relations (namely univariate and multivariate versions of stochastic dominance and likelihood ratio order), i.e. binary relations aiming to compare random variables or random vectors. These relations purpose to express for random variables/vectors the intuitive ideas of "being larger than". The main result of that section will be the stochastic monotonicity of the members of the GAM family with respect to the canonical parameter.

Section 3 is devoted to positive dependence concepts (namely, positive orthant dependence, association, conditional increasingness, multivariate total positivity and comonotonicity), i.e. ways to formalize the fact that large values for a component of a random vector tend to be associated with large values for the others.

The remaining sections contain the main results. Section 4 basically states that the structure of dependence exhibited by the components of $\mathbf{\Lambda}_i$ is transmitted to the components of $\boldsymbol{Y}_i$. Section 5 considers the a posteriori distribution of the vector $\mathbf{\Lambda}_i$ of random effects and proves its stochastic monotonicity in $\boldsymbol{Y}_i$. Section 6 intends to formalize the intuitive idea that increasing the strength of dependence between the components of $\mathbf{\Lambda}_i$ yields more positively dependent components of $\boldsymbol{Y}_i$. Section 7 deals with the predictive distributions. It aims to answer the following question: how do past claims influence future claims? We will see that in many cases, the future is increasing in the past, in a sense to be precised later. The final Section 8 concludes.

2 Stochastic increasingness of GAM's in the canonical parameter

Since the random effects have been added on the scale of the fixed effects, we must investigate whether an increase of the canonical parameter makes the GAM outcome "larger" or "smaller". In order to formalize this monotonicity, we need stochastic orderings; for more details about these probabilistic tools, we refer the reader e.g. to KAAS, VAN HEERWAARDEN & GOOVAERTS (1994), MÜLLER & STOYAN (2002) or SHAKED & SHANTHIKUMAR (1994).

2.1 Univariate likelihood ratio order

Let us recall the definition of the likelihood ratio order, and provide the reader with some intuitive interpretations of it. Given two random variables X and Y with respective (discrete or continuous) probability density functions f_X and f_Y, X is said to be smaller than Y in the likelihood ratio order, denoted as $X \preceq_{\text{lr}} Y$, if

$$f_X(u) f_Y(v) \geq f_X(v) f_Y(u) \text{ for all } u \leq v \in \mathbb{R}. \tag{2.1}$$

Considering (2.1), a ranking in the $\preceq_{\text{lr}}$-sense can be given the following nice interpretation. Provided X and Y are independent (which can be assumed without loss of generality), the left-hand side of (2.1) can be regarded as the likelihood of the event "X is small and Y is large" whereas the right-hand side of this relation reads "X is large and Y is small". Then, (2.1) expresses the fact that the latter event is less likely to occur than the first one.

2.2 Univariate stochastic dominance

Another intuitive interpretation for a ranking in the $\preceq_{\text{lr}}$-sense is as follows. Let $t \in \mathbb{R}$. Integrating (or summing in the discrete case) both sides of (2.1) over $u \in (-\infty, t]$ and $v \in (t, +\infty)$ yields

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

The latter inequality shows that Y is indeed "larger" than X since the probability for Y to be large (i.e. to exceed the treshold t) is bigger than the probability for X to be large. When $\Pr[X > t] \leq \Pr[Y > t]$ holds for all $t \in \mathbb{R}$, we will write $X \preceq_{\text{st}} Y$ and say that X is smaller than Y in the stochastic dominance. Clearly, $X \preceq_{\text{lr}} Y \Rightarrow X \preceq_{\text{st}} Y$. It can be shown that the latter implication is strict.

It is worth mentioning that since any non-decreasing function can be obtained as the uniform limit of a sequence of non-decreasing step functions, we also have $X \preceq_{\text{st}} Y \Leftrightarrow \mathbb{E}[\psi(X)] \leq \mathbb{E}[\psi(Y)]$ for any non-decreasing function ψ, provided the expectations exist.

2.3 Univariate $\preceq_{\text{lr}}$-increasingness of GAM's

The following property will be extremely useful in the remainder of our work. In the model **A1**-**A2**, all covariates being observed (fixed), it expresses the $\preceq_{\text{lr}}$-increasingness of the response in the random parameter. Henceforth, given a random variable X and an event A, we denote as $[X|A]$ a random variable with distribution function $x \mapsto \Pr[X \leq x|A]$.

Property 2.1. *In the model* ***A1-A2****,* $[Y_{it}|\Lambda_{it} = \lambda_{it}]$ *is increasing in* λ_{it} *in the* $\preceq_{lr}$*-sense, that is,*

$$\lambda_{it} \leq \lambda'_{it} \Rightarrow [Y_{it}|\Lambda_{it} = \lambda_{it}] \preceq_{lr} [Y_{it}|\Lambda_{it} = \lambda'_{it}] \text{ for any } i \text{ and } t.$$

Proof. Since $\eta_{it} + \lambda_{it} = \theta_{it}$, the announced implication is the same as

$$\theta_{it} \leq \theta'_{it} \Rightarrow [Y_{it}|\Theta_{it} = \theta_{it}] \preceq_{\mathrm{lr}} [Y_{it}|\Theta_{it} = \theta'_{it}]$$

where $\Theta_{it} = \eta_{it} + \Lambda_{it}$. The (discrete or continuous) probability density function of the response has the form:

$$f(y; \theta_{it}, \phi) = \exp\left(\frac{y\theta_{it} - b(\theta_{it})}{\phi} + c(y, \phi)\right)$$

Since the ratio

$$\frac{f(y; \theta_{it}, \phi)}{f(y; \theta'_{it}, \phi)} = \exp\left(\frac{y(\theta_{it} - \theta'_{it}) - (b(\theta_{it}) - b(\theta'_{it}))}{\phi}\right)$$

is clearly decreasing in y provided $\theta_{it} \leq \theta'_{it}$, we get the announced result from (2.1). □

2.4 Multivariate likelihood ratio order

The multivariate version of $\preceq_{\mathrm{lr}}$ is defined by extending (2.1) to joint densities. More precisely, given two n-dimensional vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$, with (discrete or continuous) probability density functions $f_{\boldsymbol{X}}$ and $f_{\boldsymbol{Y}}$, respectively, $\boldsymbol{X}$ is said to be smaller than $\boldsymbol{Y}$ in the likelihood ratio order, written as $\boldsymbol{X} \preceq_{\mathrm{lr}} \boldsymbol{Y}$, if

$$f_{\boldsymbol{X}}(\boldsymbol{x}) f_{\boldsymbol{Y}}(\boldsymbol{y}) \leq f_{\boldsymbol{X}}(\boldsymbol{x} \wedge \boldsymbol{y}) f_{\boldsymbol{Y}}(\boldsymbol{x} \vee \boldsymbol{y}), \tag{2.2}$$

where $\vee$ and $\wedge$ denote, respectively, the componentwise maximum and minimum. The inequality in (2.2) defining multivariate $\preceq_{\mathrm{lr}}$ can be interpreted as (2.1).

2.5 Multivariate stochastic dominance

In case we want to compare random vectors, an intuitively acceptable strategy consists in transforming those vectors into random variables using increasing mappings, and then to compare the resulting outcomes with univariate stochastic orderings. This yields the following definition for multivariate $\preceq_{\mathrm{st}}$: given two random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$, $\boldsymbol{X}$ is said to be smaller than $\boldsymbol{Y}$ in the stochastic dominance, written as $\boldsymbol{X} \preceq_{\mathrm{st}} \boldsymbol{Y}$, if $\psi(\boldsymbol{X}) \preceq_{\mathrm{st}} \psi(\boldsymbol{Y})$ for every non-decreasing function $\psi : \mathbb{R}^n \to \mathbb{R}$. It can be shown that, as in the univariate case, $\boldsymbol{X} \preceq_{\mathrm{lr}} \boldsymbol{Y} \Rightarrow \boldsymbol{X} \preceq_{\mathrm{st}} \boldsymbol{Y}$.

2.6 Multivariate $\preceq_{\mathrm{lr}}$-increasingness of GAM's

The next property is the multivariate counterpart of Property 2.1. It applies in particular to GAMM's. Henceforth, an inequality between two real vectors has to be interpreted componentwise.

Property 2.2. *In the model **A1**-**A2**, $[\boldsymbol{Y}_i|\boldsymbol{\Lambda}_i = \boldsymbol{\lambda}_i]$ is increasing in $\boldsymbol{\lambda}_i$ in the $\preceq_{lr}$-sense, that is,*

$$\boldsymbol{\lambda}_i \leq \boldsymbol{\lambda}'_i \Rightarrow [\boldsymbol{Y}_i|\boldsymbol{\Lambda}_i = \boldsymbol{\lambda}_i] \preceq_{lr} [\boldsymbol{Y}_i|\boldsymbol{\Lambda}_i = \boldsymbol{\lambda}'_i].$$

Proof. Under **A1-A2**, $[\boldsymbol{Y}_i|\boldsymbol{\Lambda}_i=\boldsymbol{\lambda}_i]$ is a random vector with independent components. Moreover, by virtue of Property 2.1, the stochastic inequalities

$$\begin{array}{rcl}
[Y_{i1}|\Lambda_{i1}=\lambda_{i1}] & \preceq_{\mathrm{lr}} & [Y_{i1}|\Lambda_{i1}=\lambda'_{i1}] \\
[Y_{i2}|\Lambda_{i2}=\lambda_{i2}] & \preceq_{\mathrm{lr}} & [Y_{i2}|\Lambda_{i2}=\lambda'_{i2}] \\
\vdots & \vdots & \vdots \\
[Y_{in_i}|\Lambda_{in_i}=\lambda_{in_i}] & \preceq_{\mathrm{lr}} & [Y_{in_i}|\Lambda_{in_i}=\lambda'_{in_i}]
\end{array}$$

hold true for any $\lambda_{i1}\leq\lambda'_{i1}$, $\lambda_{i2}\leq\lambda'_{i2}$, ... $\lambda_{in_i}\leq\lambda'_{in_i}$. Since a comparison in the multivariate $\preceq_{\mathrm{lr}}$-sense reduces to a componentwise $\preceq_{\mathrm{lr}}$-ranking when both random vectors have independent component (this is an immediate consequence of (2.2) when the joint probability density functions factor in the product of the marginal probability density functions), we get that $[\boldsymbol{Y}_i|\boldsymbol{\Lambda}_i=\boldsymbol{\lambda}_i]\preceq_{\mathrm{lr}}[\boldsymbol{Y}_i|\boldsymbol{\Lambda}_i=\boldsymbol{\lambda}'_i]$ holds true provided $\boldsymbol{\lambda}_i\leq\boldsymbol{\lambda}'_i$, as announced. □

3 Positive dependence concepts

There are many ways to express that the components of a random vector are positively dependent. Most often, the purpose is to formalize the fact that large values of one component tend to be associated with large values of the others. In this section, we briefly review the most relevant dependence notions, from the weakest to the strongest. For more details, the reader is referred e.g. to Müller & Stoyan (2002).

3.1 Positive orthant dependence

Random variables $X_1, X_2, \ldots, X_n$ (or the random vector $\boldsymbol{X}$) are said to be positively orthant dependent (POD, in short) when the inequalities

$$\Pr[\boldsymbol{X}\leq x]\geq\prod_{i=1}^{n}\Pr[X_i\leq x_i] \text{ and } \Pr[\boldsymbol{X}>x]\geq\prod_{i=1}^{n}\Pr[X_i>x_i] \tag{3.1}$$

simultaneously hold for any $\boldsymbol{x}\in\mathbb{R}^n$. Intuitively, (3.1) means that $X_1, X_2, \ldots, X_n$ are more likely simultaneously to have small/large values, compared with a vector of independent random variables with the same corresponding univariate marginals. It is worth mentioning that the inequalities (3.1) are usually referred to in Statistics as the Sidak inequalities, or as the first order product-type inequalities.

3.2 Association

Random variables $X_1, X_2, \ldots, X_n$ (or the random vector $\boldsymbol{X}$) are said to be associated when

$$\mathbb{C}\mathrm{ov}\left[\psi_1(\boldsymbol{X}),\psi_2(\boldsymbol{X})\right]\geq 0 \tag{3.2}$$

for all non-decreasing functions ψ_1 and $\psi_2:\mathbb{R}^n\to\mathbb{R}$ for which the covariances exist. It can be shown that $\boldsymbol{X}$ associated $\Rightarrow\boldsymbol{X}$ POD.

The abstract characterization (3.2) of association is difficult to interpret, especially in a concrete statistical model. The main interest of association is that it is usually easy to be established and that it implies the weaker POD, whose interpretation is clear.

3.3 Conditional increasingness in sequence

Random variables $X_1, X_2, \ldots, X_n$ (or the random vector $\boldsymbol{X}$) are said to be conditionally increasing in sequence (CIS, in short) if, for any $i = 2, 3, \ldots, n$, the following conditions hold:

$$[X_i | X_1 = x_1, X_2 = x_2, \ldots, X_{i-1} = x_{i-1}] \preceq_{\mathrm{st}} [X_i | X_1 = x_1', X_2 = x_2', \ldots, X_{i-1} = x_{i-1}']$$

for any $x_1 \leq x_1'$, $x_2 \leq x_2'$, $\ldots$, $x_{i-1} \leq x_{i-1}'$ in the support of the X_i's.

This dependence notion is particularly useful when the components of the random vector are naturally ordered (as in repeated measures, for instance). In this case, CIS expresses the monotonicity of the future given the past.

It can be shown that CIS is a stronger dependence notion than association, i.e. $\boldsymbol{X}$ CIS $\Rightarrow$ $\boldsymbol{X}$ associated.

3.4 Multivariate total positivity

A random vector $\boldsymbol{X}$ with (discrete or continuous) probability density function $f_{\boldsymbol{X}}$ is MTP$_2$ if $f_{\boldsymbol{X}}$ is an MTP$_2$ function, that is, if the inequality

$$f_{\boldsymbol{X}}(\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{y}) \leq f_{\boldsymbol{X}}(\boldsymbol{x} \vee \boldsymbol{y}) f_{\boldsymbol{X}}(\boldsymbol{x} \wedge \boldsymbol{y}) \tag{3.3}$$

holds for any $\boldsymbol{x}$, $\boldsymbol{y}$ in $\mathbb{R}^n$. In the case of a bivariate density function, MTP$_2$ reduces to the standard TP$_2$.

The following result, due to Kemperman (1977), will play an important role in the remainder of the work. It basically says that MTP$_2$ reduces to TP$_2$ in pairs, making MTP$_2$ much easier to establish.

Characterization 3.1. *Suppose the support of $\boldsymbol{X}$ is a lattice (that is, if $\boldsymbol{x}$ and $\boldsymbol{y}$ are in the support of $\boldsymbol{X}$, then so are $\boldsymbol{x} \vee \boldsymbol{y}$ and $\boldsymbol{x} \wedge \boldsymbol{y}$). Then $\boldsymbol{X}$ is MTP$_2$ if, and only if, its (continuous or discrete) probability density function $f_{\boldsymbol{X}}$ is TP$_2$ in each pair of its variables when the other $n-2$ variables are held fixed, that is*

$$\begin{aligned} & f_{\boldsymbol{X}}(x_1, \ldots, x_i, \ldots, x_j, \ldots, x_n) f_{\boldsymbol{X}}(x_1, \ldots, x_i', \ldots, x_j', \ldots, x_n) \\ \geq \; & f_{\boldsymbol{X}}(x_1, \ldots, x_i', \ldots, x_j, \ldots, x_n) f_{\boldsymbol{X}}(x_1, \ldots, x_i, \ldots, x_j', \ldots, x_n) \end{aligned}$$

whenever $x_i \leq x_i'$ and $x_j \leq x_j'$, x_k, $k \neq i, j$ held fixed.

The implication $\boldsymbol{X}$ MTP$_2$ $\Rightarrow$ $\boldsymbol{X}$ CIS holds true.

3.5 Comonotonicity

Random variables $X_1, X_2, \ldots, X_n$ (or the random vector $\boldsymbol{X}$) are said to be comonotonic when there exists a random variable Z and non-decreasing functions $\psi_1, \psi_2, \ldots, \psi_n : \mathbb{R} \to \mathbb{R}$ such that $\boldsymbol{X}$ is distributed as $(\psi_1(Z), \psi_2(Z), \ldots, \psi_n(Z))$. Comonotonicity can thus be regarded as perfect positive dependence: increasing any of the X_i's makes the other X_j's larger.

It can be established that $\boldsymbol{X}$ comonotonic $\Rightarrow$ $\boldsymbol{X}$ MTP$_2$.

4 Dependence in GAMM's

Our aim is to show that in many cases, GAMM's induce positive dependence between the Y_{it}'s, in the sense that "large" (or "small") values of the random variables tend to occur together. This is formally stated in the next result, which uses the increasingness of the conditional distribution in the mixing parameter established in Property 2.1.

Proposition 4.1. *In the model* ***A1-A2****,*

(i) $\boldsymbol{\Lambda}_i$ *associated* $\Rightarrow$ $\boldsymbol{Y}_i$ *associated;*

(ii) $\boldsymbol{\Lambda}_i$ *MTP*$_2$ $\Rightarrow$ $\boldsymbol{Y}_i$ *MTP*$_2$.

Proof. (i) From Property 2.2, we know that $[\boldsymbol{Y}_i|\boldsymbol{\Lambda}_i=\boldsymbol{\lambda}_i] \preceq_{\rm st} [\boldsymbol{Y}_i|\boldsymbol{\Lambda}_i=\boldsymbol{\lambda}_i']$ whenever $\boldsymbol{\lambda}_i \leq \boldsymbol{\lambda}_i'$. This is equivalent to

$$\mathbb{E}[\psi(\boldsymbol{Y}_i)|\boldsymbol{\Lambda}_i=\boldsymbol{\lambda}_i] \text{ being nondecreasing in } \boldsymbol{\lambda}_i \text{ for every nondecreasing } \psi. \tag{4.1}$$

Now, we have to establish that $\boldsymbol{Y}_i$ is associated, that is, $\mathbb{C}\text{ov}[\psi_1(\boldsymbol{Y}_i),\psi_2(\boldsymbol{Y}_i)] \geq 0$ for every pair of functions ψ_1 and ψ_2 wich are nondecreasing in each of their arguments. Reformulating the covariance yields

$$\begin{aligned}\mathbb{C}\text{ov}[\psi_1(\boldsymbol{Y}_i),\psi_2(\boldsymbol{Y}_i)] &= \mathbb{E}\Big[\mathbb{C}\text{ov}[\psi_1(\boldsymbol{Y}_i),\psi_2(\boldsymbol{Y}_i)|\boldsymbol{\Lambda}_i]\Big] \\ &\quad +\mathbb{C}\text{ov}\Big[\mathbb{E}[\psi_1(\boldsymbol{Y}_i)|\boldsymbol{\Lambda}_i],\mathbb{E}[\psi_2(\boldsymbol{Y}_i)|\boldsymbol{\Lambda}_i]\Big].\end{aligned}$$

Given $\boldsymbol{\Lambda}_i=\boldsymbol{\lambda}_i$, the components of $\boldsymbol{Y}_i$ are independent and the first term is positive (since independent random variables are associated, as shown in Barlow & Proschan (1975, Theorem 2.2, p31)). The second term is positive according to (4.1), so that $\boldsymbol{Y}_i$ is indeed associated, as announced.

(ii) We will show that the density of $\boldsymbol{Y}_i$ is TP$_2$ in pairs, whence the announced implication then results from Characterization 3.1. We show it for components s and t. According to Property 2.1, we have that the (discrete or continuous) probability density function $f_{is}(y|\lambda_{is})$ of $[Y_{is}|\Lambda_{is}=\lambda_{is}]$ and $f_{it}(y|\lambda_{it})$ of $[Y_{it}|\Lambda_{it}=\lambda_{it}]$ are TP$_2$ in (y,λ_{is}) and in (y,λ_{it}), respectively. Since $f_{is}(y|\lambda_{is})=f_{is}(y|\boldsymbol{\lambda}_i)$ and $f_{it}(y|\lambda_{it})=f_{it}(y|\boldsymbol{\lambda}_i)$, both $f_{is}(y|\boldsymbol{\lambda}_i)$ and $f_{it}(y|\boldsymbol{\lambda}_i)$ are MTP$_2$ in $(y,\boldsymbol{\lambda}_i)$. Since $\boldsymbol{\Lambda}_i$ is assumed to be MTP$_2$, its joint probability density function $g(\boldsymbol{\lambda}_i)$ is known to be MTP$_2$. Now, fixing the values of y_{ij}, $j\neq s,t$, the function

$$\begin{aligned}(y_{is},y_{it}) \mapsto f(\boldsymbol{y}_i) &= \int_{\boldsymbol{\lambda}_i\in\mathbb{R}^{n_i}} \left\{\prod_{j=1}^{n_i} f_{ij}(y_{ij}|\boldsymbol{\lambda}_i)\right\} g(\boldsymbol{\lambda}_i)d\boldsymbol{\lambda}_i \\ &= \int_{\boldsymbol{\lambda}_i\in\mathbb{R}^{n_i}} f_{is}(y_{is}|\boldsymbol{\lambda}_i)f_{it}(y_{it}|\boldsymbol{\lambda}_i)\left\{\prod_{j\neq s,t} f_{ij}(y_{ij}|\boldsymbol{\lambda}_i)\right\} g(\boldsymbol{\lambda}_i)d\boldsymbol{\lambda}_i\end{aligned}$$

is TP$_2$ in (y_{is},y_{it}) invoking (1.14) and (1.15) in Karlin & Rinott (1980). □

Let us now illustrate the usefulness of the results of Proposition 4.1 in actuarial problems.

Example 4.2. (Credibility with static random effects)

If $\mathbf{\Lambda}_i = (\Lambda_i, \ldots, \Lambda_i)$ then $\mathbf{\Lambda}_i$ is MTP_2 (since it is comonotonic), and so is $\boldsymbol{Y}_i$, invoking Proposition 4.1(ii). This first example shows that the dependence among the Y_{it}'s in the classical credibility model is very strong, though not perfect. Broadly speaking, the dependence between the observed outcomes $Y_{i1}, \ldots, Y_{in_i}$ is always weaker than the dependence between the hidden random effects $\Lambda_{i1}, \ldots, \Lambda_{in_i}$. In the present case, comonotonicity is transformed into MTP_2.

Example 4.3. (Credibility with dynamic random effects)

Assume that $\mathbf{\Lambda}_i$ is multivariate normal with covariance matrix $\mathbf{\Sigma} = \{\sigma_{st}\}$. Pitt (1982) showed that the condition $\sigma_{st} \geq 0$ for all s, t is necessary and sufficient for $\mathbf{\Lambda}_i$ to be associated. In this case, Proposition 4.1(i) ensures that $\boldsymbol{Y}_i$ is also associated.

Further, Tong (1990) established that provided $\mathbf{\Sigma}$ is invertible (denote as $\boldsymbol{R}$ its inverse, i.e. $\boldsymbol{R} = \{r_{st}\} = \mathbf{\Sigma}^{-1}$), $\mathbf{\Lambda}_i$ is MTP_2 (and hence $\boldsymbol{Y}_i$ from Proposition 4.1(ii)) if, and only if, $r_{st} \leq 0$ for all $s \neq t$.

Example 4.4. (Credibility with AR1 random effects)

Since,

$$\mathbb{C}\mathrm{ov}[\Lambda_{is}, \Lambda_{it}] = \sigma^2 \varrho^{|s-t|}$$

the elements σ_{st} of $\mathbf{\Sigma}$ are given by:

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

Therefore the off-diagonal elements of the matrix $\boldsymbol{R} = \mathbf{\Sigma}^{-1}$ are

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

and are all non positive when $\varrho \geq 0$. Hence $\varrho \geq 0 \Rightarrow \mathbf{\Lambda}_i$ MTP_2 which in turn ensures that $\boldsymbol{Y}_i$ is MTP_2 by virtue of Proposition 4.1(ii).

Example 4.5. (Credibility with exchangeable random effects)

Note that the normal distribution has a log-concave density (i.e. the logarithm of the normal probability density function is concave). Karlin & Rinott (1980, Proposition 3.8) showed that, provided each S_{it} is governed by a log-concave density function then $\mathbf{\Lambda}_i$ (where $\Lambda_{it} = R_i + S_{it}$) is MTP_2. In this case also $\boldsymbol{Y}_i$ is MTP_2 in application of Proposition 4.1(ii).

5 A posteriori distribution of the random effects

Let us now consider the posterior distribution of $\mathbf{\Lambda}_i$ given the observations $\boldsymbol{Y}_i = \boldsymbol{y}_i$, denoted as $g(\cdot|\boldsymbol{y}_i)$ (whereas $g(\cdot)$ denotes the a priori probability density function of $\mathbf{\Lambda}_i$). Let us prove the following result, which is in the vein of Whitt (1979) and Fahmy et al. (1982). It basically states that observing large outcomes $\boldsymbol{Y}_i$ increases unobservable latent variables (in the $\preceq_{\mathrm{lr}}$-sense).

Proposition 5.1. *In the model* ***A1-A2****,* $[\mathbf{\Lambda}_i|\boldsymbol{Y}_i = \boldsymbol{y}_i] \preceq_{lr} [\mathbf{\Lambda}_i|\boldsymbol{Y}_i = \boldsymbol{y}'_i]$ *whenever* $\boldsymbol{y}_i \leq \boldsymbol{y}'_i$.

Proof. We have that

$$
\begin{aligned}
f(\boldsymbol{y}_i|\boldsymbol{\lambda}_i) &= \prod_{j=1}^{n_i} f_{ij}(y_{ij}|\lambda_{ij}) \\
&= \prod_{j=1}^{n_i} \exp\left(\frac{y_{ij}\theta_{ij} - b(\theta_{ij})}{\phi}\right) \exp\left(c(y_{ij},\phi)\right) \\
&= \left\{\prod_{j=1}^{n_i} \exp\left(c(y_{ij},\phi)\right) \exp\left(\frac{y_{ij}\eta_{ij}}{\phi}\right)\right\} \left\{\prod_{j=1}^{n_i} \exp\left(\frac{-b(\lambda_{ij}+\eta_{ij})}{\phi}\right)\right\} \left\{\prod_{j=1}^{n_i} \exp\left(\frac{y_{ij}\lambda_{ij}}{\phi}\right)\right\} \\
&\equiv h_1(\boldsymbol{y}_i) h_2(\boldsymbol{\lambda}_i) \prod_{j=1}^{n_i} \exp\left(\frac{y_{ij}\lambda_{ij}}{\phi}\right).
\end{aligned}
$$

Hence, we can express the posterior probability density function of $\boldsymbol{\Lambda}_i$ given $\boldsymbol{Y}_i = \boldsymbol{y}_i$ as

$$
\begin{aligned}
g(\boldsymbol{\lambda}_i|\boldsymbol{y}_i) &= \frac{f(\boldsymbol{y}_i|\boldsymbol{\lambda}_i)g(\boldsymbol{\lambda}_i)}{f(\boldsymbol{y}_i)} \\
&= \frac{h_1(\boldsymbol{y}_i)h_2(\boldsymbol{\lambda}_i)\left\{\prod_{j=1}^{n_i}\exp\left(\frac{y_{ij}\lambda_{ij}}{\phi}\right)\right\} g(\boldsymbol{\lambda}_i)}{f(\boldsymbol{y}_i)}.
\end{aligned}
$$

Now, we have to show that for $\boldsymbol{y}_i \leq \boldsymbol{y}_i'$, the inequality

$$
g(\boldsymbol{\lambda}_i|\boldsymbol{y}_i)g(\boldsymbol{\lambda}_i'|\boldsymbol{y}_i') \leq g(\boldsymbol{\lambda}_i' \wedge \boldsymbol{\lambda}_i|\boldsymbol{y}_i)g(\boldsymbol{\lambda}_i' \vee \boldsymbol{\lambda}_i|\boldsymbol{y}_i')
$$

holds true, which immediately follows from the TP_2 property of the function

$$
(y_{ij},\lambda_{ij}) \mapsto \exp\left(\frac{y_{ij}\lambda_{ij}}{\phi}\right)
$$

together with Characterization 3.1. The proof is now complete. □

In a Bayesian framework, it is common to predict $\boldsymbol{\Lambda}_i$ by means of the posterior mean, that is

$$
\widehat{\boldsymbol{\Lambda}_i}(\boldsymbol{y}_i) = \mathbb{E}[\boldsymbol{\Lambda}_i|\boldsymbol{Y}_i = \boldsymbol{y}_i].
$$

Proposition 5.1 ensures that $\widehat{\boldsymbol{\Lambda}_i}(\boldsymbol{y}_i) \leq \widehat{\boldsymbol{\Lambda}_i}(\boldsymbol{y}_i')$ whenever $\boldsymbol{y}_i \leq \boldsymbol{y}_i'$.

6 More positively dependent Λ_{it}'s induce more positively dependent Y_{it}'s

It seems natural to expect that increasing the strength of the positive dependence between the latent Λ_{it}'s will induce more association between the observed outcomes Y_{it}'s. This section precisely aims to formalize this intuitive idea. For this purpose, we resort to the supermodular ordering (see e.g. Shaked & Shanthikumar (1997) for a general presentation and Bäuerle & Müller (1998) for applications in actuarial science), which will be seen to be an appropriate tool to compare the strength of dependence.

6.1 Supermodular order

The supermodular order is based on the comparison of expectations of supermodular functions. Let us recall that a real-valued function $\psi : \mathbb{R}^n \to \mathbb{R}$ is called supermodular if

$$\psi(\boldsymbol{x} \vee \boldsymbol{y}) + \psi(\boldsymbol{x} \wedge \boldsymbol{y}) \geq \psi(\boldsymbol{x}) + \psi(\boldsymbol{y}), \tag{6.1}$$

for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$. It is interesting to contrast the inequality (6.1) defining supermodularity with the inequality (3.3) defining multivariate total positivity; this reveals that supermodularity is a kind of log-MTP$_2$ notion.

Then, given two random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$, $\boldsymbol{X}$ is said to precede $\boldsymbol{Y}$ in the supermodular order, denoted as $\boldsymbol{X} \preceq_{\text{sm}} \boldsymbol{Y}$, if $\mathbb{E}[\psi(\boldsymbol{X})] \leq \mathbb{E}[\psi(\boldsymbol{Y})]$ for all supermodular function ψ for which the expectations exist.

Since the functions $\boldsymbol{y} \mapsto \mathbb{I}[\boldsymbol{y} > \boldsymbol{x}]$ and $\boldsymbol{y} \mapsto \mathbb{I}[\boldsymbol{y} \leq \boldsymbol{x}]$ are supermodular for each fixed $\boldsymbol{x}$, it is immediate that

$$\boldsymbol{X} \preceq_{\text{sm}} \boldsymbol{Y} \Rightarrow \Pr[\boldsymbol{X} > \boldsymbol{x}] \leq \Pr[\boldsymbol{Y} > \boldsymbol{x}] \text{ and } \Pr[\boldsymbol{X} \leq \boldsymbol{x}] \leq \Pr[\boldsymbol{Y} \leq \boldsymbol{x}]. \tag{6.2}$$

This expresses well the fact that the components of $\boldsymbol{Y}$ are more POD than those of $\boldsymbol{X}$. Note that from (6.2) it follows that if $\boldsymbol{X} \preceq_{\text{sm}} \boldsymbol{Y}$ then X_i and Y_i are identically distributed for $i = 1, 2, \ldots, n$. Therefore, if $\boldsymbol{X} \preceq_{\text{sm}} \boldsymbol{Y}$ then $\boldsymbol{X}$ and $\boldsymbol{Y}$ have necessarily the same univariate marginals.

Supermodular ordering turns out to be a useful tool for comparing dependence structures of random vectors. Indeed, since all functions $\psi(\boldsymbol{x}) = x_i x_j$ for $i \neq j$ are supermodular. Hence,

$$\boldsymbol{X} \preceq_{\text{sm}} \boldsymbol{Y} \Rightarrow \mathbb{C}\text{ov}[X_i, X_j] \leq \mathbb{C}\text{ov}[Y_i, Y_j] \text{ for any } i \neq j.$$

Further, the values of the Pearson's, Spearman's and Kendall's correlation coefficients are always larger for $\boldsymbol{Y}$ than for $\boldsymbol{X}$.

6.2 Ordering the strength of dependence

We are now in a position to state the following result, which formalizes our intuition.

Proposition 6.1. *In the model* ***A1-A2****,* $\boldsymbol{\Lambda}_i \preceq_{sm} \boldsymbol{\Lambda}_i' \Rightarrow \boldsymbol{Y}_i \preceq_{sm} \boldsymbol{Y}_i'$.

Proof. The result straightly follows from Property 2.1 together with Theorem 4.1(a) in Denuit & Müller (2002). ☐

Coming back to the interpretation of $\preceq_{\text{sm}}$ as a positive dependence order, the statement in Proposition 6.1 reads "more positively dependent Λ_{it}'s yield more positively dependent Y_{it}'s", as expected.

We apply the result of Proposition 6.1 in the two examples where it makes sense (obviously, two vectors of static random effects cannot be ordered in the $\preceq_{\text{sm}}$-sense, and the same remark applies to the exchangeable random effects, because only distributions with the same marginals can be compared in the supermodular sense).

Example 6.2. (Credibility with dynamic random effects)

Assume that $\mathbf{\Lambda}_i$ and $\mathbf{\Lambda}'_i$ both conform to the multivariate normal distribution, with identical univariate marginals and $\mathbb{C}\text{ov}[\Lambda_{is}, \Lambda_{it}] \leq \mathbb{C}\text{ov}[\Lambda'_{is}, \Lambda'_{it}]$ for all $s \neq t$. Then, Theorem 3.13.5 of Müller & Stoyan (2002) ensures that $\mathbf{\Lambda}_i \preceq_{\text{sm}} \mathbf{\Lambda}'_i$ and Proposition 6.1 yields in turn $\boldsymbol{Y}_i \preceq_{\text{sm}} \boldsymbol{Y}'_i$.

The dynamic credibility model thus conforms to the intuition: increasing the correlation between each pair of random effects makes the annual claims more dependent.

Example 6.3. (Credibility with AR1 random effects)

Since $\mathbb{C}\text{ov}[\Lambda_{is}, \Lambda_{it}] = \sigma^2 \varrho^{|s-t|}$, Proposition 6.1 together with the preceding example ensures that $\varrho \leq \varrho' \Rightarrow \boldsymbol{Y}_i \preceq_{\text{sm}} \boldsymbol{Y}'_i$. In this model, the amount of dependence is thus controlled by the parameter ϱ: a large value of ϱ will increase the importance of past claims history in the determination of future premiums.

7 Predictive distributions

Before studying predictive distributions, let us first examine the monotonicity of a subset of $\boldsymbol{Y}_i$, given the other components.

Proposition 7.1. *Let $\boldsymbol{Y}_{iJ}$ (resp. $\boldsymbol{Y}_{iK}$) be the random vector with components Y_{ij}, $j \in J$ (resp. Y_{ik}, $k \in K$). In the model* ***A1-A2****, if $\mathbf{\Lambda}_i$ is MTP$_2$ then $[\boldsymbol{Y}_{iJ} | \boldsymbol{Y}_{iK} = \boldsymbol{y}_K] \preceq_{lr} [\boldsymbol{Y}_{iJ} | \boldsymbol{Y}_{iK} = \boldsymbol{y}'_K]$ for any $\boldsymbol{y}_K \leq \boldsymbol{y}'_K \in \mathbb{R}^{\#K}$, for any partition of $\{1, 2, \ldots, n_i\}$ in J and K.*

Proof. Considering Proposition 4.1(ii), we know that $\boldsymbol{Y}_i$ is MTP$_2$. Let us denote as $f(\boldsymbol{y}_J | \boldsymbol{y}_K)$ the conditional probability density function of $\boldsymbol{Y}_{iJ}$ given $\boldsymbol{Y}_{iK} = \boldsymbol{y}_K$. We have to prove that

$$f(\boldsymbol{y}_J \wedge \widetilde{\boldsymbol{y}}_J | \boldsymbol{y}_K) f(\boldsymbol{y}_J \vee \widetilde{\boldsymbol{y}}_J | \boldsymbol{y}'_K) \geq f(\boldsymbol{y}_J | \boldsymbol{y}_K) f(\widetilde{\boldsymbol{y}}_J | \boldsymbol{y}'_K)$$

holds for any $\boldsymbol{y}_J, \widetilde{\boldsymbol{y}}_J \in \mathbb{R}^{\#J}$ provided $\boldsymbol{y}'_K \leq \boldsymbol{y}_K$. Since the joint probability density function f of $\boldsymbol{Y}_i$ is MTP$_2$, we know that

$$f(\boldsymbol{y}_J \wedge \widetilde{\boldsymbol{y}}_J, \boldsymbol{y}_K) f(\boldsymbol{y}_J \vee \widetilde{\boldsymbol{y}}_J, \boldsymbol{y}'_K) \geq f(\boldsymbol{y}_J, \boldsymbol{y}_K) f(\widetilde{\boldsymbol{y}}_J, \boldsymbol{y}'_K)$$

holds true, whence the desired inequality follows by dividing each side by $f_K(\boldsymbol{y}_K) f_K(\boldsymbol{y}'_K)$, where f_K is the probability density function of $\boldsymbol{Y}_{iK}$. □

The intuitive explanation behind the results stated in Proposition 7.1 is clear: when the components of $\mathbf{\Lambda}_i$ exhibit strong positive dependence (MTP$_2$), observing large outcomes for some of the Y_{ij}'s (those in K) makes the others (those in J) larger (in the $\preceq_{\text{lr}}$-sense).

Let us now apply these results to predictive distributions. The very aim of credibility theory is indeed to predict future claim behaviour. In that respect, predictive distributions are of prime interest: these are the distributions of claim characteristics for next year, given past observations.

Example 7.2. (Credibility with static random effects)

Since $\mathbf{\Lambda}_i$ is MTP$_2$, we have that $[Y_{in_i+1} | \boldsymbol{Y}_i = \boldsymbol{y}_i] \preceq_{\text{lr}} [Y_{in_i+1} | \boldsymbol{Y}_i = \boldsymbol{y}'_i]$ whenever $\boldsymbol{y}_i \leq \boldsymbol{y}'_i$. This indicates that a bad claim record is a worse sign for the future than a good one, whatever the distributions of the Λ_i's.

Example 7.3. (Credibility with dynamic random effects)

Provided the covariance matrix $\mathbf{\Sigma}$ fulfills the condition of Example 4.3 that ensures $(\Lambda_{i1},\ldots,\Lambda_{in_i},\Lambda_{in_i+1})$ to be MTP_2, we reach the same conclusion than for static credibility models. Again, future claims Y_{in_i+1} are increasing in the past claims $\boldsymbol{Y}_i$ in the $\preceq_{\text{lr}}$-sense.

Example 7.4. (Credibility with AR1 random effects)

Provided $\varrho \geq 0$, $\mathbf{\Lambda}_i$ is MTP_2 so that Y_{in_i+1} increases in $\boldsymbol{Y}_i$ in the $\preceq_{\text{lr}}$-sense.

Example 7.5. (Credibility with exchangeable random effects)

Since $\mathbf{\Lambda}_i$ is always MTP_2 in that case, the increasingness of the future given the past applies to this situation.

8 Conclusions

This paper studies the kind of dependence induced by the introduction of random effects. A prime example of such a construction in actuarial science is of course credibility theory, to which the paper is focussed. Other possible applications include ratemaking by geographical area in the framework of Boskov & Verrall (1994).

The main message addressed to practicians is that in the linear exponential family of distributions (with loglikelihood of the form (1.1)) most credibility systems will behave as intuitively expected, that is

1. increasing the past claims will make the unobservable random effects "larger", making the policyholders more dangerous on the unobservable characteristics;
2. increasing the past claims will increase the future claims;
3. increasing the temporal dependence (for instance by increasing the number of missing explanatory variables) will increase the dependence between past and future claims, so that the a posteriori corrections induced by the credibility model will be more severe.

This paper makes these points clear using stochastic orderings and dependence notions.

Acknowledgements

Both authors warmly thank the "Fonds Spéciaux de Recherche" of the Université catholique de Louvain, Louvain-la-Neuve, for financial support. They have benefited from stimulating discussions with Professor Philippe Lambert and Professor Moshe Shaked.

References

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

[2] Bäuerle, N., & Müller, A. (1998). Modeling and comparing dependencies in multivariate risk portfolios. *ASTIN Bulletin* **28**, 59-76.

[3] Boskov, M. & Verrall, R.J. (1994). Premium rating by geographical area using spatial models. *ASTIN Bulletin* **24**, 131-143.

[4] Denuit, M., & Müller, A. (2000). Smooth generators of integral stochastic orders. *Annals of Applied Probability*, to appear.

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

[6] Hastie, T., & Tibshirani, R. (1990). *Generalized Additive Models*. Chapman & Hall, London.

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

[8] Karlin, S., & Rinott, Y. (1980). Classes of orderings of measures and related correlation inequalities. I. Multivariate totally positive distributions. *Journal of Multivariate Analysis* **10**, 467-498.

[9] Kemperman J.H. (1977). On the FKG-Inequality for measures on a partially ordered space. *Indagationes Mathematicae* **39**, 311-331.

[10] Müller, A., & Stoyan, D. (2002). *Comparison Methods for Stochastic Models and Risks*. Wiley, New York.

[11] Nelder, J.A., & Verrall, R.J. (1997). Credibility theory and generalized linear models. *ASTIN Bulletin* **27**, 71-82.

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

[13] Pinquet, J., Guillen, M., and Bolancé, C. (2001). Allowance for the age of claims in bonus-malus systems. *ASTIN Bulletin* **31**.

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

[15] 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.

[16] Purcaru, O., & Denuit, M. (2002). On the dependence induced by frequency credibility models. II. Time-dependent random effects. Discussion Paper 02-03, Institut de Statistique, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

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

[18] Shaked, M., & Shanthikumar, J.G. (1997). Supermodular stochastic orders and positive dependence of random vectors. *Journal of Multivariate Analysis* **61**, 86-101.

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

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

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