Research Paper

A Claim Score for
Dynamic Claim Counts Modelling

**Research Council –
Academic Research Committee**

October 2019

Ce document est disponible en français

A Claim Score for Dynamic Claim Counts Modelling

Jean-Philippe Boucher and Mathieu Pigeon

Département de mathématiques, UQAM, Montréal, Québec, Canada

Emails: boucher.jean-philippe@uqam.ca and pigeon.mathieu.2@uqam.ca

Abstract

We develop a claim score based on the bonus-malus approach proposed by Boucher and Inoussa (2014). We compare the fit and predictive ability of this new model with various models for panel count data. In particular, we study in more detail a new dynamic model based on the Harvey–Fernandes (HF) approach, which gives different weight to the claims according to their date of occurrence. We show that the HF model has serious shortcomings that limit its use in practice. In contrast, the bonus-malus model does not have these defects. Instead, it has several interesting properties: interpretability, computational advantages, and ease of use in practice. We believe that the flexibility of this new model means that it could be used in many other actuarial contexts. Based on a real database, we show that the proposed model generates the best fit and one of the best predictive capabilities among the other models tested.

1. Introduction and Motivation

In a recent paper, Boucher and Inoussa proposed a new approach to merit pricing for the number of claims in automobile insurance. Instead of assuming a random effect or a copula to model the dependence between all the contracts of a policyholder, the authors directly included a bonus-malus system (BMS) in the modelling. Consequently, all the past claims experience is summarized into a single numerical value: the current level of the BMS or what we call a *claim score* (similar to a credit score). Besides strengthening links between the theory and what is currently applied in practice for pricing, the proposed approach showed an interesting flexibility that can be used, for example, to include some legal constraints in the ratemaking. However, despite the advantages of the approach, the parameter estimation procedure proposed by Boucher and Inoussa (2014) is quite cumbersome and time-consuming.

In this paper, we propose a straightforward but significant modification to this Bonus-Malus Systems Model for Panel Count Data (BMS-panel model). It changes the way the parameters

are estimated, drastically reducing both the complexity and the time of the calibration procedure. Taking advantage of this new flexibility of the model, we perform an exhaustive analysis of the model, where we do not set restriction on the size of the parameters space. In addition to showing some of the properties of the model, we compare the BMS-panel model with various panel data count models, including a new model that we introduce. We show that the fit of the BMS-panel model is the best among the considered distributions, and we analyze its prediction power.

Because the BMS-panel model has a Markovian property that greatly facilitates its use, we believe that the bonus-malus approach could be an interesting option for modelling the complex structure of claim experience in actuarial sciences. For example, in situations where we want to model dependence between cars from the same contract, or when we want to link the frequency and the severity of claims, the *claims score* produced by the BMS model could be considered. This new paradigm could then replace complex approaches such as using a series of correlated random effects (see Abdallah et al., 2016) or a multiple hierarchical copula (see Shi et al., 2016).

The paper is constructed as follows. In Section 2, we briefly present the main ratemaking approaches, where we introduce a new dynamic count distribution for panel data. In Section 3, we review the BMS-panel model, we introduce a simplified version, and we highlight some properties of the model. In Section 4, based on a database from a property and casualty insurance company, we calibrate and compare the proposed models. A very wide range of structural parameters of the BMS-panel model is tested. Finally, we conclude and present some promising generalizations in Section 5.

2. Ratemaking Techniques

We consider a property and casualty portfolio of M policyholders observed over several years. For each contract i, $i = 1, \ldots, M$, we define $N_{i,t}$, a discrete random variable counting the number of claims for the policy period t and $\mathbf{X}_{i,t}$ a column vector containing available explanatory factors at the beginning of period t. In this vector, we may include $d_{i,t}$, a scalar measuring the risk exposure. Through this project, we assume that the primary purpose of a ratemaking (or pricing) model is to provide a prediction for

$$
\mathrm{E}\left[N_{i,T_i+1} \mid \underbrace{N_{i,1}, \ldots, N_{i,T_i}}_{\mathbf{N}_{i,T_i}}, \mathbf{X}_{i,1}, \ldots, \mathbf{X}_{i,T_i+1}\right],
$$

where all policyholders are independent and T_i is the latest observed period for policyholder i. Our main objective is to investigate the strengths and weaknesses of the following classes of models:

- cross-section data models, for which independence is assumed between annual contracts for a policyholder (see Subsection 2.1);

- panel data models, for which we suppose dependence between all contracts written with a policyholder (see Subsection 2.2); and

- BMS-panel models, for which at least some of the past information is summarized using a BMS (see Section 3).

2.1. Cross-section data models

For cross-sectional models, we have independence between all policyholders as well as between all contracts, so we can write

$$\Pr\left(N_{i,t+1} = n | \mathbf{N}_{i,t}, \mathbf{X}_{i,1}, \dots, \mathbf{X}_{i,t+1}\right) = \Pr\left(N_{i,t+1} = n | \mathbf{X}_{i,t+1}\right)$$

and we can simplify our prediction problem in the following way:

$$\mathrm{E}\left[N_{i,t+1} | \mathbf{N}_{i,t}, \mathbf{X}_{i,1}, \dots, \mathbf{X}_{i,t+1}\right] = \mathrm{E}\left[N_{i,t+1} | \mathbf{X}_{i,t+1}\right] = \lambda(\mathbf{X}_{i,t+1}),$$

where $\lambda()$ is a function. Traditionally, in risk classification, we assume a log linear relationship between the mean parameter and the policyholder's and/or claim's characteristics such as sex, age, or marital status (see Denuit et al., 2007).

The base model is usually the Poisson distribution, which is part of the exponential linear family and has useful and well-known statistical properties (see McCullagh and Nelder, 1989, or Frees et al., 2014). The probability mass function is $\Pr\left(N_{i,t} = n | \mathbf{X}_{i,t}\right) = \left(\lambda_{i,t}\right)^{n} \exp(-\lambda_{i,t})/n!$, $n = 0,1,2,\dots$ and 0 elsewhere, where $\lambda_{i,t} = d_{i,t} \exp\left(\mathbf{X}_{i,t}'\boldsymbol{\beta}\right)$ and $\boldsymbol{\beta}$ is a column vector containing model parameters. Finally, we note

$$\pi_{i,t+1}^{\mathrm{Poi}} = \mathrm{E}\left[N_{i,t+1} | \mathbf{X}_{i,t+1}\right] = \lambda_{i,t+1}.$$

This equation refers to an annual premium, when the cost of each claim is 1. Because this premium does not depend on the past claim experience, we usually call $\pi_{i,t+1}$ an *a priori premium*.

This Poisson distribution implies equidispersion, i.e., $\mathrm{E}\left[N_{i,t} | \mathbf{X}_{i,t}\right] = \mathrm{Var}\left[N_{i,t} | \mathbf{X}_{i,t}\right]$ which is, usually, a too-strong assumption in non-life ratemaking. In an effort to overcome this issue, we consider the Negative Binomial (NB) distribution, which is one of the most commonly used alternatives to the Poisson model. To facilitate data analysis, we limit ourselves to the simplest forms of the NB distribution but, the reader can consult Winkelmann (2010). The most intuitive way to construct an NB distribution from a Poisson distribution is to introduce a random heterogeneity term in the mean parameter. Let Θ be a random variable following a gamma distribution, with probability density function $f_{\Theta}(\theta) = \gamma^{\alpha} \theta^{\alpha-1} \exp(-\theta\gamma)/\Gamma(\alpha)$ for $\theta > 0$, with $\alpha > 0$ and $\gamma > 0$. We suppose that $\alpha = 1/\tau$ and $\gamma = 1/\tau$, meaning that $\mathrm{E}[\Theta] = 1$ and $\mathrm{Var}[\Theta] = \tau$. We assume then that $\left(N_{i,t} | \Theta, \mathbf{X}_{i,t}\right) \sim \mathrm{Poisson}(\lambda_{i,t}\theta)$. Therefore, the random variable $\left(N_{i,t} | \mathbf{X}_{i,t}\right)$ follows an NB distribution of type 2 (NB2) with parameters τ and $\lambda_{i,t}$ and probability mass function given by

$$\Pr\left(N_{i,t} = n | \mathbf{X}_{i,t}\right) = \frac{\Gamma(n + 1/\tau)}{\Gamma(n + 1)\Gamma(1/\tau)} \left(\frac{\lambda_{i,t}}{1/\tau + \lambda_{i,t}}\right)^{n} \left(\frac{1/\tau}{1/\tau + \lambda_{i,t}}\right)^{1/\tau}, \qquad n = 0,1,2,\dots$$

and 0 elsewhere. We can directly obtain $\mathrm{E}\left[N_{i,t}\right] = \lambda_{i,t}$ and $\mathrm{Var}\left[N_{i,t}\right] = \lambda_{i,t} + \tau\lambda_{i,t}^{2}$.

We also consider a slightly different version of the NB distribution (NB1) with parameters $\lambda_{i,t}$ and τ for which the probability mass function is

$$\Pr\big(N_{i,t} = n | \mathbf{X}_{i,t}\big) = \frac{\Gamma\left(n + \frac{\lambda_{i,t}}{\tau}\right)}{\Gamma(n+1)\Gamma\left(\frac{\lambda_{i,t}}{\tau}\right)} (1+\tau)^{-\lambda_{i,t}/\tau} \left(1 + \frac{1}{\tau}\right)^{-n}, \qquad n = 0,1,2,\dots$$

and 0 elsewhere. We observe that $\mathrm{Var}\big[N_{i,t}\big] = \lambda_{i,t} + \tau\lambda_{i,t} = \phi\lambda_{i,t}$, which corresponds to the variance function of the overdispersed Poisson in the generalized linear model framework. In all cases, parameters can be easily estimated using a maximum-likelihood procedure.

The premium predicted by both the NB1 and NB2 model is

$$\pi_{i,t+1}^{\mathrm{NB}} = \mathrm{E}\big[N_{i,t+1} | \mathbf{X}_{i,t+1}\big] = \lambda_{i,t+1},$$

where independence between contracts of the same insured is still assumed.

2.2. Panel data models

Panel data models assume some dependence between all annual contracts belonging to a single policyholder. We need a model for the (conditional) random vector $[\mathbf{N}_{i,t} | \mathbf{X}_{i,1}, \dots, \mathbf{X}_{i,t}]$, $t = 1,2,\dots$ in order to predict $\mathrm{E}\big[N_{i,T_i+1} | \mathbf{N}_{i,T_i}, \mathbf{X}_{i,1}, \dots, \mathbf{X}_{i,T_i+1}\big]$. There are plenty of models for the time dependence between random variables – e.g., conditional models, marginal models, and subject-specific models – but it has been shown that random effects models were the best suited for non-life insurance data (see Boucher and Guillén, 2009). In a ratemaking model, an individual random effect may capture variability caused by the lack of information on some important classification variables such as road rage and drug use. Let Θ denote this random effect. Conditionally on Θ, all contracts of the same insured are supposed independent. The joint probability mass function is defined by

$$\Pr\big(N_{i,1} = n_{i,1}, \dots, N_{i,t} = n_{i,t} | \mathbf{X}_{i,1}, \dots, \mathbf{X}_{i,t}\big)$$

$$= \int_{-\infty}^{\infty} \Pr\big(N_{i,1} = n_{i,1}, \dots, N_{i,t} = n_{i,t} | \theta_i, \mathbf{X}_{i,1}, \dots, \mathbf{X}_{i,t}\big) \, dG(\theta_i | \mathbf{X}_{i,1}, \dots, \mathbf{X}_{i,t}),$$

$$= \int_{-\infty}^{\infty} \left(\prod_{k=1}^{t} \Pr\big(N_{i,k} = n_{i,k} | \theta_i, \mathbf{X}_{i,1}, \dots, \mathbf{X}_{i,t}\big) \right) dG(\theta_i).$$

where $G(\theta_i)$ is the cumulative distribution function of the random effect. We assume that the distribution of the random effect θ_i does not depend on covariates $\mathbf{X}$; see Boucher and Denuit (2006) for a discussion about this conventional assumption in actuarial science. Finally, note that the joint distribution can also be expressed as the product of all predictive distributions of each insurance contract for insured i:

$$\Pr\big(N_{i,1} = n_{i,1}, \dots, N_{i,t} = n_{i,t} | \mathbf{X}_{i,1}, \dots, \mathbf{X}_{i,t}\big)$$

$$= \Pr\big(N_{i,1} = n_{i,1} | \mathbf{X}_{i,1}\big) \Pr\big(N_{i,2} = n_{i,2} | N_{i,2} = n_{i,1}, \mathbf{X}_{i,1}, \mathbf{X}_{i,2}\big)$$

$$\times \cdots \times \Pr\big(N_{i,t} = n_{i,t} | N_{i,1} = n_{i,1}, \dots, N_{i,t-1} = n_{i,t-1}, \mathbf{X}_{i,1}, \dots, \mathbf{X}_{i,t}\big).$$

2.2.1. Negative Multinomial

Following the construction of the NB2, the simplest random effects model is given by

$$\left(N_{i,t}|\Theta_i = \theta_i, \mathbf{X}_{i,1}, \dots, \mathbf{X}_{i,t}\right) \sim \mathrm{Poisson}\left(\theta_i \lambda_{i,t}\right)$$

and $\Theta_i \sim$ amma $(\alpha = \kappa, \gamma = \kappa)$, which lead to

$$\Pr\left(N_{i,1} = n|\mathbf{X}_{i,1}\right) = \frac{\Gamma(n+\kappa)}{\Gamma(n+1)\Gamma(\kappa)}\left(\frac{\lambda_{i,1}}{\kappa + \lambda_{i,1}}\right)^n \left(\frac{\kappa}{\kappa + \lambda_{i,1}}\right)^\kappa$$

$$\Pr\left(N_{i,t+1} = n|\mathbf{N}_{i,t}, \mathbf{X}_{i,1}, \dots, \mathbf{X}_{i,t+1}\right) = \frac{\Gamma(n+\alpha)}{\Gamma(n+1)\Gamma(\alpha)}\left(\frac{\lambda_{i,t+1}}{\gamma + \lambda_{i,t+1}}\right)^n \left(\frac{\gamma}{\gamma + \lambda_{i,t+1}}\right)^\alpha,$$

with $\alpha = \kappa + \sum_{k=1}^t n_{i,k}$, $\gamma = \kappa + \sum_{k=1}^t \lambda_{i,k}$, and $\lambda_{i,k} = d_{i,k}\exp(\mathbf{X}_{i,k}'\boldsymbol{\beta})$. Both probability distributions have a NB2 form (for all t), and then can be simply noted as $\mathrm{NB2}_{i,t}(\lambda_{i,t}, \alpha, \gamma)$.

The predictive distribution comes from the classical credibility theory (see Bühlmann and Gisler, 2005), where the *a posteriori* distribution of the heterogeneity term Θ_i is a Gamma distribution with updated parameters $\kappa + \sum_{k=1}^t n_{i,k}$ and $\kappa + \sum_{k=1}^t \lambda_{i,k}$. The joint distribution, called Multivariate NB distribution (MVNB), or Negative Multinomial, is often applied in non-life insurance. Again, parameters can be estimated using a maximum-likelihood method. Therefore, the requested prediction is

$$\pi_{i,1}^{\mathrm{MVNB}} = \mathrm{E}\left[N_{i,1}|\mathbf{X}_{i,1}\right] = \lambda_{i,1}\left(\frac{\kappa}{\kappa}\right) = \lambda_{i,1}$$

$$\pi_{i,t+1}^{\mathrm{MVNB}} = \mathrm{E}\left[N_{i,t+1}|\mathbf{N}_{i,t}, \mathbf{X}_{i,1}, \dots, \mathbf{X}_{i,t+1}\right] = \lambda_{i,t+1}\left(\frac{\kappa + \sum_{k=1}^t n_{i,k}}{\kappa + \sum_{k=1}^t \lambda_{i,k}}\right).$$

Given that each insurance contract of the same policyholder has the same random effects, $N_{i,j}$ and $N_{i,t+j}$ are dependent:

$$\mathrm{Cov}\left[N_{i,t}, N_{i,t+j}|\mathbf{X}_{i,1}, \dots, \mathbf{X}_{i,t+j}\right] = \lambda_{i,t}\lambda_{i,t+j}(1/\kappa), \quad j > 0.$$

2.2.2. NB with random effects

As pointed out by Boucher et al. (2008), the NB with random effects Beta is well suited to model the number of claims, resulting in what we call a NBBeta distribution. From the probability mass function of the NB2 distribution, we assume that $(1/\tau)/(\lambda + 1/\tau) \sim \mathrm{Beta}(a, b)$. Then, we have

$$\Pr\left(N_{i,1} = n|\mathbf{X}_{i,1}\right) = \frac{\Gamma(a+b)\Gamma(a+\lambda_{i,1})\Gamma(b+n)}{\Gamma(a)\Gamma(b)\Gamma(a+b+\lambda_{i,1}+n)}\frac{\Gamma(\lambda_{i,1}+n)}{\Gamma(\lambda_{i,1})\Gamma(n+1)}$$

$$\Pr\left(N_{i,t+1} = n|\mathbf{N}_{i,t}, \mathbf{X}_{i,1}, \dots, \mathbf{X}_{i,t+1}\right) = \frac{\Gamma(\alpha+\gamma)\Gamma(\alpha+\lambda_{i,t+1})\Gamma(\gamma+n)}{\Gamma(\alpha)\Gamma(\gamma)\Gamma(\alpha+\gamma+\lambda_{i,t+1}+n)}\frac{\Gamma(\lambda_{i,t+1}+n)}{\Gamma(\lambda_{i,t+1})\Gamma(n+1)}$$

with $\alpha = a + \sum_{k=1}^t \lambda_{i,k}$, $\gamma = b + \sum_{k=1}^t n_{i,k}$ and $\lambda_{i,k} = d_{i,k}\exp(\mathbf{X}_{i,k}'\boldsymbol{\beta})$. Both probability distributions have an NBBeta form (for all t), and then can be simply noted as

$NBB_{i,t}(\lambda_{i,t}, \alpha, \gamma)$.

We can easily show that $E[N_{i,t}] = \lambda_{i,t}(b/(a-1))$, and

$$\text{Var}[N_{i,t}] = \lambda_{i,t} \frac{(a+b-1)b}{(a-1)(a-2)} + \lambda_{i,t}^2 \left(\frac{(b+1)b}{(a-1)(a-2)} - \frac{b^2}{(a-1)^2} \right)$$

and derive the *a posteriori* distribution of the random effects, which is a Beta distribution with updated parameters $\sum_t \lambda_{i,t} + a$ and $\sum_t n_{i,t} + b$. Thus, we have

$$\pi_{i,1}^{\text{NBB}} = E[N_{i,1}|\mathbf{X}_{i,1}] = \lambda_{i,1} \left(\frac{b}{a-1} \right)$$

$$\pi_{i,t+1}^{\text{NBB}} = E[N_{i,t+1}|\mathbf{N}_{i,t}, \mathbf{X}_{i,1}, ..., \mathbf{X}_{i,t+1}] = \lambda_{i,t+1} \left(\frac{b + \sum_{k=1}^{t} n_{i,k}}{a + \sum_{k=1}^{t} \lambda_{i,k} - 1} \right).$$

Finally, as for the MVNB, the covariance between the number of claims of annual contracts of the same insured can be shown to be equal to:

$$\text{Cov}[N_{i,t}, N_{i,t+j}|\mathbf{X}_{i,1}, ..., \mathbf{X}_{i,t+j}] = \lambda_{i,t}\lambda_{i,t+j} \left(\frac{b}{a-1} \right) \left(\frac{b+1}{a-2} - \frac{b}{a-1} \right), \quad j > 0.$$

Obviously, for random effects models, other choices than the MVNB or the NBBeta are possible to construct.

2.3. Dynamic panel data models

At this point, it is worth mentioning one major drawback of the use of a classic random effects model, such as the MVNB or the NBBeta. By analyzing the predictive premiums, one can see that all past claims have the same importance, i.e., the same weight, in predicting the future premium. This means that a 10-year-old claim is as important as a one-year-old one. In practice, it is generally accepted that this is not a realistic scenario: drivers evolve over time, and recent experience should have a greater impact than older experience when estimating a driver's risk.

Unfortunately, it is not easy to include such temporal dynamics in a panel data model. A random effects approach must assume a random process for $\Theta_{i,t}$. Thus, models where the random effects $\Theta_{i,t}$, $t = 1, ..., T_i$ evolve over time would need a T-dimensional integral to express the joint distribution of all claims of a single policyholder. Therefore, complex numerical procedures or approximated inference methods are needed (see Jung and Liesenfeld, 2001, for example). Other approaches have been proposed to include a dynamic effect into count models: evolutionary credibility models in Albrecht (1985), Poisson residuals in Pinquet et al. (2001), or, more recently, copulas with the jittering method in Shi and Valdez (2016).

In this paper, we consider the Harvey–Fernandes model, or HF model, proposed by Harvey and Fernades (1989) and introduced in actuarial science by Bolancé et al. (2007). Let $\mathcal{H}_{i,t}$ denote the claim history up to time t for contract i. This approach includes random effects that develop over time according to a two-step procedure:

- P-step: the conditional distribution of the random effect $\Theta_{i,t}|\mathcal{H}_{i,t}$ is *predicted*; and

- U-Step: the distribution is *updated* according to some function U and $\Theta_{i,t+1} \sim U(\Theta_{i,t}|\mathcal{H}_{i,t})$.

If we select a conjugate distribution for the random effects (such as the gamma for the MVNB model, and the beta for the NBBeta model), $\Theta_{i,t}$ and $\Theta_{i,t}|\mathcal{H}_{i,t}$ should be from the same distribution. Thus, the selected function U can be directly applied to the parameters of the distribution of $\Theta_{i,t}|\mathcal{H}i,t$ in order to obtain the distribution of $\Theta_{i,t+1}$ (U-step). Consistent with Bolancé et al. (2007), we follow this path and select a function U where the parameters α_t, τ_t of the posterior distribution $\Theta_{i,t}|N_{i,t}$ will be modified as $\alpha_t^* = v\alpha_t$ and $\tau_t^* = v\tau_t$ with starting values α_0 and τ_0. From this structure we can derive the distribution of each $\Theta_{i,1}, \dots, \Theta_{i,t}$.

Given that the joint distribution can be expressed as the product of predictive distributions, the following result can be shown (we remove $\mathbf{X}_i$ for simplicity):

$$\Pr(N_{i,1} = n_{i,1}, \dots, N_{i,t} = n_{i,t}) =$$

$$\Pr(N_{i,1} = n_{i,1}|\alpha_{i,1}, \gamma_{i,1})\Pr(N_{i,2} = n_{i,2}|\alpha_{i,2}, \gamma_{i,2}) \times \dots \times \Pr(N_{i,t} = n_{i,t}|\alpha_{i,t}, \gamma_{i,t}),$$

where only updated parameters are explicitly mentioned in each distribution:

$$\alpha_{i,t} = (v)^{t-1}\alpha_0 + \sum_{k=1}^{t-1} (v)^k n_{i,t-k}$$

$$\gamma_{i,t} = (v)^{t-1}\gamma_0 + \sum_{k=1}^{t-1} (v)^k \lambda_{i,t-k}. \tag{1}$$

Thus, in line with Bolancé et al. (2007), we can use a dynamic MVNB (noted HF-MVNB) by using the product of $\mathrm{NB2}_{i,t}(\lambda_{i,t}, \alpha_{i,t}, \gamma_{i,t})$ distributions for $t = 1, \dots, T_i$. Similarly, we can construct a new dynamic distribution, based on the NBB distribution and then noted HF-NBB, as the product of $\mathrm{NBB}_{i,t}(\lambda_{i,t}, \alpha_{i,t}, \gamma_{i,t})$ distributions. We can see that both dynamic joint distribution put larger weight on recent claims in the predictive premium calculation:

$$\pi_{i,t+1}^{\mathrm{HF-MVNB}} = \mathrm{E}[N_{i,t+1}|\mathbf{N}_{i,t}, \mathbf{X}_{i,1}, \dots, \mathbf{X}_{i,t+1}]$$

$$= \lambda_{i,t+1} \left(\frac{(v)^t\kappa + \sum_{k=1}^{t} (v)^k n_{i,t-k+1}}{(v)^t\kappa + \sum_{k=1}^{t} (v)^k \lambda_{i,t-k+1}} \right)$$

and

$$\pi_{i,t+1}^{\mathrm{HF-NBB}} = \mathrm{E}[N_{i,t+1}|\mathbf{N}_{i,t}, \mathbf{X}_{i,1}, \dots, \mathbf{X}_{i,t+1}]$$

$$= \lambda_{i,t+1} \left(\frac{(v)^t b + \sum_{k=1}^{t} (v)^k n_{i,t-k+1}}{(v)^t a + \sum_{k=1}^{t} (v)^k \lambda_{i,t-k+1} - 1} \right).$$

We discuss the covariance between the number of claims of annual contracts of the same policyholder in Section 4.

3. BMS-Panel Models

Bonus-malus systems have been introduced in ratemaking procedures for many practical reasons (see Lemaire, 1995, and Denuit et al., 2007). The basic idea underlying BMS models is to summarize past claims experience into a *claim score*: a discrete value going from 1 to s, where 1 represents the lowest risk and s the highest. In practice, when a new policyholder enters the portfolio, the insurance company gives that individual a selected entry level as an initial claim score. Each year, depending on the insured's claim experience, the policyholder will move into the bonus-malus scale: toward high values for frequent claims, toward lower values in the opposite case.

Definition 3.1 (Bonus-Malus System): *A bonus-malus system is a three-parameter model* (Ψ, s, ℓ^*) *where the level of the system at the beginning of period* $t+1$ *in the database is given by*

$$L(t+1) = \min\big(\max\big(L(t) - \mathbb{I}\big(N_{i,t} = 0\big) + \Psi N_{i,t}, 1\big), s\big), \tag{2}$$

where $\mathbb{I}\big(N_{i,t} = 0\big)$ is a dummy variable indicating a period without claims, s is the highest level of the system, and Ψ is the *jump parameter*. The parameter ℓ^* corresponds to the entry level of the system for a policyholder without experience. This kind of structure of the BMS is commonly denoted by $-1/+\Psi$.

Figure 1 presents an example of a BMS.

Figure 1: Example of a BMS. At the beginning of the first period, the new policyholder i enters the system at level ℓ^*. If no accidents are filed during the first insurance contract, the policyholder moves from level ℓ^* to level $\ell^* - 1$ at the beginning of the second period. For each claim, the policyholder's BMS level increases by Ψ.
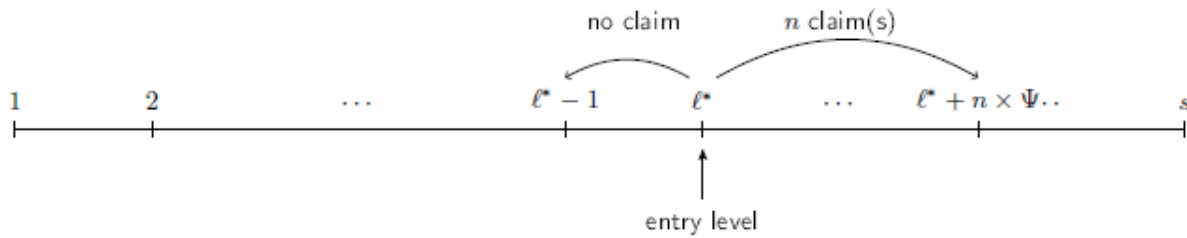
Intuitively, by averaging the claims frequency for each claim score, the insurance company could then construct a form of merit rating, representing relativities for BMS (see, for example, Figure 5 of Boucher and Inoussa, 2014). However, we can improve the rating system by choosing other approaches than averaging the frequency of each claim score. An extensive body of literature deals with how to calibrate a BMS with cross-section insurance data (see Denuit et al., 2007, for an overview). However, when we observe many annual contracts for a single insured in a data set (i.e., panel data structure), Boucher and Inoussa (2014) presents a more suitable approach, where the authors construct what they call a BMS-panel model.

3.1. BMS-panel model

The main objective of the BMS-panel model is to estimate the parameters needed for the *a priori* and the *a posteriori* ratemaking simultaneously, by including the BMS structure directly in the modelling. In order to estimate parameters, we need a model for the conditional random vector $[\mathbf{N}_{i,t}|\mathbf{X}_{i,1}, \dots, \mathbf{X}_{i,t}]$. First, let $[N_{i,1}, N_{i,2}]$ be a random vector in a portfolio with a BMS as defined in Definition 3.1 and suppose that we know $\ell_{i,1}$, the BMS level of insured i at time 1. The conditional joint probability mass function is given by (we drop the dependence on $\mathbf{X}$ to simplify the presentation)

$$\Pr(N_{i,1} = n_{i,1}, N_{i,2} = n_{i,2}|L(1) = \ell_{i,1})$$

$$= \Pr(N_{i,1} = n_{i,1}|L(1) = \ell_{i,1})\Pr(N_{i,2} = n_{i,2}|N_{i,1} = n_{i,1}, L(1) = \ell_{i,1})$$

$$= \Pr(N_{i,1} = n_{i,1}|L(1) = \ell_{i,1})$$

$$\times \left(\sum_{y=1}^{s} \Pr(N_{i,2} = n_{i,2}|N_{i,1} = n_{i,1}, L(1) = \ell_{i,1}, L(2) = y) \Pr(L(2) = y|N_{i,1} = n_{i,1}, L(1) = \ell_{i,1}) \right).$$

Given the BMS, past information is captured by the last level reached by the system, we obtain

$$= \Pr(N_{i,1} = n_{i,1}|L(1) = \ell_{i,1}) \left(\sum_{y=1}^{s} \Pr(N_{i,2} = n_{i,2}|L(2) = y)\Pr(L(2) = y|N_{i,1} = n_{i,1}, L(1) = \ell_{i,1}) \right)$$

$$= \Pr(N_{i,1} = n_{i,1}|L(1) = \ell_{i,1})\Pr(N_{i,2} = n_{i,2}|L(2) = \ell_{i,2}),$$

where $\Pr(L(2) = y|N_{i,1} = n_{i,1}, L(1) = \ell_{i,1}) = 0$ for all y, except for $y = \ell_{i,2}$, the level reached by the BMS at time 2 after $n_{i,1}$ claims were observed during the year. Consequently,

$$\Pr(N_{i,1} = n_{i,1}, N_{i,2} = n_{i,2}, \dots, N_{i,t} = n_{i,t}|L(1) = \ell_{i,1}) = \prod_{k=1}^{t} \Pr(N_{i,k} = n_{i,k}|L(k) = \ell_{i,k}). \quad (3)$$

Several options are available to model the conditional distribution $\Pr(N_{i,t} = n_{i,t}|L(t) = \ell_{i,t})$: we can consider the distributions introduced in subsections 2.1 or 2.2, as well as the hurdle or the zero-inflated distributions (see Boucher and Guillén, 2009, for an extensive overview of count distributions in ratemaking). For a selected distribution, the BMS level should be used to model the mean parameter of the conditional distribution. Based on the Poisson, the NB2 or the NB1 introduced earlier, we assume that the mean parameter will be modelled as

$$\pi_{i,t+1}^{\text{BMS}} = \mathrm{E}[N_{i,t+1}|\mathbf{N}_{i,t+1}, \mathbf{X}_{i,1}, \dots, \mathbf{X}_{i,t+1}] = \lambda_{i,t+1} r_{\ell_{i,t+1}},$$

where $\lambda_{i,t}$ is an *a priori* premium based on the characteristics of the insured (sex, age, etc.) for the period $[t, t+1)$. Several structures for r_ℓ can be chosen, but we will restrict ourselves to studying linear relativities for the BMS, as defined below.

Definition 3.2 (Linear Bonus-Malus System): *A linear relativity is associated with each step of the scale according to the equation*

$$r_{L(t)} = 1 + \delta(L(t) - 1), \tag{4}$$

where δ is the *penalty parameter*.

Equation (4) implies that $r_1 = 1$ and defines the basis risk. These linear relativities, initially proposed by Gilde and Sundt (1989), prevent unwanted situations such as $r_i > r_j$ for $i < j$.
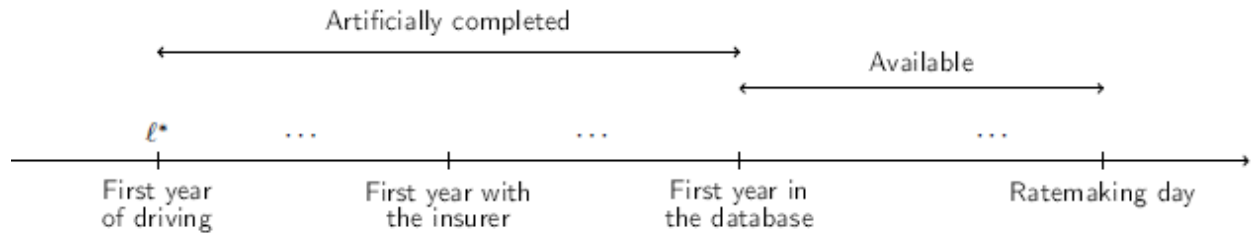
3.2. Entry level

The distributions previously mentioned were defined conditionally on the knowledge of the BMS level of insured i at time 1. Consequently, the joint probability mass function of policyholder i at time t is given by

$$\Pr\big(N_{i,1} = n_{i,1}, \dots, N_{i,t} = n_{i,t} | \mathbf{X}_{i,1}, \dots, \mathbf{X}_{i,t}\big)$$

$$= \sum_{y=1}^{s} \Pr\big(N_{i,1} = n_{i,1}, \dots, N_{i,t} = n_{i,t} | L(1) = y, \mathbf{X}_{i,1}, \dots, \mathbf{X}_{i,t}\big)\Pr(L(1) = y)$$

$$= \sum_{y=1}^{s} \prod_{k=1}^{t} \Pr(N_{i,k} = n_{i,k} | L(k) = \ell_{i,k}, \mathbf{X}_{i,1}, \dots, \mathbf{X}_{i,k})\Pr(L(1) = y),$$

where $\Pr(L(1) = y)$ is the probability distribution of the BMS level at time $t = 1$, i.e, the first year an insured appears in the database. One must not confuse the first year of driving with the first year an insured is observed in the database, as illustrated in Figure 2.

Figure 2: Example of a policyholder with partial unknown information. In order to estimate ℓ^*, the insurer must complete the trajectory followed by the BMS between the first year of driving and the first year in the database.

If we want to model the joint distribution for a new driver, $\ell_{i,1}$ can be found directly: it is ℓ^*, the entry level selected in the construction of the model. Thus $\Pr(L(1) = y) = 0$ for all y, except $y = \ell^*$. For experienced drivers, this situation is more difficult to deal with. Insurers must be careful and should not suppose that all new insureds did not have claims in past years, nor must they automatically suppose that they should give policyholders an entry level ℓ^*. In Boucher and Inoussa (2014), the authors propose to recreate all the possible events of each policyholder from the first year of driving to the first year in the database, to create the distribution of L_1. By taking the average of L_1, they assigned each driver in the database a value of ℓ_1. Although the method is intuitive, it is a very complex to implement, in addition to

requiring a lot of time and IT resources. In this paper, we propose a much simpler method based on the fact that the probability of not filing an accident during a single year is very high, often in the range of 80–95%.

Proposition 3.3: *For a policyholder i, let $S_i(t)$ be the level reached by a (Ψ, s, ℓ^*) BMS after t periods from the first exposure year and u_i be the first observed year in the portfolio. Thus, if*

$$\Pr(N_{i,t} = 0 | L(t) = \ell_{i,t}, \mathbf{X}_{i,1}, \dots, \mathbf{X}_{i,t}) > 0.5, \qquad \forall t, \tag{5}$$

then

$$\operatorname*{argmax}_{\ell \in \{1,\dots,s\}} \left(\Pr(S_i(1) = \ell^*, \dots, S_i(u_i + 1) = \ell | \mathbf{X}_{i,1}, \dots, \mathbf{X}_{i,t}) \right) = \max(\ell^* - u_i, 1).$$

Proof. Based on Equation (2), the transition probabilities are given by (for a policyholder with a claim frequency of $\lambda_{i,t}$ for the period t)

$$p_{k,j}(\lambda_{i,t}) = \Pr(L(t+1) = j | L(t) = k, \mathbf{X}_{i,1}, \dots, \mathbf{X}_{i,t})$$

$$= \Pr(\min(\max(k - \mathbb{I}(N_{i,t} = 0) + \Psi N_{i,t}, 1), s) = j | \mathbf{X}_{i,1}, \dots, \mathbf{X}_{i,t}), \tag{6}$$

for $j = 1, \dots, s$ and $k = 1, \dots, s$. This implies that $p_{k,\max(k-1,1)}(\lambda_{i,t}) > 0.5$ and $p_{k,j}(\lambda_{i,t}) < 0.5$, $j \neq \max(k - 1, 1)$ if $\Pr(N_{i,t} = 0 | L(t) = \ell_{i,t}, \mathbf{X}_{i,1}, \dots, \mathbf{X}_{i,t}) > 0.5$. Even if the insurance company does not know the policyholders' characteristics before they enter the portfolio, it is reasonable to assume that they are such that Equation (5) is satisfied. Thus,

$$p_{\ell^*,\max(\ell^*-1,1)}(\lambda_{i,1}) p_{\max(\ell^*-1,1),\max(\ell^*-2,1)}(\lambda_{i,2}) \times \dots \times p_{\max(\ell^*-u_i+1,1),\max(\ell^*-u_i,1)}(\lambda_{i,u_i})$$

is the most likely path from level ℓ^* to level $\max(\ell^* - u_i, 1)$. Finally,

$$\operatorname*{argmax}_{\ell \in \{1,\dots,s\}} (\Pr(S_i(1) = \ell^*, \dots, S_i(u_i + 1) = \ell)) = \max(\ell^* - u_i, 1).$$

Figure 3: Example of a policyholder with partial unknown information (u_i years are unknown and T_i years are known).
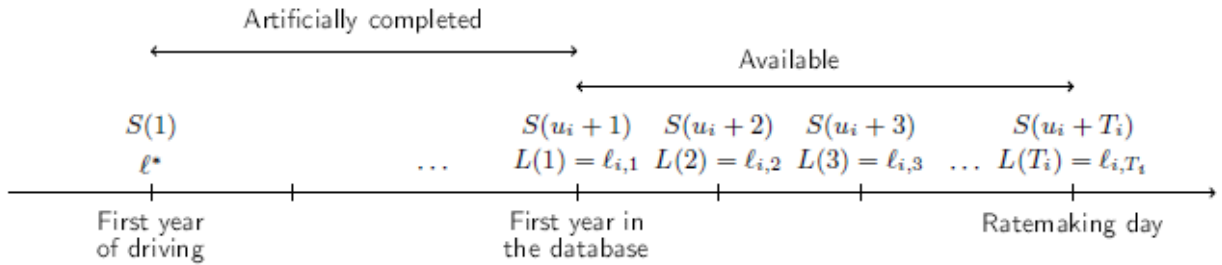

Figure 3 illustrates the situation, and the new variables introduced. Based on the result of the last proposition, we then select the entry level of a new policyholder with u years of experience as $\max(\ell^* - u, 1)$, or in other words $\Pr(L(1) = y) = 0$ for all y, except $y = \max(\ell^* - u, 1)$. Therefore, each year of driving experience results in a decrease of one level of the BMS, which is already one of the ways insurance companies deal with experienced drivers

in merit rating plans. So, this way of selecting $\ell_{i,1}$ not only simplifies the BMS-panel model, but it theoretically justifies the procedure already in use.

3.3. Parameters

The joint distribution of N_{i,T_i} is now completely defined for all insured in the database. We can thus summarize the steps needed to estimate all the parameters of the model. First, the actuary has to select, or estimate, the following structural components of the BMS-panel model:

- the number of levels s of the system;
- the jump parameter Ψ for each claim;
- the entry level ℓ^* for a new driver;
- the risk characteristics $\mathbf{X}_{i,1}, \dots, \mathbf{X}_{i,T_i}$ and the associated parameters $\boldsymbol{\beta}$; and
- the underlying count distribution.

Then, for this specific BMS-panel model, N_{i,T_i}, the *a priori* parameters ($\boldsymbol{\beta}$) linking the covariates with the expected value and the *penalty parameter* δ must be estimated by maximizing the likelihood function. When structural components of the BMS-panel model are selected, parameter estimation is direct and can be done using standard statistical software. Other parameters linked to the underlying count distribution should also be estimated. For example, if an NB1 or a NB2 distribution is used, an estimator of the overdispersion parameter α should also be found.

Note that for parameters s, Ψ and ℓ^*, the structure of the model provides us with some additional information that allows us to reduce the parameter space to a lattice. This is a well-known problem in statistics; for example, see Hammersley (1950) for simple cases or Choirat and Raffaello (2012) for asymptotic properties. To obtain the best BMS-panel model, we adjust all combinations of s, Ψ and ℓ^* and we select the model that generates the best likelihood profile and/or the best prediction capacity (based on out-of-sample analysis).

3.4. Properties of the BMS-panel model

Among the properties of the BMS-panel model, we are interested in evaluating the covariance between premiums paid by a policyholder for two periods. To do this, it is necessary to study the covariance between N_t and N_{t+k}, $k = 1,2,\dots$, conditional to $L(t) = \ell_t$.

For a BMS-panel model, the one-year probability of the random variable L going from BMS level $\ell_{i,t}$ to BMS level $\ell_{i,t+1}$ is denoted by $p_{\ell_{i,t},\ell_{i,t+1}}(\lambda_{i,t})$ as defined by equation (6). For a policyholder, we can construct a transition probability matrix

$$\mathbf{P}(\lambda_{i,t}) = \begin{bmatrix} p_{1,1}(\lambda_{i,t}) & p_{1,2}(\lambda_{i,t}) & \cdots & p_{1,s}(\lambda_{i,t}) \\ p_{2,1}(\lambda_{i,t}) & p_{2,2}(\lambda_{i,t}) & \cdots & p_{2,s}(\lambda_{i,t}) \\ \vdots & \vdots & \ddots & \vdots \\ p_{s,1}(\lambda_{i,t}) & p_{s,2}(\lambda_{i,t}) & \cdots & p_{s,s}(\lambda_{i,t}) \end{bmatrix}.$$

We can show that for all $K = 1, 2, \ldots$, we have

$$\mathbf{P}^{(K)}(\lambda_{i,t}) = \mathbf{P}^K(\lambda_{i,t}),$$

meaning that the transition probability matrix over K time periods is simply the Kth power of the annual transition probability matrix $\mathbf{P}(\lambda_{i,t})$.

Proposition 3.4: *In a BMS-panel model, at the beginning of a period t, the conditional covariance between $N_{i,t}$ and $N_{i,t+j}$, $j = 1, 2, \ldots$ is*

$$\text{Cov}\big[N_{i,t}, N_{i,t+j} | \ell_{i,t}, \mathbf{X}_{i,1}, \ldots, \mathbf{X}_{i,t}\big] = \lambda_{i,t+j}$$

$$\times \sum_{m=1}^{s} r_m \left(\text{E}\left[N_{i,t} p^{(j-1)}_{\min(\max(\ell_{i,t} - \mathbb{I}(N_{i,t}=0) + \Psi N_{i,t}, 1), s), m}(\lambda_{i,t}) | \ell_{i,t}, \mathbf{X}_{i,1}, \ldots, \mathbf{X}_{i,t}\right] \right.$$

$$\left. - \lambda_{i,t} r_{\ell_{i,t}} p^{(j)}_{\ell_{i,t}, m}(\lambda_{i,t}) \right),$$

where the transition probabilities are extracted from the transition matrices $\mathbf{P}(\lambda_{i,t})^{(j)}$ and $\mathbf{P}(\lambda_{i,t})^{(j-1)}$ assuming that $\mathbf{X}_{i,t} = \mathbf{X}_{i,t+1} = \mathbf{X}_{i,t+2} = \cdots$.

Proof. We have (we drop the reference to i and to $\mathbf{X}_{i,1}, \ldots, \mathbf{X}_{i,t}$ in order to ease the presentation)

$$\text{E}[N_t | \ell_t] = \lambda_t r_{\ell_t}$$

$$\text{E}\big[N_{t+j} | \ell_t\big] = \sum_{n=0}^{\infty} n \Pr\big(N_{t+j} = n | \ell_t\big)$$

$$= \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} n \Pr\big(N_{t+j} = n | L(t+j) = m, \ell_t\big) \Pr(L(t+j) = m | \ell_t).$$

The inner sum is

$$\sum_{n=0}^{\infty} n \Pr\big(N_{t+j} = n | L(t+j) = m, \ell_t\big) = \sum_{n=0}^{\infty} n \Pr\big(N_{t+j} = n | L(t+j) = m\big)$$

$$= \text{E}\big[N_{t+j} | L(t+j) = m\big]$$

$$= \lambda_{t+j} r_m.$$

Thus,

$$\text{E}\big[N_{t+j} | \ell_t\big] = \sum_{m=0}^{\infty} \lambda_{t+j} r_m \Pr(L(t+j) = m | \ell_t)$$

$$= \sum_{m=0}^{\infty} \lambda_{t+j} r_m \sum_{q_1=0}^{\infty} \Pr(L(t+j) = m | L(t+j-1) = q_1, \ell_t) \Pr(L(t+j-1) = q_1 | \ell_t)$$

And, assuming that $\mathbf{X}_{i,t} = \mathbf{X}_{i,t+1} = \mathbf{X}_{i,t+2} = \cdots,$ we have

$$= \lambda_{t+j} \sum_{m=0}^{\infty} r_m \sum_{q_1=0}^{\infty} p_{q_1,m}(\lambda_{i,t}) \Pr(L(t+j-1) = q_1|\ell_t).$$

Recursively, we obtain

$$= \lambda_{t+j} \sum_{m=0}^{\infty} r_m \sum_{q_1=0}^{\infty} p_{q_1,m}(\lambda_{i,t}) \cdots \sum_{q_{j-1}=0}^{\infty} p_{q_{j-1},q_{j-2}}(\lambda_{i,t}) p_{\ell_t,q_{j-1}}(\lambda_{i,t}),$$

and

$$= \lambda_{t+j} \sum_{m=0}^{\infty} r_m p_{\ell_t,m}^{(j)}(\lambda_{i,t}).$$

$$\mathrm{E}\big[N_t N_{t+j}|\ell_t\big] = \sum_{n_0=0}^{\infty} \sum_{n_j=0}^{\infty} n_0 n_j \Pr\big(N_t = n_0, N_{t+j} = n_j|\ell_t\big)$$

$$= \sum_{n_0=0}^{\infty} \sum_{n_j=0}^{\infty} n_0 n_j \Pr(N_t = n_0|\ell_t)\Pr\big(N_{t+j} = n_j|\ell_t, N_t = n_0\big)$$

$$= \sum_{n_0=0}^{\infty} \sum_{n_j=0}^{\infty} n_0 n_j \Pr(N_t = n_0|\ell_t) \sum_{m=0}^{\infty} \Pr\big(N_{t+j} = n_j|L(t+j) = m, \ell_t, N_t = n_0\big)$$

$$\times \Pr(L(t+j) = m|\ell_t, N_t = n_0)$$

$$= \sum_{n_0=0}^{\infty} \sum_{n_j=0}^{\infty} \sum_{m=0}^{\infty} n_0 n_j \Pr(N_t = n_0|\ell_t)\Pr\big(N_{t+j} = n_j|L(t+j) = m\big)$$

$$\times \sum_{q_1=0}^{\infty} \Pr(L(t+j) = m|L(t+j-1) = q_1)\Pr(L(t+j-1) = q_1|\ell_t, N_t = n_0).$$

Recursively, we obtain

$$= \sum_{n_0=0}^{\infty} \sum_{n_j=0}^{\infty} \sum_{m=0}^{\infty} n_0 n_j \Pr(N_t = n_0|\ell_t)\Pr\big(N_{t+j} = n_j|L(t+j) = m\big)$$

$$\times p_{\min(\max(\ell_t - \mathbb{I}(n_0=0) + \Psi n_0, 1), s), m}^{(j-1)}(\lambda_{i,t})$$

$$= \lambda_{t+j} \sum_{n_0=0}^{\infty} \sum_{m=0}^{\infty} r_m n_0 \Pr(N_t = n_0|\ell_t) p_{\min(\max(\ell_t - \mathbb{I}(n_0=0) + \Psi n_0, 1), s), m}^{(j-1)}(\lambda_{i,t})$$

$$= \lambda_{t+j} \sum_{m=0}^{\infty} r_m \mathrm{E}\left[N_t p_{\min(\max(\ell_t - \mathbb{I}(N_t=0)+\Psi N_t, 1), s), m}^{(j-1)}(\lambda_{i,t}) \mid \ell_t\right]$$

and the result follows directly from the definition of a covariance.

For $j = 1$, i.e., for two successive periods, the result of Proposition 3.4 simplifies to

$$\mathrm{Cov}\left[N_{i,t}, N_{i,t+1} \mid \ell_{i,t}, \mathbf{X}_{i,1}, \ldots, \mathbf{X}_{i,t}\right] = \sum_{n=0}^{\infty} \sum_{q=0}^{\infty} nq \mathrm{Pr}\left(N_{i,t} = n \mid \ell_{i,t}, \mathbf{X}_{i,1}, \ldots, \mathbf{X}_{i,t}\right)$$

$$\times \mathrm{Pr}\left(N_{i,t+1} = q \mid \ell_{i,t+1}, \mathbf{X}_{i,1}, \ldots, \mathbf{X}_{i,t}\right)$$

$$- \lambda_{i,t} \lambda_{i,t+1} r_{\ell_{i,t}} \sum_{m=1}^{s} r_m p_{\ell_{i,t}, m}(\lambda_{i,t}),$$

with $\ell_{i,t+1} = \ell_{i,t} - \mathbb{I}(n = 0) + n\Psi$.

4. Empirical Illustrations

We analyze a database from a portfolio of general liability insurance policies for private individuals from a major Canadian insurance company. Because bonus-malus systems are generally built for auto insurance products, we only consider the private use of vehicles in our project. The sample contains information about 429,333 contracts from 140,714 policyholders, and runs from the year 2012 to 2016. The sample has the following properties:

1. We only keep policyholders with a maximum of one insured car through all their observed contracts in order to avoid possible within-contract dependence;

2. We only keep policyholders with complete coverages on all their observed contracts, to avoid censorship where some claims would not be covered or observed; and

3. We remove from the database policyholders who use their car for commercial purposes, since they might exhibit different driving behaviours.

These selection criteria might affect the results. For example, by choosing only insureds with one car, we might have a higher proportion of single insured than normally. However, the purpose of this study is not to explain the accident process by covariates, nor the claiming process, but only to present interesting count models that can be considered for ratemaking. To evaluate the performance of our models, we split the database into two components: a fitting set (98,589 policyholders) and a validation set (42,125 policyholders). Table 1 describes the eight covariates selected in the modelling. For every contract, we have initial information at the beginning of the period and we are interested in predicting the number of claims (excluding comprehensive claims). The average claim frequency is approximately 6.5% and we observe a maximum of five claims per contract. The distribution of the number of contracts observed per insured is shown in Table 2, for an average of 3.05 contracts per policyholder.

Table 1: Dichotomous variables in the database.

Variable	Description
$X_1 = 1$	The policyholder is female.
$X_2 = 1$	The policyholder is married.
$X_3 = 1$	The policyholder is less than 30 years old.
$X_4 = 1$	The policyholder is between 30 and 50 years old.
$X_5 = 1$	The distance driven per year is less than 10,000 km.
$X_6 = 1$	The distance driven per year is between 10,000 km and 20,000 km.
$X_7 = 1$	The distance driven per year is between 20,000 km and 30,000 km.
$X_8 = 1$	The vehicle is used to commute.

Table 2: Distribution of the number of contracts

Number of year	Percentage
1	21.42%
2	17.73%
3	11.66%
4	32.57%
5	16.62%

4.1. Count distributions

We present, in Table 3, fitting results for various models. We observe that the Harvey–Fernandes modification of both the MVNB and the NBBeta distribution did not improve the fitting, resulting in a value of $v = 1$ for both models. We suppose that this is explained by the average of 3.05 contracts per policyholder, where the weight of past claims can still be supposed equal in the prediction of new claims.

Table 3: Fitting results

Model	# parameters	Logl.	AIC	BIC
Poisson	9	−85526,64	171071.28	171280.34
NB2	10	−85238.04	170496.08	170728.37
NB1	10	−85149.08	170318.16	170550.45
MVNB	10	−84812.54	169645.08	169877.37
NBBeta	11	−84727.13	169476.26	169731.78
HF-MVNB	11	−84812.54	169647.08	169902.60
HF-NBBeta	12	−84727.13	169478.26	169757.01
MVNB *	10	−84611.88	169243.76	169476.05
NBBeta *	11	−84426.63	168875.26	169130.78
HF-MVNB *	11	−84460.82	168943.64	169199.16
HF-NBBeta *	12	$\mathbf{-84355.07}$	$\mathbf{168734.14}$	$\mathbf{169012.89}$

We can observe two results for each panel data count distribution (with and without *). This can be explained by additional information in our database. Indeed, the province of Ontario uses Autoplus, a database that provides detailed automobile claims and policy history. Thus, it is possible for insurers to have access to the past claims history of a policyholder for up to 10 years. In the construction of the database used for this project, we are then able to know the past 10 years of claim experience for each insured (before their first insurance contract observed in the database). However, only the claim experience for each insured is available, and the risk characteristics ($\mathbf{X}$) for those past 10 years are unknown. Because we will use this information in the BMS-panel model, we also tried to include this past experience in panel data models. Thus, we propose a slightly modified version of the MVNB model and the NBBeta model, which we name MVNB * and NBBeta *. Specifically, we assume that the distribution of the random variable Θ_i has already been adapted to account for past claims. Following a similar development to that described above, we could show that for a policyholder with m years of experience before entering the database, the distribution of Θ_i for the first observed contract of an insured (gamma for the MVNB, beta for the NBBeta) will have the following parameters

$$\alpha^* = \alpha + \sum_{j=1}^{m_i} n^*_{i,-j} \quad \text{and} \quad \gamma^* = \gamma + \sum_{j=1}^{m_i} \lambda_{i,-j},$$

where m_i is the minimum value between 10 and the driving experience (in years) for insured i, and $n^*_{i,-t}$ is the observed number of claim(s) t year(s) before the entry of the policyholder into the database. Because we do not observe the $\lambda_{i,-j}$, $j = 1, \dots, m_i$, we approximate them by $\bar{\lambda} = 6.5\%$, the average frequency of the database.

Finally, the predicted premium for the MVNB* can be shown to be equal to

$$\pi_{i,t+1}^{\text{MVNB}^*} = \lambda_{i,t+1}\left(\frac{\sum_{k=1}^{t} n_{i,k} + \alpha^*}{\sum_{k=1}^{t} \lambda_{i,k} + \gamma^*}\right)$$

$$= \lambda_{i,t+1}\left(\frac{\sum_{j=1}^{m_i} n_{i,-j}^* + \sum_{k=1}^{t} n_{i,k} + \alpha}{m_i\lambda + \sum_{k=1}^{t} \lambda_{i,k} + \gamma}\right), \tag{7}$$

where all the available claim experience can be used to estimate the future premium. The predicted premium for the NBBeta* can also be computed straightforwardly. Note that modified Harvey–Fernandes approaches (HF-MVNB* and HF-NBBeta*) can also be constructed using the same procedure, with

$$\alpha_{i,1}^* = (v)^{m_i}\alpha_0 + \sum_{k=1}^{m_i} (v)^k n_{i,-k} \text{ and } \gamma_{i,1}^* = (v)^{m_i}\gamma_0 + \sum_{k=1}^{m_i} (v)^k \lambda_{i,-k}.$$

When we include this new information, the HF-NBBeta* outperforms all the other distributions in fitting statistics, even if we considerer penalized criteria such as the AIC or the BIC. We also use the out-of-sample data to compare the models. Table 4 shows the results where we compute two measures evaluating the prediction capacity of each model: the mean square error (MSE), and $-$ because we are dealing with count data and not a continuous distribution $-$ a loglikehood statistic from the Poisson distribution. Based on both out-of-sample statistics, the HF-MVNB* seems to offer a better prediction capacity.

Table 4: Out-of-sample statistics.

Model	Logl. (Poisson)	MSE
Poisson	-36966.48	10609.54
NB2	-36860.55	10611.03
NB1	-36815.96	10609.53
MVNB	-36731.55	10567.37
NBBeta	-36830.10	10601.31
HF-MVNB	-36731.55	10567.37
HF-NBBeta	-36830.10	10601.31
MVNB *	-36588.47	10552.34
NB-Beta *	-36628.14	10564.14
HF-MVNB *	$\mathbf{-36543.69}$	$\mathbf{10543.84}$
HF-NBBeta *	-36628.04	10570.82

4.2. BMS-panel models

We adjust a BMS-panel model with Poisson, NB1, and NB2 underlying distributions. As we did for the previous models marked with $*$, we also consider past claim experience from *Autoplus*. This 10-year history of past claims allows us to find ℓ_1, the BMS level of each insured when they are observed in the database at the first time. We choose several combinations of structure parameters on a grid given by $s = 2,3,\dots,S$, $\Psi = 1,2,\dots,s$ and $\ell^* = 1,\dots,s$, where $S = 22$ for the Poisson distribution, $S = 16$ for the NB1 distribution and $S = 15$ for the NB2 distribution. These values of S cover between 1,000 and 3,000 possibilities for each underlying distribution. Because each estimation step takes at least two to five minutes on a personal computer, covering all those possibilities is very time-consuming. We are looking for a procedure that could help us to narrow the space of $\{s, \Psi, \ell^*\}$ to find the best combinations for a specific underlying distribution.

Table 5: BMS-panel data statistics.

Distribution	# par.	Ψ	s	ℓ^*	Loglikelihood	AIC	BIC
Poisson	13	6	11	1	-84672.04	169370.08	169508.07
	13	6	11	2	-84672.63	169371.26	169509.25
	13	6	10	1	-84673.27	169372.54	169510.53
	13	6	10	2	-84673.89	169373.78	169511.77
NB1	14	6	11	1	$\mathbf{-84331.55}$	$\mathbf{168691.11}$	$\mathbf{168839.71}$
	14	6	10	1	-84331.92	168691.84	168840.45
	14	6	11	2	-84332.13	168692.27	168840.87
	14	6	10	2	-84332.52	168693.05	168841.65
NB2	14	6	11	1	-84442.99	168913.97	169062.58
	14	6	11	2	-84443.56	168915.13	169063.73
	14	6	10	1	-84444.49	168916.98	169065.58
	14	6	10	2	-84445.09	168918.19	169066.79

For each model, we estimate $\boldsymbol{\beta}$, δ, and an overdispersion parameter (for both NB distributions) using the estimation set. Finally, as we did for the other distributions, we calculate the value of the MSE, as well as the loglikelihood of a Poisson distribution for each fitted model based on the test sample, in both cases to prevent over-adjustment. In Table 5, we present the results of the four best BMS-panel models for all three underlying distributions, evaluated on the estimation dataset. We observe that the model $-1/+6$ with $s = 11$ is the best model for the Poisson, the NB1, and the NB2. The four best models for each underlying distribution have the same structural parameters. Only the second-best model of the NB1 ranks third for the Poisson and the NB2. The value of δ does not seem to depend on the underlying distribution but rather depends on the three structural parameters s, Ψ and ℓ^*. Finally, we

can see that the underlying distribution NB1 offers better fitting statistics than the NB2, reflecting what we already observed for cross-section data distributions. Formally, some statistical tests could be done to determine if the overdispersion parameter of the NB1 and the NB2 is statistically significant, but the huge differences between loglikelihoods already show that is the case. As done previously, we also compute out-of-sample statistics to assess the prediction capacity of BMS models. Table 6 presents the results.

Table 6: Out-of-sample statistics for the BMS-panel model.

Distribution	s	Ψ	ℓ^*	MSE	Logl. (Poisson)
Poisson	11	6	1	10550.35	-36587.65
	11	6	2	10550.41	-36587.94
	10	6	1	10549.90	-36586.96
	10	6	2	10549.91	-36587.22
NB1	11	6	1	10550.41	-36587.57
	10	6	1	10549.93	-36586.85
	11	6	2	10550.42	-36587.80
	10	6	2	10549.95	-36587.11
NB2	11	6	1	10552.06	-36586.44
	11	6	2	10552.14	-36586.74
	10	6	1	10551.54	-36585.73
	10	6	2	10551.61	-36586.04

The best model selected by the estimation dataset is the NB1, $-1/+6$ with $s = 11$ and $\ell^* = 1$. Out-of-sample statistics show that this model has an interesting predictive capacity, only outperformed by the HF-MVNB [*]. We also see that the BMS-panel data model makes stable predictions, whereas the 15 other BMS-panel models selected show similar out-of-sample statistics (the greatest difference is less than 0.005%).

Finally, because the best model selected by the data is the NB1, $-1/+6$ with $s = 11$ and $\ell^* = 1$, we point out that

- a policyholder reaches the top of the BMS after two claims in three years, resulting in an annual premium 2.2 times higher than the premium of a policyholder located at level 1;

- after one claim, it takes six years for an insured revert to their initial position; and

- new drivers are assigned an entry level of 1, the best of the BMS. This result highly depends on the data used, but is in total opposition to what Boucher and Inoussa (2014) obtained. Indeed, these authors concluded that new drivers should be assigned to the worst BMS scale.

4.3. Parameters analysis

Table 7 presents estimated values and standard errors for all β parameters, for some of the best models namely the NB1 and the HF-NBB models, and the BMS-NB1 model with $s = 11$, $\ell^* = 1$, and $\Psi = 6$. For the latter, we do not consider the variability of the structural parameters s, ℓ^*, and Ψ in the analysis. As already noticed by Boucher and Inoussa (2014), it is interesting to note the difference between the $\hat{\beta}$, particularly when comparing cross-section data models (such as the NB1) and the other two models that allow for merit rating. Estimated parameters $\widehat{\beta_5}$, $\widehat{\beta_6}$ and $\widehat{\beta_7}$, associated with the distance driven, show the largest difference between models. Smaller differences between those $\hat{\beta}$ can be observed between the HF-NBBeta model and the BMS-NB1 model. It means that the form of the premium penalties for claiming has an impact on the *a priori* rating.

Table 7: Estimated β (std.err.) parameters for specified models

Parameters	NB1		HF-NBB		BMS-NB1	
	Est.	Std. err.	Est.	Std. err.	Est.	Std. err.
$\widehat{\beta_0}$	−2.103	(0.036)	1.527	(0.106)	−2.356	(0.031)
$\widehat{\beta_1}$	0.037	(0.010)	0.037	(0.015)	0.031	(0.013)
$\widehat{\beta_2}$	−0.026	(0.014)	−0.036	(0.015)	−0.031	(0.014)
$\widehat{\beta_3}$	0.424	(0.018)	0.417	(0.019)	0.403	(0.018)
$\widehat{\beta_4}$	0.345	(0.015)	0.323	(0.016)	0.309	(0.016)
$\widehat{\beta_5}$	−0.579	(0.041)	−0.497	(0.071)	−0.481	(0.028)
$\widehat{\beta_6}$	0.453	(0.042)	−0.397	(0.070)	−0.385	(0.032)
$\widehat{\beta_7}$	−0.245	(0.050)	−0.222	(0.074)	−0.216	(0.039)
$\widehat{\beta_8}$	0.029	(0.013)	0.047	(0.016)	0.035	(0.016)

4.4. Predictive and covariance analysis

Table 8: Estimated parameters for specified models

Models	Parameter	Estimation	Std. Err
NB1	τ	0.067	(0.004)
HF-NBB	a	264.818	(12.881)
	b	5.500	(0.471)
	ν	0.900	(0.008)
BMS-NB1	τ	0.062	(0.003)
	δ	0.120	(0.004)

Table 8 shows the other estimated parameters for the same models. For the NB1 distribution, the value of τ mainly measures the overdispersion of the count distribution and does not model the dependence between number of claims. Indeed, the cross-section models suppose independence between all contracts and a null covariance between $N_{i,t}$ and $N_{i,t+j}$. This means that no merit rating is possible for this class of models. By comparison, for the more classic panel data models, such as the MVNB or the NBBeta, we established that the covariance between $N_{i,t}$ and $N_{i,t+j}$ is constant and does not depend on j. This results in the fact that the age of a claim is not being considered in the merit rating plan scheme.

The HF-NBB models, generalized quite directly by something similar to a Kalman filter approach, were designed to allow different weights depending on the age of the claim. An analytic solution to compute the covariance is too complex, so we simply simulate the values. Figure 4 shows two graphs for the covariance between the number of claims. The one on the left-hand side shows $\text{Cov}[N_{i,t}, N_{i,t+j}]$ for $t = 1, \dots, 10$ for lag $j = 1, \dots, 15$, while the other shows $\text{Cov}[N_{i,t-j}, N_{i,t}]$ for $t = 15$ and for lag $j = 1, \dots, 14$.

Figure 4: Covariances of the HF models with $\text{Cov}[N_{i,t}, N_{i,t+j}]$ on the left-hand side and $\text{Cov}[N_{i,15-j}, N_{i,15}]$ on the right.
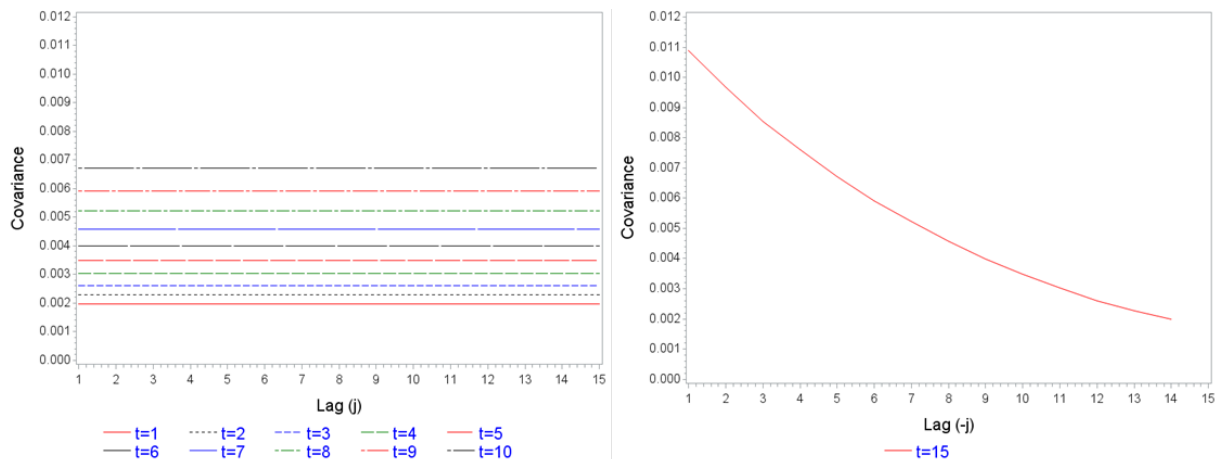
Analysis of the graphs indicates, for example, that the number of claims observed in the first contract of insured i will have the same impact on all future contracts. However, the covariance between claim counts will grow as t becomes larger. Therefore, the following relation holds:

$$\text{Cov}[N_{i,t_1}, N_{i,t_1+j}] \leq \text{Cov}[N_{i,t_2}, N_{i,t_2+j}], \text{for } t_1 < t_2.$$

The HF model was constructed to give unequal weight on the predictive premium depending on the age of the claim. Now, we see that the model does not give less weight to older claims, but gives more and more weight to newer claims. This pattern can be seen more clearly by the following proposition.

Proposition 4.1: *Based on an insured who has never claimed, the premium increase resulting from a claim will be higher if the insured has a long driving experience. In addition, the greater the driving experience of an insured, the greater the impact of a claim on the following year's premium.*

Proof. First, based on the HF-NBBeta model, let us compute the premium at time $t+1$ for an insured who did not report at all:

$$\pi_{i,t+1} = \lambda_{i,t+1} \left(\frac{v^t b}{(v)^t a + \sum_{k=1}^{t} (v)^k \lambda_{i,t-k+1} - 1} \right).$$

Then, we compute the premium of an insured who had only one claim at time $w < t$:

$$\pi_{i,t+1}^{(w)} = \lambda_{i,t+1} \left(\frac{v^t b + v^{t-w+1}}{(v)^t a + \sum_{k=1}^{t} (v)^k \lambda_{i,t-k+1} - 1} \right).$$

The impact of the claim at time w, or the increase in premium, is then calculated as:

$$\frac{\pi_{i,t+1}^{(w)}}{\pi_{i,t+1}} = \left(\frac{v^t b + v^{t-w+1}}{v^t b} \right) = 1 + \frac{v^{-w+1}}{b} = 1 + \left(\frac{v}{b} \right) v^{-w}$$

Because $v < 1$ for the HF model, which means that the premium increase will amplify as w grows. Moreover, we note that the increase does not depend on t or on the age of the claim $(t-w)$. In other words, the HF models suppose that the impact of a claim at time w will stay the same for all future contracts t.

In classic panel data models, such as the MVNB or the NBBeta models, it is impossible for an insured who claims at least once to ever have a premium equal to an insured who has never claimed, even after several years since the first and only claim. With the HF generalization and the introduction of the weight parameter v, we would expect that the impact of old claims gradually becomes insignificant in future premiums. We just saw that this is not the case. For the HF-NBBeta used with our data, we obtained $\hat{b} = 5.5$ and $\hat{v} = 0.9$, which means that a claim at time $w = 1$ will cause an 18.2% increase in the premium for all future contracts, while a claim at time $w = 15$ will always increase the premium by 78.9% compared with an insured

who never files an accident. Even if the fit of the HF models is interesting, this property of the model means that it cannot be seriously considered in practice: the longer the insureds' driving experience, the higher their penalty for filing an accident. The BMS-panel model does not have this property: for example, with the maximum number of levels $s = 11$, an insured without any claims for 11 consecutive years will have the same premium as a similar insured who did not report at all. We think that this property is more realistic, and more desirable for insurers.

The BMS-NB1 model with $s = 11, \ell^* = 1$, and $\Psi = 6$ supposes a linear relativity for claim penalty with $\hat{\delta} = 0.12$. Depending on the BMS level of each insured, it means that the premium will be equal to $1.12, 1.24, \ldots, 2.08, 2.20$ times the basic premium (level 1). For example, for an insured at level 2, a claim during the year results in a jump of $\Psi = 6$ levels, which represents an increase of approximately 64% of the premium. Conversely, a year without claim generates a premium reduction of 11%.

To evaluate more precisely the impact of the current level and the dependence between claim counts on different contracts, based on the result of Proposition 3.4, we compute the covariance of the BMS-panel model, as shown in Figure 5. To compare the covariances, we also included the covariance implied by the NBBeta model, which stays constant over time. Unlike in the HF model, the covariance depends on ℓ_t, the BMS level at time t. The dynamic property of the BMS-panel model can be observed: the dependence between annual contracts decreases as the lag grows. We think that this is one of the most important properties of a merit rating plan. The impact of ℓ_1 (the level of the BMS at time 1) on the covariance is clear. For example, as an insured located on $\ell_1 = 1$ cannot attain a lower level at time 2, the covariance between N_1 and N_2 is limited. As ℓ_1 grows, the covariance also increases, but because the impact of one claim on the next BMS level is $\Psi = 6$, we observe that the covariance begins to decrease for $\ell_1 > 5$. Finally, insured on $\ell_1 = s = 11$ are also limited, this time by the fact that they cannot attain a higher BMS level. As the lag between two numbers of claims grows, the covariance decreases and the impact of ℓ_1, the level of the BMS at time 1, vanishes.

Figure 5: Covariance of the BMS-panel depending on the level ℓ.
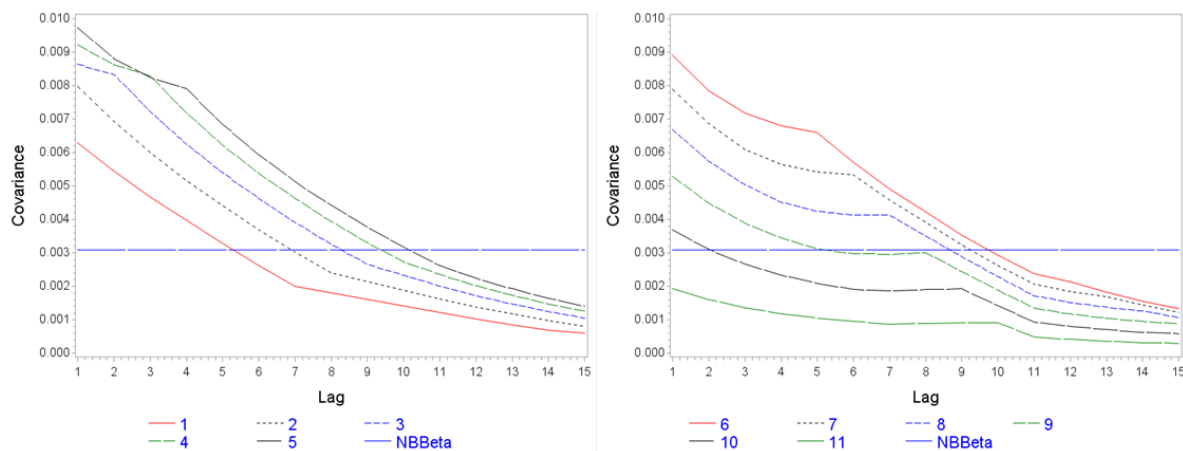
5. Conclusion

The BMS approach proposed in this paper generates a claim score that is easy to explain. The claim score is also easy to use because it summarizes the whole claim experience of an insured, by also taking into account the age of each claim. While we initially wanted to show that the BMS-panel model was practical and flexible, we were surprised to see that it also generated a much better statistical fit than some of the most popular counting distributions known in the actuarial literature. The decreasing covariance of the model, which assigns a lower weight to older claims, seems to best explain the interesting predictive power of the model. Indeed, the vast majority of existing models do not show such flexibility. The HF model proposed in this paper, which appeared to be a straightforward generalization of classic panel data models, seems an interesting solution. However, by analyzing the predictive premium of those HF models, we have shown that they become completely unpractical when a claimant's insurance history grows. Indeed, the impact of a single claim for an insured with a long driving experience becomes so important that no insurer would be interested in implementing such an approach.

Many studies and generalizations of the BMS-panel models are now possible:

- we can study a two-dimensional generalization of the BMS approach, where several types of claims (such as at-fault or no-fault accidents) could be modelled;

- achieving a better understanding of the multivariate dynamics in insurance; and

- many properties of the BMS, already developed and understood, can now be applied to BMS-panel models, e.g., asymptotic properties of the BMS, tools to compare merit rating plans, and hunger for bonus.

Acknowledgements

References

Abdallah, A., J.-P. Boucher, H. Cossette, and J. Trufin (2016). Sarmanov family of bivariate distributions for multivariate loss reserving analysis. *North American Actuarial Journal*, 20(2), 184–200.

Albrecht, P. (1985). An evolutionary credibility model for claim numbers. *ASTIN Bulletin*, 15(1), 1–17.

Bolancé, C., M. Denuit, M. Guillén, and P. Lambert (2007). Greatest accuracy credibility with dynamic heterogeneity: the Harvey–Fernandes model. *Belgian Actuarial Bulletin*, 7(1), 14–18.

Boucher, J.-P., and M. Denuit (2006). Fixed versus random effects in Poisson regression models for claim counts: case study with motor insurance. *ASTIN Bulletin*, 36, 285–301.

Boucher, J.-P., M. Denuit, and M. Guillén (2008). Models of insurance claim counts with time dependence based on generalisation of Poisson and negative binomial distributions. *Variance*, 2(1), 135–162.

Boucher, J.-P., and M. Guillén (2009). A survey on models for panel count data with applications to insurance. *Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales*, 103(2), 277–295.

Boucher, J.-P., and R. Inoussa (2014). *A posteriori* ratemaking with panel data. *ASTIN Bulletin*, 44(3), 587–612.

Bühlmann, H., and A. Gisler (2005). *A Course in Credibility Theory and its Applications*. Berlin, Heidelberg, and New York: Springer.

Choirat, C., and S. Raffaello (2012). Estimation in discrete parameter models. *Statistical Science*, 27(2), 278–293.

Denuit, M., X. Maréchal, S. Pitrebois, and J.-F. Walhin (2007). *Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Scales*. New York: Wiley.

Frees, E.W., R.A. Derrig, and G. Meyers (2014). *Predictive Modeling Applications in Actuarial Science*. Cambridge: Cambridge University Press.

Gilde, V., and B. Sundt (1989). On bonus systems with credibility scales. *Scandinavian Actuarial Journal*, 1989(1), 13–22.

Gourieroux, C., and J. Jasiak (2004). Heterogeneous INAR(1) model with application to car insurance. *Insurance: Mathematics and Economics*, 34(2), 177–192.

Harvey, A.C., and C. Fernandes (1989). Time series models for count or qualitative observations. *Journal of Business & Economics Statistics*, 7, 407–422.

Hammersley, J.M. (1950). On estimating restricted parameters (with discussion). *J. Roy. Statist. Soc. Ser. B*, 12, 192–240.

Jung, R., and R. Liesenfeld (2001). Estimating time series models for count data using efficient importance sampling. *AStA Advances in Statistical Analysis*, 4(85), 387–407.

Lemaire, J. (1995). *Bonus-Malus Systems in Automobile Insurance*. Boston: Kluwer.

McCullagh, P., and J.A. Nelder (1989). *Generalized Linear Models*. London: Chapman and Hall, 2nd ed.

Pinquet, J., M. Guillén, and C. Bolancé (2001). Allowance for the age of claims in bonus-malus systems. *ASTIN Bulletin*, 31(2), 337–348.

Shi, P., and E. Valdez (2016). Longitudinal modeling of insurance claim counts using jitters. *Scandinavian Actuarial Journal*, 2014(2), 159–179.

Shi, P., X. Feng, and J.-P. Boucher (2016). Multilevel modeling of insurance claims using copulas. *Annals of Applied Statistics*, 10(2), 834–863.

Winkelmann, R. (2010). *Econometric Analysis of Count Data*. Berlin and Heidelberg: Springer-Verlag, 5th ed.