A POSTERIORI RATEMAKING WITH PANEL DATA

BY

JEAN-PHILIPPE BOUCHER AND ROFICK INOUSSA

ABSTRACT

Ratemaking is one of the most important tasks of non-life actuaries. Usually, the ratemaking process is done in two steps. In the first step, a priori ratemaking, an a priori premium is computed based on the characteristics of the insureds. In the second step, called the a posteriori ratemaking, the past claims experience of each insured is considered to the a priori premium and set the final net premium. In practice, for automobile insurance, this correction is usually done with bonus-malus systems, or variations on them, which offer many advantages. In recent years, insurers have accumulated longitudinal information on their policyholders, and actuaries can now use many years of informations for a single insured. For this kind of data, called panel or longitudinal data, we propose an alternative to the two-step ratemaking approach and argue this old approach should no longer be used. As opposed to a posteriori models of cross-section data, the models proposed in this paper generate premiums based on empirical results rather than inductive probability. We propose a new way to deal with bonus-malus systems when panel data are available. Using car insurance data, a numerical illustration using at-fault and non-at-fault claims of a Canadian insurance company is included to support this discussion. Even if we apply the model for car insurance, as long as another line of business uses past claim experience to set the premiums, we maintain that a similar approach to the model proposed should be used.

KEYWORDS

A priori ratemaking, a posteriori ratemaking, bonus-malus systems, cross-section data, panel data, count data, Poisson, credibility, inductive probabilities.

1. INTRODUCTION

Modeling the number of claims and the cost of claims is an essential task for actuaries. Parametric modeling of these random variables in insurance allows identification of the claims process, helps in understanding the behavior of the

Astin Bulletin 44(3), 587-612. doi: 10.1017/asb.2014.11

insureds, and can be useful for solvency purposes. Parametric modeling also allows computing of premiums for ratemaking.

Traditionally in actuarial sciences, ratemaking was separated into two steps called a priori ratemaking and a posteriori ratemaking. By defining homogeneous classes of risk with exogenous information about each policyholder, the first part of the ratemaking process uses classification models (such as regression techniques) to compute a priori premiums that depend on the characteristics of the insureds. In a posteriori ratemaking, the insurers consider the past claim experience of each insured to update the a priori premiums. In doing this, insurers suppose that the number of claims reported in the past can be used to improve the estimation of the conditional expectation of the insurance cost for the following year.

This dual approach to ratemaking has a long history in actuarial sciences and is related to old techniques such as the minimum bias technique or the credibility theory. For practical reasons, an approach called *bonus-malus system* (BMS) theory was developed for a posteriori ratemaking in automobile insurance. The BMS has been used to estimate the premiums of each insured conditionally on the BMS level occupied. While the two-step approach to ratemaking is logical and coherent when working with cross-section data, we believe that the BMS may be subject to philosophical criticisms regarding the induction problem.

Nevertheless, when a panel or longitudinal data structure is available for actuaries, we think that the two-step approach to ratemaking should no longer be used. In this paper, we will show that a complete ratemaking structure should be used with a single step in the modeling. We will show that classic a priori ratemaking and a posteriori ratemaking are no longer coherent in a panel data setting. In the following sections, we will no longer estimate a priori premiums and a posteriori or BMS premiums separately, but instead estimate premiums that depend on risk characteristics and on claim experience simultaneously. The purpose of the model is to propose a technique for ratemaking. Consequently, the proposed models are not the best models for adjusting the data. More sophisticated models would certainly fit the data better. However, because of practical reasons (see Section 2.2), many of these models would not generate premiums that could be used directly in the industry, unlike the proposed models.

Because the premiums will be estimated with classic statistical techniques, the fitting of the data does not need a wide range of special methods to select the best a posteriori system, but only needs the use of standard statistical tests. As opposed to the BMS of cross-section data, the models proposed in this paper generate premiums based on empirical considerations and not on inductive probabilities.

Although the first sections of the paper seem to only review well-known theories, we believe that such a review is needed to highlight the conceptual problems in old techniques. In Section 2, the ratemaking process with cross-section data is reviewed. A priori and a posteriori ratemaking is clearly defined in the context where only independent insurance contracts are available for modeling. To solve practical problems in the computation of predictive premiums, actuaries have created BMSs that will also be explained in this section. Section 3 contains a short survey of models that can fit the panel data structure. We will show why the BMS methods developed for cross-section data can no longer be used in the panel data framework. In Section 4, we will propose a coherent way to compute the premiums by using what we called the bonus-malus system for panel data model (BMS-panel model). An empirical illustration with real insurance data illustrating how the BMS-panel model can be used will be presented in Section 5. The interpretation of BMS-panel models compared with other distributions will be discussed in Section 6. The induction problem of the credibility theory models will also be discussed in this section, where we will show that BMS-panel models partly avoid this problem. Section 7 concludes this paper.

1.1. Basis of ratemaking

Before the theoretical advances in statistical sciences and their application in actuarial sciences, a method called *the minimum bias technique* was used to find the premiums that should be offered to insureds with different risk characteristics. The idea behind these techniques was to find the parameters of the premiums that minimize the bias of the premium by iterative algorithms.

With the development of the generalized linear models (GLM; McCullagh and Nelder, 1989), actuaries now use statistical techniques to estimate the regression parameters for ratemaking. It has been shown that the results obtained from this theory are very close to the ones obtained by the minimum bias technique (see Brown, 1988, for example).

Using specific probability distributions for claim counts and the amounts of claims, the premium is typically calculated by obtaining the conditional expectation of the number of claims given the risk characteristics (even if other premium principles are possible), combined with the expected claim amount. In other words, the premium is calculated by multiplying the expected frequency of claims with the expected amount of the claim. In this paper, we will focus on the number of claims because merit rating plans are better suited to the frequency part of the premium, but the severity of claims can easily be added to the model (Lemaire, 1995).

2. Cross-Section Data

Historically, insurance companies began compiling their data in a cross-section form. This means that the database used for ratemaking analysis consisted of a list of insurance contracts. Each observation of the database contained information about each contract, such as the sex of the driver, the age of the driver, the type of car used, etc. This data structure was intended to ensure that each contract was supposed to be independent. An example of a cross-section database is provided in Table 1.

TABLE 1

CROSS-SECTION DATASET EXAMPLE.

Observations (i)	Sex	Civil Status	No. of Claims
1	W	S	1
2	W	S	0
3	W	M	2
4	M	M	3
5	M	M	1
6	W	M	0
...	...	...	...

To parametrically model the number of claims, conditionally on the covariates, the actuary has to select a count distribution. Commonly, the starting point for the modeling of count data is the Poisson distribution. The Poisson has the following probability function:

$$\Pr[N_i = n_i | X_i] = \frac{\lambda_i^{n_i} e^{-\lambda_i}}{n_i!}.$$

The characteristics of the insured that should influence their premium are included as regressors in the mean parameter of the count distribution. This exogenous information can be coded with binary variables. In insurance, an exponential function is commonly used to have $\lambda_i = t_i \exp(\boldsymbol{x}_i' \boldsymbol{\beta})$, where t_i represents the risk exposure of insured i. Because $E[N_i | X_i] = \lambda_i$, by knowing the characteristics of the insured, the actuary is then able to compute the premium for an insured.

In this framework, it is important to note that the actuary does not consider the past claim experience of the insureds in the modeling. Covariates that use past claim experience and merit-rating systems cannot be considered in this framework. Those models are discussed in Section 3. Because the premiums are calculated without using any merit rating, this form of ratemaking is usually called a priori ratemaking.

2.1. A posteriori ratemaking

Except for new drivers who have no driving experience, the insurance company is not only interested in the a priori premium, but rather in a premium that also considers driving experience. Formally, the actuary is interested in finding the predictive mean $E[N_{i,T} | n_{i,1}, \ldots, n_{i,T-1}, \boldsymbol{X}_{i,T}]$, where the vector $\boldsymbol{X}_{i,T} = \{X_{i,1}, \ldots, X_{i,T}\}$ corresponds to all characteristics of the insured from time $t = 1$ to $t = T$.

As mentioned earlier, because the cross-section database does not contain information about past claims experience, the actuary cannot compute the premium directly. To obtain the premium, the actuary must make some assumptions. A classic approach is the linear credibility framework (e.g., Bühlmann, 1967), where the actuary supposes that each insured has his own random heterogeneity component Θ_i that will affect all his future insurance contracts.

This component accounts for dependence among all the contracts of the same insured. One common interpretation of this random heterogeneity component is that it comes from the lack of some important classification variables (swiftness of reflexes, aggressiveness behind the wheel, consumption of drugs, etc.). In the credibility framework, these hidden features are supposed to be captured by this individual's random heterogeneity term. With cross-section data, to compute predictive premiums, the idea is to find the distribution of the random variable Θ_i and to use it to predict the future number of claims.

One way to obtain this distribution is to work with conditional probability distribution and add the random heterogeneity term Θ_i to the conditional distribution. The classic count model for the number of claims supposes a conditional Poisson distribution such as:

$$N_{i,t}|\Theta_i = \theta \sim \text{Poisson}(\lambda_{i,t}\theta). \tag{2.1}$$

The actuary then has to suppose a distribution for Θ_i. Many parametric distributions can be chosen for this distribution. The most popular distribution to use with a conditional Poisson distribution is the gamma (α, α) distribution (see Boucher *et al.*, 2007, for example) that leads to a negative binomial type 2 (negative binomial 2 or simply NB2) distribution with the following form:

$$\Pr[N_{i,t} = n] = \frac{\Gamma(\alpha + n)}{\Gamma(\alpha) n!} \left(\frac{\lambda_i}{\alpha + \lambda_i}\right)^n \left(\frac{\alpha}{\alpha + \lambda_i}\right)^\alpha. \tag{2.2}$$

It can be shown that the NB2 distribution has the following moments:

$$E[N_{i,1}|\boldsymbol{X}_{i,1}] = \lambda_{i,1,} \tag{2.3}$$

$$E[N_{i,T}|n_{i,1}, ..., n_{i,T-1}, \boldsymbol{X}_{i,T}] = \lambda_{i,T} \frac{\alpha + \sum_t^{T-1} n_{i,t}}{\alpha + \sum_t^{T-1} \lambda_{i,t}}. \tag{2.4}$$

Using the notations introduced earlier, the first expectation represents the a priori premium (for time $t = 1$), while the second expectation is called the predictive premium (for time $t \geq 2$). For the NB2 distribution, the predictive premium can be seen as the a priori premium times a correction for the past number of claims. This a posteriori ratemaking procedure penalizes insureds with claims and rewards insureds without claims.

In this modeling, we are not directly observing the predictive expectation. For cross-section data, the insurer only observes $n_{i,t}$, $X_{i,t}$ for $t = 1$. To calculate predictive premiums, we must use inference techniques (such as maximum likelihood estimators) with a count distribution (such as the NB2) to find $\widehat{\alpha}$ and $\widehat{\boldsymbol{\beta}}$. We replace α and $\boldsymbol{\beta}$ by their estimators and compute the premiums expressed in (2.3) and (2.4). In other words, the value of $\widehat{\alpha}$ is the key for time-dependent modeling.

For illustration purposes, the gamma distribution has been chosen for Θ_i. However, other distributions can be used with a conditional Poisson distribution, such as the inverse-Gaussian, the lognormal, or other mixed distributions (see Boucher *et al.*, 2007, for example). Note, however, that in these cases, the predictive premium obtained will not be necessarily equal to (2.4). Similarly, other conditional count distributions can also be chosen instead of the Poisson distribution (Boucher *et al.*, 2007).

Predictive premiums are strongly related to the linear credibility theory (see, for example, Bühlmann, 1967; Bühlmann and Straub, 1970; Hachemeister, 1975; Jewell, 1975). The objective of the linear premium of Bühlmann (1967) is to approximate the predictive premium. It has the following form:

$$P_{i,T}^{\text{LinCred}} = \nu_i \overline{N}_i + \pi_i P_{i,0}, \tag{2.5}$$

where $P_{i,T}^{\text{LinCred}}$ is the approximation of the predictive premium at time T, $\overline{N}_i$ is the average claim number, and $P_{i,0}$ is the a priori premium. By estimating the parameters in minimizing the quadratic error, Bühlmann (1967) was able to obtain closed form values of $\widehat{\nu}_i$ and $\widehat{\pi}_i$.

2.2. Practical application and bonus-malus systems

By working with cross-section data, even if the actuary cannot observe $E[N_{i,T}|n_{i,1}, ..., n_{i,T-1}, \boldsymbol{X}_{i,T}]$, they are able to construct a ratemaking procedure at the cost of making some assumptions.

In practice, however, by looking at the form of the predictive premium in (2.4), we can see that the insurer can encounter serious problems if he wants to use this equation. Indeed, it is not clear how the insurer would obtain some of the information needed to compute the premium. For example, how can the insurer obtain $\sum_t^{T-1} n_{i,t}$? Moreover, how could an insurer calculate $\sum_t^{T-1} \lambda_{i,t}$, which needs all $X_{i,t}$ from time $t = 1$ to $t = T - 1$? For an insured who is 50 years old, the insurer would need to know the insured's past 30 years of driving experience, as well as all his past risk characteristics (such as the kind of car he had, the places he lived, etc.) to compute all his last a priori premiums $\lambda_{i,t}$. Moreover, the form of the predictive premium supposed in the NB2 model does not put any time weight for past claims, meaning that an old claim increases the premium as much as a newer claim does. This is certainly not something the insurer would want.

To solve this practical problem, actuaries have created BMSs (Norberg, 1976). BMSs are class systems where the insured's level increases or decreases depending on the number of reported accidents. Table 2 illustrates an example of a BMS, where $P_{i,t}^{\text{BMS}}$ is the BMS premium, $\lambda_{i,t}$ can be considered as the a priori seen in (2.3). Level 1 (respectively s) offers the lowest (highest) premium.

https://doi.org/10.1017/asb.2014.11 Published online by Cambridge University Press

TABLE 2

ILLUSTRATION OF A BONUS-MALUS SYSTEM.

Level L	$P_{i,t}^{\text{BMS}}$	Relativities
s	$\lambda_{i,t} r_s$	r_s
...	...	...
...	...	...
...	...	...
l	$\lambda_{i,t} r_l$	r_l
...	...	...
...	...	...
...	...	...
1	$\lambda_{i,t} r_1$	r_1

A specific entry level is determined for new insureds. Each year, the level of each insured is adjusted depending on that person's claim experience. As already mentioned, only the number of claims is considered in determining the level of the BMS, but the amount of the claim could also be considered.

As for predictive premiums, the parameters first need to be estimated using the marginal count distribution, such as the NB2 distribution of (2.2). With the estimated parameters, there are many techniques to calculate the values of the relativities r_ℓ for each level $\ell = 1, ..., s$. We follow Denuit *et al.* (2007) to introduce the notations needed to calibrate a BMS:

- We note $p_{\ell_{t-1}\ell_t}(\psi)$ the one-year probability of the random variable L to go from BMS level ℓ_{t-1} to BMS level ℓ_t (for an insured with a claim frequency of ψ), where ℓ_t represents the level at time t. Consequently, we have:

$$p_{\ell_{t-1}\ell_t}(\psi) = \Pr[L(t; \psi) = \ell_t | L(t-1; \psi) = \ell_{t-1}],$$

- With all possible $p_{\ell_{t-1}\ell_t}(\psi)$, we construct a transition probability matrix. Formally, we have:

$$\boldsymbol{P}(\psi) = \begin{pmatrix} p_{11}(\psi) & p_{12}(\psi) & \dots & p_{1s}(\psi) \\ p_{21}(\psi) & p_{22}(\psi) & \dots & p_{2s}(\psi) \\ \dots & \dots & \dots & \dots \\ p_{s1}(\psi) & p_{s2}(\psi) & \dots & p_{ss}(\psi) \end{pmatrix}. \tag{2.6}$$

- We can show that for all $K = 0, 1, ...,$ we have:

$$\boldsymbol{P}^{(K)}(\psi) = \boldsymbol{P}^K(\psi),$$

meaning that the transition probability matrix over K time periods is simply the Kth power of the annual transition probability matrix $\boldsymbol{P}(\psi)$.

- We also introduce the probability vector

$$\boldsymbol{p}^{(K)}(\psi) = \begin{pmatrix} \Pr[L(K;\psi)=1] \\ \Pr[L(K;\psi)=2] \\ \dots \\ \Pr[L(K;\psi)=s] \end{pmatrix}^T$$

  that can be used to express the entry level condition. Indeed, the probability vector at time $t = 1$, i.e., $\boldsymbol{p}^{(1)}(\psi)$, is filled with zero except for the line corresponding to the entry level, which is filled with value of one.
- We then have the following relation:

$$\boldsymbol{p}^{(K)}(\psi) = \boldsymbol{p}^{(1)}(\psi)\boldsymbol{P}^{K}(\psi),$$

  meaning that the probability vector at time K is the product of the probability vector at time 1 and the K-step transition probability matrix.
- Finally, we have the limit theorem $\lim_{K\to\infty} \boldsymbol{P}^{(K)}(\psi) = \boldsymbol{\Pi}(\psi)$, which expresses the stationary distribution of the BMS.

To compute the values of the relativities, the system of Norberg (1976; see Denuit *et al.*, 2007 for details) that considers the a priori classification proposes the following equation:

$$r_\ell = \frac{\sum_{j=1}^{J} w_j \int \theta \boldsymbol{\pi}_\ell(\lambda_j\theta) g(\theta) d\theta}{\sum_{j=1}^{J} w_j \int \boldsymbol{\pi}_\ell(\lambda_j\theta) g(\theta) d\theta}, \ell = 1, \dots, s, \tag{2.7}$$

where w_j is the proportion of insureds having λ_j as a priori premium, J is the number of profiles in the portfolio, $g(\theta)$ is the heterogeneity distribution, and $\boldsymbol{\pi}_\ell(.)$ is the line component of $\boldsymbol{\Pi}(.)$. The probability matrix $\boldsymbol{\Pi}(.)$ can be considered as the stationary distribution of the BMS for the model of Norberg (1976). However, as proposed by Borgan *et al.* (1981), the probability matrix can also depend on the age of the policy because the distribution of insureds over all BMS levels is often far from the stationary distribution.

To calculate the relativities, the actuary first needs to select the form of the BMS. More precisely, before computing the relativities, the actuary must select the characteristics of the BMS and should select:

1. the number of levels s of the BMS;
2. the value of the penalty for each claim, i.e., the penalty structure of the BMS;
3. the entry level for new insureds.

Depending on the chosen form of the BMS, the values of r_ℓ, $\ell = 1, ..., s$ will not be the same. The actuary must then choose the best BMS available, i.e., the BMS that will offer the best prediction of the future costs of claims. To select the best BMS, a variety of methods has been proposed in the literature: the coefficient of variation, the mean-square error of prediction, the elasticity of the BMS, the speed of convergence, etc. (see Lemaire, 1995 or Denuit *et al.*, 2007

TABLE 3

LONGITUDINAL (OR PANEL) DATASET EXAMPLE.

	Year 2003				Year 2004		...	Year 2008	
Observations (i)	Sex	Civil Status	...	No. of Claims	...	No. of Claims	...	...	No. of Claims
1	W	S	...	0	...	1		...	0
2	M	S	...	1	...	2		...	1
3	W	S	...	0	...	0		...	.
4	M	M	...	0	...	0		...	0
5	W	S	...	0	...	0		...	1
6	W	M	...	0	...	0		...	0
...	...	...	...	...	...	...	...	...	...

for an overview of these techniques). Each of these methods has its advantages and their disadvantages. However, even more important is that these methods are clearly not similar and consequently do not necessarily select the same BMS as ideal for predicting the number of claims, or the total costs of the claims.

3. Longitudinal Information

In recent years, insurers have decided to accumulate longitudinal information on their policyholders. Sometimes, this additional information comes from a central authority. In this case, the insurers are able to ask the central authority for the past claim experience of new insureds entering the database. In an intuitive approach, the insured's past experience is used as a covariate (Gerber and Jones, 1975; Sundt, 1988). This model interprets past claims as a factor that changes the mean of the distribution. It would result in Poisson or Negative Binomial distributions with a mean equal to $\exp(X_{i,t}\beta + c_1 n_{i,t-1} + \ldots + c_{t-w} n_{i,t-w})$. However, as stated in Gourieroux and Jasiak (2004), the stationarity properties of this model cannot be established, and premiums for new insureds or for insureds with less than w years of experience then cannot be computed.

More generally, insurers now compile their own information about their insured. Consequently, actuaries can track a single insured over many years. Instead of working with cross-section data, the actuaries now work with what is called panel data or longitudinal data. Table 3 presents an example of a panel dataset.

In this data setting, an insured i is observed over T_i consecutive years (for simplicity, T will be used instead of T_i). The vector of random variables $(N_{i,1}, \ldots, N_{i,T})$ is the random counts to be modeled. Therefore, in contrast with cross-section data, dependence between all the contracts of the same insured can be concretely incorporated in the modeling.

There are many ways to model count data in a panel data setting. For non-Gaussian random variables, and particularly for discrete data, following the

https://doi.org/10.1017/asb.2014.11 Published online by Cambridge University Press

categories of Molenberghs and Verbeke (2005), we can classify three kinds of models for panel data:

- conditional models;
- marginal models;
- subject-specific models (e.g., using random effects).

Conditional models are models where the values of past realizations of the random variables to be modeled can be used directly. Generalizations for time series of count can be considered as members of this family of panel count data models. For example, the integer autoregressive model (INAR) and all related generalizations can also be used to model panel count data. Boucher *et al.* (2008) adjusted these models and expressed the form for the predictive premiums. The authors showed that for some specific models of this family, the fitting of insurance claim count data is poor, but it does not mean that no other conditional model can generate a good fitting.

The objective of marginal models for panel data count is to find a multivariate distribution that will fit all the contracts of a single insured. We can include the common shock model in this category (Boucher *et al.*, 2008, who shows that this model is not well suited for insurance panel data) or the generalized estimating equations (GEE) models. Purcaru *et al.* (2004) proposed a linear credibility approach to compute the predictive premium for this model.

The subject-specific models approach can be seen as a generalization of the heterogeneity approach seen in Section 2.1. Indeed, conceptually, the approach is similar. In a panel data setting, the dependence between all the contracts of the same insured still comes from a common random effect term Θ_i, such as

$$\begin{aligned}\Pr[n_{i,1}, ..., n_{i,T}|\boldsymbol{X}_{i,T}] &= \int \Pr[n_{i,1}, ..., n_{i,T}|\theta_i, \boldsymbol{X}_{i,T}]g(\theta_i)d\theta_i \\ &= \int \left(\prod_{t=1}^{T} \underbrace{\Pr[n_{i,t}|\theta_i, \boldsymbol{X}_{i,t}]}_{\text{conditional distribution}}\right) \underbrace{g(\theta_i)}_{\text{random effects distribution}} d\theta_i.\end{aligned}$$

The same combination as in Section 2.1, i.e., a conditional Poisson $(\lambda_{i,t}\theta)$ with a gamma(α, α) for random effects distribution, can be chosen. In this case, we obtain a distribution that can be seen as a generalization of the NB2 distribution:

$$\Pr[N_{i,1} = n_{i,1}, ..., N_{i,T} = n_{i,T}|\boldsymbol{X}_{i,T}] = \\ \left(\prod_{t=1}^{T} \frac{(\lambda_{i,t})^{n_{i,t}}}{n_{i,t}!}\right) \frac{\Gamma(n_{i,\bullet} + \nu)}{\Gamma(\nu)} \left(\frac{\nu}{\sum_{i=1}^{T} \lambda_{i,t} + \nu}\right)^{\nu} \left(\sum_{i=1}^{T} \lambda_{i,t} + \nu\right)^{-n_{i,\bullet}}, \quad (3.1)$$

where $n_{i,\bullet} = \sum_{t=1}^{T} n_{i,t}$. This distribution is called multivariate negative binomial (MVNB). With this distribution, we can show that

$$E[N_{i,1}|\boldsymbol{X}_{i,1}] = \lambda_{i,1}$$

$$E[N_{i,T}|n_{i,1}, ..., n_{i,T-1}, \boldsymbol{X}_{i,T}] = \lambda_{i,T}\frac{\alpha + \sum_{t}^{T-1} n_{i,t}}{\alpha + \sum_{t}^{T-1} \lambda_{i,t}}.$$

Consequently, we obtain the same form of premiums as with the cross-section data (see (2.3) and (2.4)).

4. Bonus-Malus System Model

For the same practical reasons as the ones described in Section 2.2, i.e., the difficulty of compiling all past information for all insureds, actuaries can be tempted to use BMS. Indeed, with the parameters estimated using a joint distribution, such as the MVNB of (3.1), we have the values of $\widehat{\alpha}$ and $\widehat{\boldsymbol{\beta}}$. Consequently, λ_k, w_k, $g(\theta)$, and $\boldsymbol{\pi}_{\ell}(\lambda_k\theta)$ can be computed, and it seems easy to estimate the relativities of a BMS, for example using (2.7).

We think that doing this would be a conceptual error. Indeed, we cannot go from the maximum likelihood estimates (MLE) of the joint distribution of $(N_{i,1}, \ldots, N_{i,T})$ to the BMS (panel data) as we did with cross-section data. By using panel data, we are fitting all our parameters using the joint distribution (3.1), which can be decomposed into

$$\begin{aligned}&\Pr(N_{i,1}, N_{i,2}, ..., N_{i,T}|\boldsymbol{X}_{i,T}) =\\ &\Pr(N_{i,1}|\boldsymbol{X}_{i,1}) \times \Pr(N_{i,2}|n_{i,1}, \boldsymbol{X}_{i,2}) \times ... \times \Pr(N_{i,T}|n_{i,1}, ..., n_{i,T}, \boldsymbol{X}_{i,T}).\end{aligned}$$

Consequently, for $t = 2, ..., T$, a predictive distribution is already used in the modeling, and $E[N_{i,t}|n_{i,1}, ..., n_{i,t-1}, \boldsymbol{X}_{i,t}]$ is then directly supposed in the joint distribution. In other words, if for practical reasons we do not want to use the predictive expected value as premium but instead use the BMS premium $P_{i,t}^{\text{BMS}}$ or the linear credibility premium $P_{i,t}^{\text{LinCred}}$, the joint distribution of $(N_{i,1}, \ldots, N_{i,T})$ should include this premium form for the estimation of all parameters.

Another way to justify this use would be to approximate a random effects count model by BMS. However, this means that we would be dealing with two consecutive approximations: from the data to the count model and from the count model to the BMS. It seems more logical and precise to fit a BMS from the data directly.

To do this with a BMS, the joint distribution that should be maximized to find the parameters will be expressed as (covariates $\boldsymbol{X}_{i,1}, ..., \boldsymbol{X}_{i,T}$ have been removed for simplicity)

$$\Pr(N_{i,1}, N_{i,2}, ..., N_{i,T}) = \sum_{\ell=1}^{s} \Pr(N_{i,1}, N_{i,2}, ..., N_{i,T}|L(1) = \ell_1) \Pr(L(1) = \ell_1). \quad (4.1)$$

https://doi.org/10.1017/asb.2014.11 Published online by Cambridge University Press

We can evaluate the joint probability $\Pr(N_{i,1}, N_{i,2}, \ldots, N_{i,T}|L(1) = \ell_1)$ by using the following result:

$$\begin{aligned}
&\Pr(N_{i,1}, N_{i,2}|L(1) = \ell_1) \\
&= \Pr(N_{i,1}|L(1) = \ell_1)\Pr(N_{i,2}|n_{i,1}, L(1) = \ell_1) \\
&= \Pr(N_{i,1}|L(1) = \ell_1) \\
&\quad \times \left(\sum_{y=1}^{s} \Pr(N_{i,2}|n_{i,1}, L(1) = \ell_1, L(2) = y)\Pr(L(2) = y|n_{i,1}, L(1) = \ell_1) \right) \\
&= \Pr(N_{i,1}|L(1) = \ell_1) \left(\sum_{y=1}^{s} \Pr(N_{i,2}|L(2) = y)\Pr(L(2) = y|n_{i,1}, L(1) = \ell_1) \right) \\
&= \Pr(N_{i,1}|L(1) = \ell_1)\Pr(N_{i,2}|L(2) = \ell_2),
\end{aligned}$$

where $\Pr(L(2) = y|n_{i,1}, L(1) = \ell_1) = 0$ for all y, except for $y = \ell_2$. We note ℓ_2, the BMS level obtained at time $t = 2$, by applying the transition rule of the BMS with $n_{i,1}$ claims, and $L(1) = \ell_1$. Consequently, $\Pr(L(2) = \ell_2|n_{i,1}, L(1) = \ell_1) = 1$. It can then be easily shown that

$$\Pr(N_{i,1}, N_{i,2}, \ldots, N_{i,T}|L(1) = \ell_1) = \prod_{t=1}^{T} \Pr(N_{i,t}|L(t) = \ell_t),$$

where $L(t)$, $t = 1, \ldots, T$ is the BMS level occupied by the insured at time t. We call this a *BMS-panel model*. For example, if we suppose that $N_{i,t}|L(t) = \ell_t, \boldsymbol{X}_{i,t}$ is Poisson distributed, the distribution expressed in (4.1) will be a Poisson-BMS-panel model (or simply Poisson-BMS) having the following probability function:

$$\Pr(N_{i,t} = n|L(t) = \ell, \boldsymbol{X}_{i,t}) = \frac{(\lambda_{i,t} r_\ell)^n e^{-\lambda_{i,t} r_\ell}}{n!}.$$

For more details about BMS, we refer the reader to Denuit *et al.* (2007). Finally, note that because we are no longer working with random effects (or fixed effects), but instead with a Markov chain that summarizes all the information on past realizations, we are no longer in the classic subject-specific family of models. Rather, we can say that the BMS-panel model is rather a member of the conditional family of models for panel data count.

4.1. Entry level

The past number of claims is thus replaced by the BMS level in the conditioning. Based on a specific pre-determined $L(1)$ value (BMS level at time 1), the idea of the modeling is to recreate the bonus-malus levels for the entire life of an insured, from the beginning of his experience in the insurance company to the present. Given the rules of transition from one level to another and the claim experience of each driver, the recomposition of the BMS levels for all insureds in the database is not a complex task.

However, it is only possible to obtain a driver's claim experience from recent years, not for his entire coverage period. Indeed, insurers do not have information on what happened to their insureds from before they were entered in the database.

Consequently, we will look for the distribution of the BMS level at time $t = 1$, i.e., the first year an insured appears in the database. Mathematically, we are looking for $\Pr(L(1) = \ell_1 | \boldsymbol{X}_{i,1})$.

The computation of the probability function is easy and direct for new drivers. In this case $L(1)$ corresponds to the entry level of the BMS, which is selected by the actuary in the construction of the BMS. The complex task is to find the distribution of $L(1)$ for drivers who have insurance experience before being entered in the insurance database. In this case, $L(1)$ will be related to the total number of insured years for each driver and will be calculated using equations expressed in Section 2.2.

If we suppose a specific structure for the BMS, we know the entry level of all insureds when they began to drive: it is the entry level selected by the actuary in the construction of the BMS. The idea is then to recreate all the possible events of each driver from his first year of driving to his first contract in the database. This probability is calculated using a small transformation of the transition probability matrix expressed in (2.6). Indeed, the probabilities of transition are computed using

$$\boldsymbol{P}_t(\lambda_{i,t}) = \begin{pmatrix} p_{11}(\lambda_{i,t}r_1) & p_{12}(\lambda_{i,t}r_1) & \dots & p_{1s}(\lambda_{i,t}r_1) \\ p_{21}(\lambda_{i,t}r_2) & p_{22}(\lambda_{i,t}r_2) & \dots & p_{2s}(\lambda_{i,t}r_2) \\ \dots & \dots & \dots & \dots \\ p_{s1}(\lambda_{i,t}r_s) & p_{s2}(\lambda_{i,t}r_s) & \dots & p_{ss}(\lambda_{i,t}r_s) \end{pmatrix}, \tag{4.2}$$

where r_ℓ, $\ell = 1, \dots, s$ are the relativities (to be estimated) for the BMS. With this matrix, we can obtain the distribution of $L(1)$ expressed in the vector as

$$\boldsymbol{p}_{\text{NLic}_i}(\boldsymbol{\lambda}_i) = \boldsymbol{p}^{(1)}(\lambda_{i,1}) \left(\prod_{t=1}^{\text{NLic}_i} \boldsymbol{P}_t(\lambda_{i,t}) \right) \tag{4.3}$$

where NLic is the driving experience (in years) for insured i at his first contract in the database. The probability vector $\boldsymbol{p}(.)$ was introduced in Section 2.2. Note that to compute the probability distribution of $L(1)$, we need to calculate the probability distribution of the number of claims from all the years in which each insured has been driving. This generates the problem of knowing the precise average frequency of insureds for all past years of insurance. In other words, to be perfectly coherent with our method, we must use the average frequencies $\lambda_{i,t}$

for all homogeneous classes of insureds for each year between 1950 to 2010. In this paper, to approximate all these frequencies, we supposed that the average frequency of all past years of insurance were equal to λ, the average frequency of our first insured year in the database. In other words, $\lambda_{i,t} = \lambda$, for all $i = 1, ..., n$, $t = 1, ... T_i$ in (4.2) and (4.3). This is clearly an approximation because the claims frequency is known to be cyclical.

Each insured must have both a constant premium and a distribution of BMS levels. Consequently, we must select a BMS level ℓ_1 and not a distribution of L_1. There are many ways to select the BMS level for an insured entering the database. The method that we propose uses the average of the first relativity, or formally,

$$\overline{r} = \sum_{\ell=1}^{s} \Pr(L(1) = \ell) r_{\ell}.$$

To determine the BMS level ℓ_1 of an insured, we choose the level j such as $r_j \leq \overline{r} < r_{j+1}$.

5. Empirical Illustration

We use this BMS-panel model with a sample of insurance data that comes from a major Canadian insurance company. Only private use cars have been considered in this sample. The unbalanced panel data contains information from 2003 to 2008. The sample contains 167,859 insurance contracts, which come from 57,037 policyholders.

We used 11 exogeneous variables, described in Table 4. For every policy we have the initial information at the beginning of the period. We are interested in modeling the number of claims for the collision coverage. We divided the claims into two categories: at-fault ($N^{(1)}$) and non-at-fault ($N^{(2)}$) claims. The average claim frequencies are 2.874% ($N^{(1)}$) and 3.741% ($N^{(2)}$). The maximum number of claims is three for both types of claims.

For some provinces in Canada, insurers can consult the claim files of all their insureds for the last six years. This means that the insurer is able to obtain the past number of claims for the last six years for each new insured. Using the variable indicating the number of years since the insured obtained his licence (NLic), we used this information in the BMS-panel model. As explained in Section 4.1, we wanted to model $L(1)$, the BMS level of an insured in his first year of coverage in the database. Because we can determine the number of claims for the last six years, we can model $L(-x + 1)$, with $x = \min(6, \text{NLic})$, the BMS level of an insured six years before the beginning of the database (or when he started driving).

https://doi.org/10.1017/asb.2014.11 Published online by Cambridge University Press

TABLE 4

BINARY VARIABLES SUMMARIZING THE INFORMATION AVAILABLE ABOUT EACH POLICYHOLDER.

Variable	Description
X1	Equals 1 if the insured is between 16 and 25 years old.
X2	Equals 1 if the insured is between 26 and 60 years old.
X3	Equals 1 if the vehicle is 0 years old.
X4	Equals 1 if the vehicle is 1–3 years old.
X5	Equals 1 if the vehicle is 4–5 years old.
X6	Equals 1 if the insured has a house.
X7	Equals 1 if there is only one driver.
X8	Equals 1 if there are two drivers.
X9	Equals 1 if the insured is single.
X10	Equals 1 if the insured is divorced.
X11	Equals 1 if the insured does not have any minor convictions.

5.1. Form of the BMS-panel model

The BMS-panel model is very flexible and can handle complex ratemaking structures. Indeed, the construction of the BMS-panel model is similar to what was described in Section 4, even taking into account dependence between types of claims. In our empirical illustration, we are interested in at-fault ($N^{(1)}$) and non-at-fault ($N^{(2)}$) claims. Modeling these types of claims using a parametric model could require special estimation techniques, particularly if the actuary wants to consider time weight of claims, as the BMS-panel model proposes. BMS-panel models avoid such situations.

Similarly, BMS-panel models can be used with complex penalty structures, where each claim can be penalized differently. In such a case, only the transition probability expressed in equation (4.2) will change. In our example, we have to deal with legal constraints. Indeed, in Canada, we cannot increase the premium for non-at-fault claims. To overcome this difficulty, we decided to eliminate the no-claim bonus at the end of the year if the insured had a non-at-fault claim. This leaves us with three possibilities:

1. an insured has no claims: his BMS level will be lowered by 1;
2. an insured has at-fault claims: his BMS level will increase by the number of claims times the decided penalty (XX);
3. an insured has no at-fault claims but has non-at-fault claims: his BMS level will stay the same.

We called this BMS-panel model structure –1/0/+XX.

5.2. Parametric relativities

We have to compute s relativities, i.e., r_ℓ for $\ell = 1, ..., s$. To avoid unwanted situations, where $r_i > r_j$ for $i < j$, we used linear relativities (Gilde and Sundt,

https://doi.org/10.1017/asb.2014.11 Published online by Cambridge University Press

1989) for the BMS. More precisely, we used two forms of linear relativities:

$$r_l^{\text{BMS1}} = 1 + \tau(l - 1), \quad (5.1)$$

$$r_l^{\text{BMS2}} = \begin{cases} 1 \text{ for } l = 1 \\ \delta + \tau(l - 1) \text{ otherwise} \end{cases}, \quad (5.2)$$

where $l = 1, \ldots, s$ represents the BMS level. These two linear forms imply that $r_1 = 1$, meaning that the BMS relativities can be seen as the premium ratio with an insured at level 1.

To use the joint distribution (4.1), we also have to choose the conditional distribution for $N^{(1)}$ and $N^{(2)}$. We chose two distributions: the Poisson and the NB2 distributions. Many count distributions can be used; see Boucher *et al.* (2007) for a list of possible distributions for claim counts.

The choice of the count distribution specifies the joint distribution of $\boldsymbol{N}_{i,t} = \{N_{i,t}^{(1)}, N_{i,t}^{(2)}\}$ for $t = 1, \ldots, T_i$:

$$\Pr(\boldsymbol{N}_{i,1}, \boldsymbol{N}_{i,2}, \ldots, \boldsymbol{N}_{i,T_i} | L(1) = \ell_1, \boldsymbol{X}_{i,T}) = \prod_{t=1}^{T_i} \Pr(\boldsymbol{N}_{i,t} | L(t) = \ell_t, \boldsymbol{X}_{i,t})$$

$$= \prod_{t=1}^{T_i} \Pr(N_{i,t}^{(1)} | L(t) = \ell_t, \boldsymbol{X}_{i,t}) \times \prod_{t=1}^{T_i} \Pr(N_{i,t}^{(2)} | L(t) = \ell_t, \boldsymbol{X}_{i,t}).$$

Note that the chosen count distribution is also related to the distribution of $\boldsymbol{p}_{\text{NLic}_i}(\lambda)$ (see equation 4.3).

5.3. Selecting the best BMS structure

As mentioned in Section 2.2, in the classic theory of calibrating BMS, the actuaries have to use many techniques to find the best BMS among all possible penalty structures, number of levels, entry levels, etc. With the BMS-panel models proposed here, all the characteristics of the BMS are selected using classic statistical fitting. Indeed, by using standard statistical tests, or simply by using profile loglikelihoods, the actuary can then determine

- the entry level;
- the penalty structure;
- the number of levels;
- the form of the relativities;
- the conditional distributions.

Therefore we rely on well-defined statistical theory to find the structure of the BMS instead of creating completely new metrics that were used only in actuarial sciences.

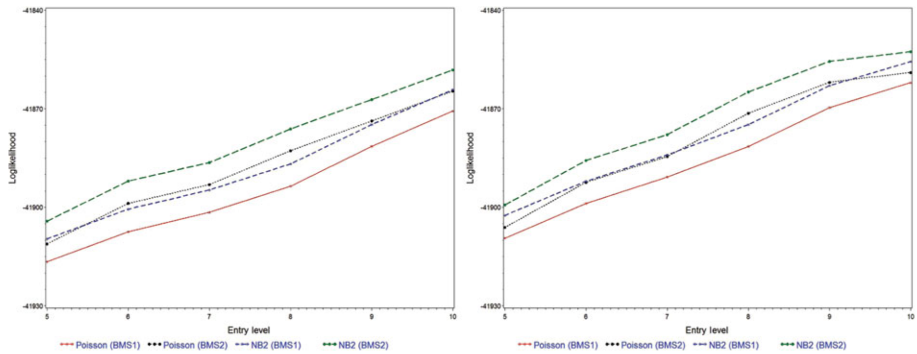

FIGURE 1: Loglikelihood for different entry levels, with penalties $XX = 1$ and $XX = 2$. (Color online)

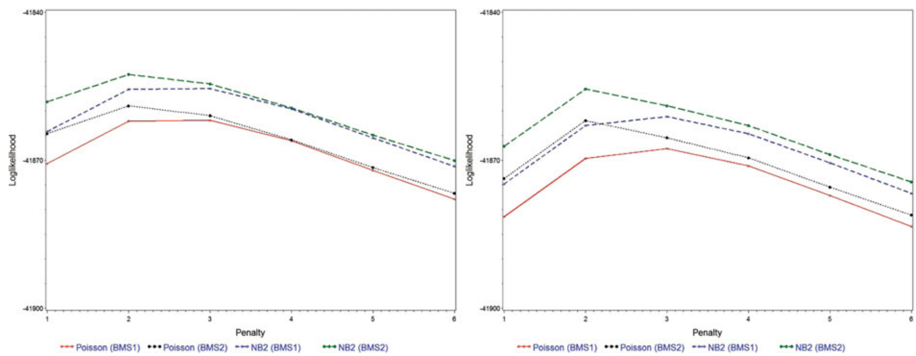

FIGURE 2: Loglikelihood for different penalties, with entry levels 10 and 9. (Color online)

It would have been possible to extend the choice of the best BMS by using loglikelihood analysis with a different number of levels. However, for simplicity, we restrict ourselves to BMS with 10 levels. Figure 1 illustrates the loglikelihoods obtained for different entry levels for penalty structures $XX = 1$ and $XX = 2$ (levels of penalty for each claim). Figure 2 illustrates the loglikelihoods obtained for different values of penalties XX, with entry levels 10 and 9.

Our analysis of loglikelihoods shows that for the linear form BMS1, the best system with 10 levels is a –1/0/+3 with an entry level at 10 for the Poisson distribution (P.-BMS1), and a –1/0/+3 with an entry level at 10 for the NB2 distribution (NB2-BMS1). For the other linear form (BMS2), the best systems with 10 levels have been shown to be a –1/0/+2 with an entry level of 10 for the Poisson distribution (P.-BMS2), and also a –1/0/+2 with an entry level of 10 for the NB2 distribution (NB2-BMS2).

Loglikelihood analysis also helps us select the best relativity structure. BMS2 of (5.2) converges to BMS1 of (5.1) for $\delta \to 1$. The value of $\widehat{\delta}$ can be tested to compare BMS1 and BMS2, for BMS with the same structure. In our example, for each specific count distribution, to compare the best BMS1 model with

https://doi.org/10.1017/asb.2014.11 Published online by Cambridge University Press

the best BMS2 model, we compare two BMS structures: a BMS1 with –1/0/+3 versus a BMS2 with –1/0/+2. In this case, we can compare their loglikelihood and choose the best models (both BMS-panel models have the same number of parameters). The BMS2 seems to offer a better fit for Poisson and NB2 distributions.

5.4. Covariates and a priori ratemaking

The classic ratemaking technique in insurance separates the estimation of the premiums into two components: a priori ratemaking and a posteriori ratemaking. It is well known that this two-step process ratemaking has the consequence of penalizing bad drivers' profiles twice (see Denuit *et al.*, 2007 for details). Indeed, because of their characteristics, their a priori premiums are higher than the premiums of other profiles. Also, because they claim more than other profiles, they will be more affected by merit ratings and BMS. Young and DeVylder (2000) propose linear credibility models to correct this problem.

The BMS-panel model proposed here does not split ratemaking into two components. BMS relativities and β parameters for the risk classification are estimated together. This is another advantage of this method. The estimated premiums for all profiles and for all levels of the BMS are coherent. However, it does not mean that some profiles will not pay more than they need to. It only means that the proposed BMS-panel model is sufficiently flexible to find ways to minimize this error.

We use the –1/0/+3 BMS-panel model with 10 levels for Poisson and NB2 distributions. Table 5 shows the estimated β parameters for Poisson-BMS models and for independent Poisson. Table 6 shows the same parameters for NB2 distributions. We can see that the parameters β_1 and β_2 of BMS-panel models change compared to independent Poisson and NB2 distributions. These parameters are used for young drivers. Because we select the BMS-panel model with an entry level at 10, young drivers need 10 years without claim to reach the best level of the BMS. Consequently, this is the reason why parameters for the age are directly affected.

Parameter β_9, which is used to identify the marital status of the insured, is also affected by the model. Compared with the independent Poisson distribution, we can observe a difference of about 15% to 20% for both $N^{(1)}$ and $N^{(2)}$. Because the BMS panel-model proposed here seems to primarily affect the premium of young drivers, we think that the value of the β_9 parameter is also related to the age of driver. Further, young drivers are not often married. The value of the parameter β_{11} increases relative to the independent Poisson distribution. Because it is used to identify drivers without convictions, we also think that the increasing value of β_{11} is also caused by the way the BMS-panel model deals with young drivers.

Other β parameters are also modified by the BMS-panel model, but the change is less important.

TABLE 5

ESTIMATED PARAMETERS FOR INDEPENDENT POISSON, POISSON-BMS1, AND POISSON-BMS2 MODELS.

	$N^{(1)}$						$N^{(2)}$					
	Independent Poisson		P.-BMS1		P.-BMS2		Independent Poisson		P.-BMS1		P.-BMS2	
	Estimate	Std. Err.	Estimate	Std. Err.	Estimate	Std. Err.	Estimate	Std. Err.	Estimate	Std. Err.	Estimate	Std. Err.
β_0	−3.3532	(0.086)	−3.4152	(0.091)	−3.3984	(0.090)	−3.3488	(0.092)	−3.4127	(0.084)	−3.3964	(0.087)
β_1	0.4672	(0.082)	0.0929	(0.080)	0.0545	(0.081)	0.5329	(0.080)	0.1590	(0.079)	0.1207	(0.079)
β_2	0.0500	(0.056)	−0.0064	(0.050)	−0.0137	(0.051)	0.3136	(0.051)	0.2579	(0.049)	0.2508	(0.049)
β_3	0.3657	(0.045)	0.3471	(0.044)	0.3473	(0.044)	0.2231	(0.039)	0.2051	(0.039)	0.2051	(0.038)
β_4	0.3118	(0.043)	0.3046	(0.042)	0.3047	(0.042)	0.1476	(0.038)	0.1404	(0.038)	0.1405	(0.037)
β_5	0.2653	(0.045)	0.2517	(0.043)	0.2519	(0.044)	0.1184	(0.038)	0.1049	(0.038)	0.1050	(0.038)
β_6	0.1998	(0.035)	0.1743	(0.035)	0.1715	(0.035)	0.0729	(0.031)	0.0467	(0.031)	0.0438	(0.031)
β_7	−0.4639	(0.056)	−0.4453	(0.057)	−0.4548	(0.056)	−0.2690	(0.050)	−0.2499	(0.053)	−0.2591	(0.049)
β_8	−0.3809	(0.059)	−0.3866	(0.059)	−0.3945	(0.056)	−0.1828	(0.049)	−0.1891	(0.054)	−0.1967	(0.050)
β_9	0.1040	(0.044)	0.0845	(0.042)	0.0812	(0.043)	0.0842	(0.039)	0.0645	(0.037)	0.0611	(0.037)
β_{10}	0.3189	(0.064)	0.3264	(0.064)	0.3299	(0.064)	0.2321	(0.056)	0.2394	(0.057)	0.2429	(0.057)
β_{11}	−0.1880	(0.059)	−0.1448	(0.059)	−0.1457	(0.062)	−0.1758	(0.056)	−0.1312	(0.051)	−0.1319	(0.051)

TABLE 6

ESTIMATED PARAMETERS FOR INDEPENDENT NB2, NB2-BMS1, AND NB2-BMS2 MODELS.

	$N^{(1)}$						$N^{(2)}$					
	Independent NB2		NB2-BMS1		NB2-BMS2		Independent NB2		NB2-BMS1		NB2-BMS2	
	Estimate	Std Err.	Estimate	Std Err.	Estimate	Std Err.	Estimate	Std Err.	Estimate	Std Err.	Estimate	Std Err.
β_0	−3.3520	(0.092)	−3.4138	(0.093)	−3.3971	(0.093)	−3.3478	(0.081)	−3.4120	(0.084)	−3.3954	(0.080)
β_1	0.4684	(0.081)	0.0934	(0.082)	0.0551	(0.087)	0.5331	(0.076)	0.1590	(0.081)	0.1206	(0.082)
β_2	0.0498	(0.050)	−0.0064	(0.051)	−0.0136	(0.052)	0.3135	(0.045)	0.2579	(0.049)	0.2508	(0.047)
β_3	0.3652	(0.044)	0.3468	(0.044)	0.3470	(0.044)	0.2227	(0.038)	0.2047	(0.039)	0.2047	(0.039)
β_4	0.3117	(0.042)	0.3047	(0.043)	0.3049	(0.043)	0.1475	(0.037)	0.1404	(0.038)	0.1406	(0.037)
β_5	0.2650	(0.044)	0.2515	(0.044)	0.2518	(0.044)	0.1181	(0.038)	0.1047	(0.039)	0.1049	(0.039)
β_6	0.2000	(0.036)	0.1744	(0.035)	0.1716	(0.034)	0.0729	(0.031)	0.0466	(0.031)	0.0437	(0.031)
β_7	−0.4643	(0.056)	−0.4461	(0.055)	−0.4556	(0.053)	−0.2692	(0.047)	−0.2499	(0.052)	−0.2592	(0.050)
β_8	−0.3813	(0.058)	−0.3874	(0.056)	−0.3951	(0.055)	−0.1829	(0.046)	−0.1890	(0.052)	−0.1967	(0.051)
β_9	0.1039	(0.043)	0.0843	(0.043)	0.0810	(0.043)	0.0842	(0.039)	0.0643	(0.038)	0.0610	(0.038)
β_{10}	0.3191	(0.064)	0.3266	(0.064)	0.3301	(0.063)	0.2321	(0.056)	0.2392	(0.057)	0.2428	(0.057)
β_{11}	−0.1883	(0.058)	−0.1452	(0.059)	−0.1461	(0.056)	−0.1763	(0.049)	−0.1317	(0.050)	−0.1326	(0.048)
α	0.3326	(0.180)	0.2985	(0.167)	0.2967	(0.182)	0.3646	(0.125)	0.3121	(0.149)	0.3110	(0.115)

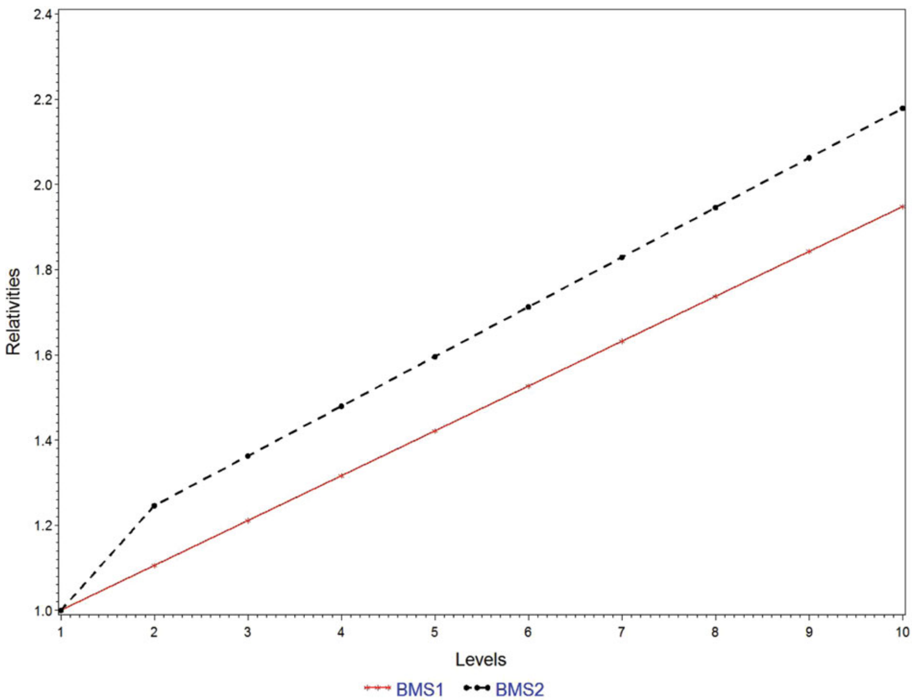

FIGURE 3: Relativities for the BMS1 (linear) and BMS2 (linear with intercept). (Color online)

With the selected models, the estimated relativities for the BMS are illustrated in Figure 3. Poisson-BMS and NB2-BMS models generate the same relativities for each of the BMS1/BMS2 levels.

Finally, as explained in Section 4.1, even if new drivers enter the BMS at level 10, an entry level must be assigned to all insureds entering the database. In the model we propose, the entry level is selected depending only on driving experience (NLic), but covariates (such as the sex) can also be used to select different entry levels. Figure 4 shows the selected level depending on NLic.

For example, the insurer will assign a BMS level of 3 to a new insured with 30 years of driving experience. However, in our application, the insurer can consult the last six years of driving experience for each insured entering the database. Consequently, to obtain the current BMS level for a first insurance contract in the company, the insurer will calculate the BMS level by supposing that the insured was on the third level six years ago, and will update the BMS level depending on the claim experience.

5.5. Selecting the conditional distribution

The conditional distribution for the BMS-panel model should be selected prudently. Some models, such as the NB2-BMS and the Poisson-BMS models, seem

https://doi.org/10.1017/asb.2014.11 Published online by Cambridge University Press

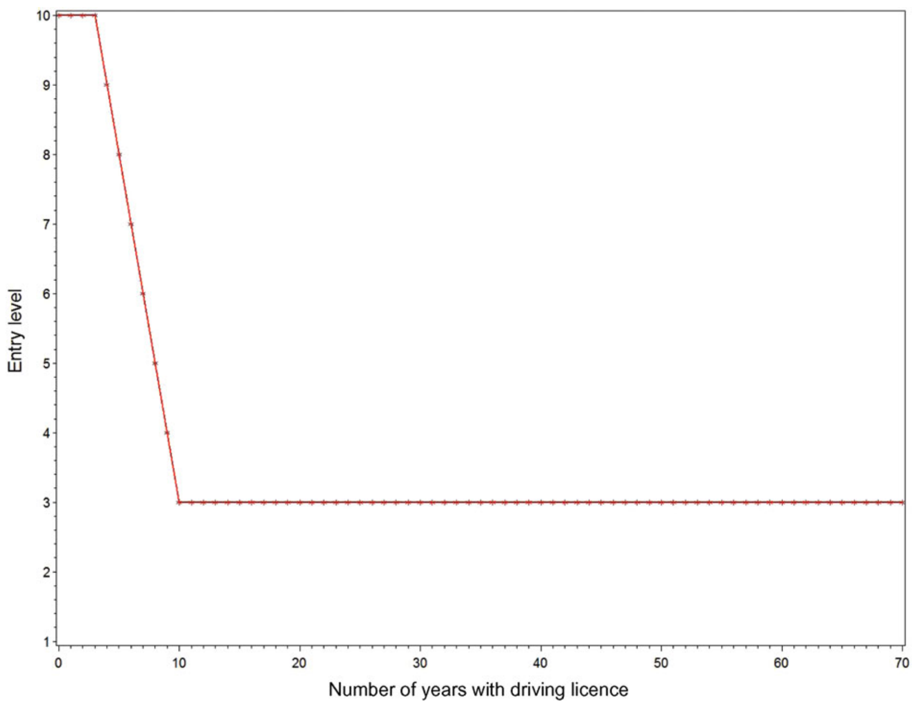

FIGURE 4: Entry Level for different years with driving licence. (Color online)

to be nested, because an NB2 distribution converges to a Poisson distribution when $\alpha \to 0$. However, these two BMS-panel models are only nested when they have the same number of levels, the same entry level, and the same penalty structure. For different forms of BMS, the NB2-BMS and the Poisson-BMS are no longer nested, and information criteria such as the AIC or the BIC should be used.

For similar forms of the BMS, the Poisson-BMS and the NB2-BMS are nested models because in this case, the NB2-BMS will converge to a Poisson-BMS when $\alpha \to 0$. This corresponds to the border of the parameter space of the NB2-BMS. It is well known that a problem with standard specification tests (the Wald or the log-likelihood ratio tests) happens when, as in our case, the null hypothesis is on the boundary of the parameter space (see Boucher *et al.*, 2007). When testing at a level δ, one must reject the H_0 hypothesis if the test statistic exceeds $\chi^2_{1-2\delta}(1)$, rather than $\chi^2_{1-\delta}(1)$, meaning that in this situation, a one-sided test must be used. Analogous results hold for the Wald test given that parameter distribution consists of a mass of one half at zero and a normal distribution for the positive values. Again in this case, the usual one-sided test critical value of $z_{1-\delta}$ should be used.

https://doi.org/10.1017/asb.2014.11 Published online by Cambridge University Press

In our example, comparing the p-value of α means that the Poisson–BMS2 is rejected against the NB2-BMS2. Therefore, of all the models we tested, the NB2-BMS2 is the best.

6. Discussion

The BMS-panel models presented in this paper allow the construction of premiums that are realistic and are directly applicable. Using an insurance database, the BMS-panel models allow us to directly calculate the premiums charged to insureds, which depends not only on their characteristics but also on their claim experience. We are clearly aware that the BMS-panel models are not the best models for adjusting the data. More sophisticated models such as the count distributions proposed in Boucher and Guillén (2009), the Tweedie model for the modeling of the pure premium of Smyth and Jorgensen (2002), the hierarchical model for the cost of the claims, and the number of claims for various coverages of Frees and Valdez (2008), or the Harvey-Fernandes model from Bolancé *et al.* (2007) would certainly fit the data better.

However, we think that the purpose of these models could serve to clarify the claims process (for example, Boucher *et al.* (2010) proposed a hunger for bonus interpretation for the zero-inflated count distribution) rather than account for ratemaking. The main purpose of ratemaking models should be to compute the most accurate premiums under practical constraints. In the complex models, the predictive premiums proposed (which are more precise than those of the BMS-panel models) cannot really be applied for practical reasons. We showed at the end of Section 3 that if an insurer cannot use the exact predictive expected value for premiums, it makes no sense to suppose other premium forms in the fitting of the model and in the inference step without including this premium form in the distribution.

We can refer to our empirical illustration of Section 5 to explain the situation. Indeed, even if it was shown by advanced statistical methods that there is time dependence between the number of claims of successive insurance contracts for non-at-fault accidents (justifying experience rating for this insurance coverage), the law does allow the insurers to increase premiums if an insured claims this kind of accident. However, if positive dependence between at-fault and non-at-fault claims exists, it is possible, at least partially, to correct this situation by increasing the penalties of at-fault claims. Only a model designed to compute the most accurate premiums can calculate this correction, while advanced statistical models (designed to find the best fit) will only find an accurate measure of time dependence between non-at-fault claims.

Finally, compared to the BMS that comes from cross-section data, the BMS-panel models avoid a part of the philosophical controversy regarding the induction problem. Although the subject was more controversial in the second half of the 20th century, this problem is still under discussion (for example, see Gelman and Shalizi, 2013, or Gelman and Robert, 2013 who discuss this problem

https://doi.org/10.1017/asb.2014.11 Published online by Cambridge University Press

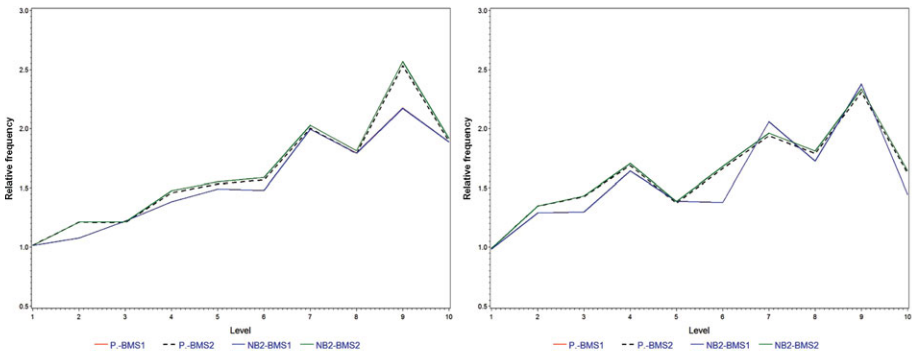

FIGURE 5: Observed relative claim frequency for at-fault and not-at-fault claims, depending for different BMS levels. (Color online)

for Bayesian inference purposes). In the old BMS application, the controversy comes from the prediction part of the model. Using distribution (2.1), the actuary can use any prior distribution for Θ, and by a correct application of the Bayes theorem, he will be able to compute predictive premiums. The choice of the prior and what it represents (ignorance, degree of belief, etc.) is controversial. However, more controversial is the use of inductive probability to calculate the premium. Note that the inductive probability was called probability$_1$ by Rudolf Carnap and *credibility* by the great philosopher Bertrand Russell. See Ryder (1976) who describes this controversy in an actuarial science context.

The purpose of this paper is not to solve this controversy or even discuss it, but rather to propose a model that partly avoids this problem in the calculation of premiums. One obvious reason for this is that the BMS-panel models no longer use unobserved random effects, nor Bayesian probability structures that need the subjective selection of a prior. However, we think that the real advantage of the BMS-panel models is the direct link between the observed statistics (from the database) and the proposed premiums.

In the BMS-panel models, the premiums that will be charged to insureds do not represent a projection of past data into the future, but a direct (but modeled) observation of claim experience. In other words, in the BMS-panel models the actuary will choose to charge a specific value for a premium because he can *observe* a claim experience in his database that is close to this value. For example, Figure 5 shows the relative observed average claim frequency for each BMS level. The frequency is calculated as $\sum_t \sum_i n_{i,t} / \sum_t \sum_i \lambda_{i,t}$. This was computed using the entry levels of Figure 4, meaning that the relativity forms BMS1 and BMS2 are partly used in the procedure. It would have been possible to obtain the same figure for unconstrained relativities. However, the purpose of Figure 5 is only to illustrate the difference between the use of the BMS-panel models compared with inductive probabilities.

From Figure 5, we can see that the estimated values of the BMS relativity can be justified by empirical considerations and not only by a Bayesian projection of past values into the future. Consequently, compared with classic BMS of Section 2.2, we think that the link between the insurance dataset and the premiums is easier to explain with BMS-panel models.

7. Conclusion

The traditional approach of calculating a priori premiums and BMS premiums (or linear credibility premiums) has been shown to be inconsistent for panel data. We demonstrate that BMS-panel models are more appropriate to compute the premiums than other complex models because they represent the real rating structure of an insurer. Consequently, the estimated premiums obtained with BMS-panel models can be used without other approximations. BMS-panel models have been shown to be very flexible: where any penalty structures can be proposed, and any count distributions can be used without complex numerical procedures. Using statistical theory, the BMS-panel model allows one to choose between all possible BMS structures without using new metrics that are only known in actuarial sciences. We also showed that the BMS-panel models are easier to understand. Finally, even if we applied the model for car insurance, as long as another line of business uses past claim experience to set the premiums, we think that a similar approach to the model proposed should be used.

Acknowledgements

The authors would like to thank an anonymous referee for his/her careful reading of the manuscript and the resulting comments that greatly helped to improve the paper. Jean-Philippe Boucher would like to thank the financial support from the Natural Sciences and Engineering Research Council of Canada.

References

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

Borgan, Ø., Hoem, J.M. and Norberg, R. (1981) A nonasymptotic criterion for the evaluation of automobile bonus systems. *Scandinavian Actuarial Journal*, **1981**(3), 265-178.

Boucher, J.-P., Denuit, M. and Guillén, M. (2007) Risk classification for claim counts: A comparative analysis of various zero-inflated mixed Poisson and hurdle models. *North American Actuarial Journal*, **11**(4), 110–131.

Boucher, J.-P., Denuit, M. and Guillén, M. (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., Denuit, M. and Guillén, M. (2009) Number of accidents or number of claims? An approach with zero-inflated poisson models for panel data. *Journal of Risk and Insurance*, **76**(4), 821–846.

BOUCHER, J.-P. and GUILLÉN, M. (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.

BROWN, R.L. (1988) Minimum bias with generalized LinearSSS models. *PCAS*, **LXXV**, 187–217.

BÜHLMANN, H. (1967) Experience rating and credibility. *ASTIN Bulletin*, **4**, 199–207.

BÜHLMANN, H. and STRAUB, E. (1970). GlaubwÄurdigkeit fÄur SchadensÄatze. *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker*, **70**, 111–133.

DENUIT, M., MARÉCHAL, X., PITREBOIS, S. and WALHIN, J.-F. (2007) *Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Scales*. New York: Wiley.

FREES, E. and VALDEZ, E.A. (2008) Hierarchical insurance claims modeling. *Journal of the American Statistical Association*, **103**(484), 1457–1469.

GELMAN, A. and ROBERT, C.P. (2013) "Not only defended but also applied": The perceived absurdity of Bayesian inference. *The American Statistician,* **67**(1), 1–5.

GELMAN, A. and SHALIZI, C.R. (2013) Philosophy and the practice of Bayesian statistics. *British Journal of Mathematical and Statistical Psychology,* **66**, 8–38.

GERBER, H.U. and JONES, D. (1975) Credibility formulas of the updating type. *Transactions of the Society of Actuaries*, **27**, 31–52.

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

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

HACHEMEISTER, C. (1975) Credibility for regression models with applications to trend. In *Credibility: Theory and Applications* (ed. P.M. Kahn) pp. 129–163. New York: Academic Press.

JEWELL, W. (1975) The use of collateral data in credibility theory: A hierarchical model. *Giornale dell Istituto Italiano degli Attuari*, **38**, 1–16.

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

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

NORBERG, R. (1976) A credibility theory for automobile bonus system. *Scandinavian Actuarial Journal*, **1976**, 92–107.

PURCARU, O., GUILLEN, M. and DENUIT, M. (2004) Linear credibility models based on time series for claim counts. *Belgian Actuarial Bulletin*, **4**, 62–74.

RYDER, J.M. (1976) Subjectivism - A reply in defense of classical actuarial methods (with discussion). *Journal of the Institute of Actuaries*, **103**, 59–112.

SMYTH, G. and JORGENSEN, B. (2002) Fitting Tweedie's compound Poisson model to insurance claims data: Dispersion modelling. *ASTIN Bulletin*, **32**(1), 143–157.

SUNDT, B. (1988) Credibility estimators with deometric weights. *Insurance: Mathematics and Economics*, **7**, 113–122.

YOUNG, V. and DEVYLDER, E.F. (2000) Credibility in favor of unlucky insureds. *North American Actuarial Journal*, **4**(1), 107–113.

JEAN-PHILIPPE BOUCHER (Corresponding author)
Quantact/Département de mathématiques, UQAM, Montréal, Québec, Canada
E-Mail: boucher.jean-philippe@uqam.ca

ROFICK INOUSSA
Quantact/Département de mathématiques, UQAM, Montréal, Québec, Canada
E-Mail: inoussa.rofick_ayinde@uqam.ca