BONUS-MALUS SCALE MODELS: CREATING ARTIFICIAL PAST CLAIMS HISTORY

A PREPRINT

Jean-Philippe Boucher
Chaire Co-operators en analyse des risques actuariels
Departement de mathematiques
Universite du Quebec a Montreal
boucher.jean-philippe@uqam.ca

January 4, 2022

ABSTRACT

In recent papers, Bonus-Malus Scales (BMS) estimated using data have been considered as an alternative to longitudinal data and hierarchical data approaches to model the dependence between different contracts for the same insured. Those papers, however, did not discuss in detail how to construct and understand BMS models, and many of the BMS's basic properties were not discussed. The first objective of this paper is to correct this situation by explaining the logic behind BMS models, and by describing those properties. More particularly, we will explain how BMS models are linked with simple GLM models that have covariates associated with the past claims experience. This study could help actuaries to understand how and why they should use BMS models for experience rating. The second objective of this paper is to create artificial past claims history for each insured. This is done by combining recent panel data theory with BMS models. We show that this addition significantly improves the prediction capacity of the BMS, and provides a temporary solution for insurers who do not have enough historical data. We apply the BMS model to real data from a major Canadian insurance company. Results are analyzed deeply to identify specific aspects of the BMS model.

Keywords Experience Rating, Count Data, Bonus-Malus Systems, Panel Data, Artificial Data

1 Introduction

Insurers have long known that policyholders who make claim will tend to make claim again in the future. There are several explanations for this. Some policyholders exhibit riskier behavior than others, others live in more disaster-prone regions, and some insured property is more susceptible to damage. The individual characteristics of each insured, often used as segmentation variables in regression models, may partly explain this situation. However, many of these character-defining elements simply cannot be measured and used in pricing. For example, a negligent or reckless insured will make more claims than a conscientious and attentive insured. Thus, the past claims experience can be used to approximate the effect of these unmeasurable characteristics on the premium.

Insurers also justify experience-rating models with the fact that normal policyholders will often have a fear of making claims from their insurer. When insureds have to make a claim, they often realize that it is very difficult. Consequently, these policyholders often become more willing to make claim in the future, and report accidents that they would not

have made claims for before, or by being less careful, knowing that the insurer will compensate them without issues and that the consequences of making a claim are not so grave (moral hazard). We can therefore see that from the insurer's point of view, it is quite important to put in place a rating structure that "penalizes" insureds who make claims by increasing their premium, and by rewarding insureds who do not claim. By sending the message that a claim will impact the premium, the insurer makes clear that it wants its insureds to be cautious and to not make claim when they experience minor accidents.

The general idea of any experience-based ratemaking model is quite simple: the insurer computes the premiums of each insured based on their past claims experience. Many approaches have been considered to achieve this goal, starting long ago with the individual credibility models of Bühlmann (1967) or Albrecht (1985), for example. However, credibility models were often difficult to use in practice. The Bonus-Malus Scales (BMS) model, popularized by Lemaire (2012), was shown to be a great alternative. A BMS corresponds to a class-system with a finite number of levels, where a relativity is assigned to each level. Depending on the transition rule of the BMS, an insured usually moves down by a level if he does not claim during his contract, and moves up a specific number of levels for each claim made. The insured's new level at the end of the year is then used to compute the next annual premium. BMSs can be easily generalized ny adding other transition rules between levels, where the number of consecutive years without making a claim can be used or by using the cost of each claim, for example. Actuaries traditionally used aggregate data and transition matrices based on assumptions about the heterogeneity distribution to find the level relativities (see Lemaire (2012) or Denuit et al. (2007) for details).

With detailed data about each contract for each insured which has become available for insurers, new longitudinal models have been proposed to improve and generalize credibility models. Complex approaches have been proposed to model claim counts: models based on series of correlated random effects (Bolancé et al. (2007), Abdallah et al. (2016) and Pechon et al. (2019b)), jitter models (Shi and Valdez (2014)), models with pair copulas or multiple hierarchical copulas (Shi and Yang (2018) and Shi et al. (2016)), time series for count data (Gourieroux and Jasiak (2004), Bermúdez et al. (2018) or Bermúdez and Karlis (2021)), etc.

However, like using credibility models in a practical context, these longitudinal models are often difficult for insurers to implement. To adapt to the new reality, Boucher and Inoussa (2014) tried to see how the BMS models theory could be corrected and generalized using the longitudinal insurance data now available. Boucher and Pigeon (2019) subsequently showed that these new BMS approaches using panel data were not only simple and easy to use but that they offer better fit statistics and predictive measures than those obtained with many advanced panel data models. Finally, Verschuren (2021) generalized the BMS approach with more flexible estimation methods, using generalized additive models (GAM) theory.

1.1 Objectives of the Paper

Despite the generalizations of those recent approaches, the general way BMS models are introduced and how they can now be used with panel or hierarchical insurance data remains confusing. Indeed, even if BMS models have been correctly defined, it is not always easy to understand how the BMS model might compare to standard pricing techniques. The **first objective** of this paper is to correct this situation by proposing the reconstruction of the whole BMS approach. By using a step-by-step methodology, we think it will help actuaries to understand precisely what the BMS approach is, and how and why BMS should be used in practice. This way of approaching BMS will also help to more precisely identify how the BMS models can be generalized and improved in the future.

After better defining the BMS model, the **second objective** of this paper is to create artificial past claims history for each insured. Indeed, a recurring practical problem with experience rating models is the lack of loss history for some insurers. We show that creating an artificial past claim history can be a temporary solution before waiting to have enough past years to estimate the full model. The artificial past claims history is created by combining panel data

models with BMS models. We show that this generalization of the BMS model significantly increases the prediction capacity of the BMS model.

1.2 Structure of the Paper

The paper has the following structure. In Section 2, we will carefully explain how we can construct BMS models with panel data information where basic ratemaking models that use standard GLM theory will be linked with BMS. A detailed description of the BMS will be provided, as well as a graphical visualization of all BMS parameters. In Section 3, the BMS model will be applied to real insurance data from a major Canadian insurance company. In Section 4, we will see how BMS can be generalized by creating artificial past claims history for each insured. The last section concludes with a list of potential generalizations that should be developed in the future to improve BMS models.

2 Experience Rating

2.1 General Definitions

Experience Rating and *a posteriori* ratemaking refer to ratemaking models that use past claims information to predict the future total amount of claims (also known as "loss costs"). In other words, the idea of experience rating is to compute a premium for insured i, for a contract of period T, that will consider all the insured's past insurance contracts or past claims experience. Formally, by supposing that the expected value corresponds to the premium (see Denuit et al. (2007) or Frees et al. (2014) for details about the link between random variables and the premium), it means that we are looking to compute:

$$E[S_{i,T}|\boldsymbol{S}^*_{i,(1:T-1)}, \boldsymbol{X}_{i,(1:T)}] \tag{1}$$

where :

- $S_{i,T}$ is a general random variable related to claim cost. Usually, it will correspond to the number of claims, the severity of claims or the total loss costs of contract T of insured i.
- $\boldsymbol{S}^*_{i,(1:T-1)}$ is a vector containing detailed past claim experience for insured i, through contract 1 to $T-1$. For example, this vector could include the cost of each claim, the type of claims, the date of each claims, etc.
- $\boldsymbol{X}_{i,(1:T)}$ is a vector containing all covariates used in the ratemaking, from contract 1 to $T-1$. This usually corresponds to information about the age of the insured, the marital status of the insured, etc. The vector also contains $X_{i,T}$ the vector of actual covariates of insured i for contract T.

Other specific terms must be correctly defined to avoid confusion and to construct a more easily understood approach to ratemaking. A **policy** is usually associated with a single **insured**. In insurers' databases, a policy is usually identified by a unique number. A policy is often made up of several insurance **contracts**. An insurance contract is also often referred to as an **insurance term** and it is usually one year long. Insurance contracts are sometimes shorter, for example three or six months. Some insurance policies contain only one term, but a significant portion of policies contain multiple contracts. An insurance contract can also contain several **items**, such as houses or vehicles. Finally, for each item, several **coverages** can apply.

Mathematically, using the standard approach to ratemaking, the premium P of a single insurance contract of a specific policy can be expressed as:

$$P \quad = \quad E[S] = \sum_{c=1}^{C} P_c = \sum_{c=1}^{C} E[S_c] = \sum_{c=1}^{C}\sum_{j=1}^{I_c} P_c^{(j)} = \sum_{c=1}^{C}\sum_{j=1}^{I_c} E[S_c^{(j)}]$$

where C is the number of coverages, I_c is the number of items for coverage c (for $c = 1, \ldots, C$), and $S_c^{(j)}$ is the total loss cost of item j for coverage c. To avoid working with too many subscripts, we will use a simplified approach in this section and will only work with P_c for $c = 1$ for each insured i. In other words, we assume that we will analyze only one coverage per contract. Generalizing to other items and other coverages will be discussed later.

2.2 Constructing the Model

We try to estimate the joint distributions of $S_{i,t}$ between all contracts $t = 1, \ldots, T_i$ of the same insured i. This kind of modeling approach is often called panel data modeling, or longitudinal modeling, which are much more complex than univariate modeling. In the scientific literature, we find a number ways to model panel data:

1. The Bayesian approach, where a common random effect affects all the contracts of the same insured. The classic credibility model of Bulhmann, based on a credibility factor Z, can be seen as a Bayesian approach.
2. Copulas can be used to model the dependence between random variables;
3. A Bonus-Malus Scales approach;
4. etc.

We used risk characteristics as covariates in regression models, so the computed premium of each insured can differ depending on sex, age, the value of the goods to be insured, etc. By the same token, we could use past insured experience as a covariate to compute the premium. For an insured i with $T-1$ years of experience, we could then use the following form:

$$E[S_{i,T}|\boldsymbol{S}^*_{i,(1:T-1)}, \boldsymbol{X}_{i,(1:T)}] = \mu_{i,T} = g(X'_{i,T}\boldsymbol{\beta} + W'_{i,T}\boldsymbol{\gamma})$$

for a link function $g(.)$, where:

- $X_{i,T}$ is the vector of actual covariates of insured i for contract t;
- $\boldsymbol{\beta}$ is a vector of parameters for covariates $X_{i,T}$ to be estimated;
- $W_{i,T}$ is a $k \times 1$ vector containing all useful past claim experience that could be used to predict $S_{i,T}$.
- $\boldsymbol{\gamma}$ is a $k \times 1$ vector of experience rating parameters to be estimated.

For example, as we commonly see in the literature, we can consider the past number of claims as good predictors when computing the premium. Other past information about claims, such as the cost of the claims, the number of claims from a specific coverage, the number of claims higher than a certain amount, etc. can also be used. However, for the sake of simplicity, we will focus on $W'_{i,t} = (n_{i,T-1}, n_{i,T-2}, \ldots, n_{i,1})$, meaning we use the number of claims for each $T-1$ past annual contracts. In this case, the predictive expected value of contract T of insured i could be expressed as:

$$\mu_{i,T} = \exp(X'_{i,T}\beta + \gamma_1 n_{i,T-1} + \gamma_2 n_{i,T-2} + ... + \gamma_{T-1} n_{i,1}),$$

where parameters $\gamma_1, \gamma_2, \ldots, \gamma_{T-1}$ are used to measure the impact of past claims. However this approach would involve using many parameters if T is large. We can probably do better. One of the purposes of statistical theory is to summarize information. One classic approach employed by the actuarial community is to use a summary of past claims. For example, we can instead use:

$$\mu_{i,T} = \exp(X'_{i,T}\beta + \gamma n_{i,\bullet})$$

where, for insured i, $n_{i,\bullet} = \sum_{t=1}^{T-1} n_{i,t}$ corresponds to the total number of past claims.

The major problem with this approach is that we cannot differentiate new insureds from insureds with many years of experience. Indeed, like a good insured without claims, a new insured will also have $n_{i,\bullet} = 0$. However, not

having claims is not the same as not having past insurance experience. Actuaries know that insureds without insurance experience claim a lot more than the average. Consequently, they should not be confused with experienced insureds without claims (who usually claim less than other insureds). To distinguish between these two types of insureds, by introducing the indicator variable $\kappa_{i,t} = I(n_{i,t} = 0)$, a solution would be to use:

$$\mu_{i,T} = \exp(X'_{i,T}\beta + \gamma_0(-\kappa_{i,\bullet}) + \gamma_1 n_{i,\bullet}) \tag{2}$$

where $\kappa_{i,\bullet} = \sum_{t=1}^{T-1} \kappa_{i,t} = \sum_{t=1}^{T-1} I(n_{i,t} = 0)$ is the sum of policy periods without claims for insured i. Using $-\kappa_{i,\bullet}$ as the covariate instead of $\kappa_{i,\bullet}$ will have a purpose later. With this model, we can differentiate new insureds from insureds without claims. Indeed, if we suppose we are dealing with a new insured, the insured will have $\kappa_{1,\bullet} = n_{1,\bullet} = 0$. An insured with insurance experience, would also have $n_{2,\bullet} = 0$, but will have $\kappa_{2,\bullet} > 0$.

2.3 Kappa-N Model

Another way of understanding the mean parameter of the model with $\kappa_\bullet$ and $n_\bullet$ is to rewrite $\mu_{i,t}$ as follows:

$$\begin{aligned}
\mu_{i,t} &= g(X'_{i,t}\beta + \gamma_0(-\kappa_{i,\bullet}) + \gamma_1 n_{i,\bullet}) \\
&= g(X'_{i,t}\beta^* + \gamma_0(100 - \kappa_{i,\bullet}) + \gamma_1 n_{i,\bullet}) \\
&= g(X'_{i,t}\beta^* + \gamma_0 \underbrace{(100 - \kappa_{i,\bullet} + \frac{\gamma_1}{\gamma_0} n_{i,\bullet})}_{\text{Claim Score } \ell_{i,t}}) \\
&= g(X'_{i,t}\beta^* + \gamma_0 \ell_{i,t}), \text{ with } \ell_{i,t} = (100 - \kappa_{i,\bullet} + \frac{\gamma_1}{\gamma_0} n_{i,\bullet})
\end{aligned}$$

A constant of 100 has been added at the second line, which changes the intercept β_0 (thus explaining β^* instead of β) but not the overall model. With this parametrization, the new covariate $\ell_{i,t}$, which summarizes all past claims, can be seen as a **claim score**. In other words, based on their past claims experience:

- Insureds with a high value of $\ell_{i,t}$ indicates an insured with bad past claim experience;
- Insureds with a low value of $\ell_{i,t}$ indicates an insured with good past claim experience.

Using this simple regression model, called a **Kappa-N model**, we can obtain good and usable properties for the implied ratemaking structure:

- For an insured i without insurance experience, we would have $n_{i,\bullet} = 0$, and $\kappa_{i,\bullet} = 0$, which implies an initial claim score of 100. In other words, we can suppose that a new insured entering the portfolio should have a claim-score of 100.
- Each annual contract without a claim will decrease the claim score by 1;
- Each claim increases the claim score by $\Psi = \frac{\gamma_1}{\gamma_0}$, called the **jump-parameter**. For convenience, without losing too much precision, we can even round Ψ to obtain an integer value. This could help to better explain the ratemaking structure to insureds, brokers and administrators, among others.
- The impact of a single claim on the premium is then roughly equal to Ψ years without claims. In other words, if an insured claims, it would take Ψ years without claims to return to the premium the insured had prior to the claim;
- The penalty for a claim is an increase of $(\exp(\Psi\gamma_0) - 1)\%$ of the premium.
- Each year without a claim decreases the premium by $(1 - \exp(-\gamma_0))\%$.

These basic results are found and computed easily. This way of computing surcharges and discounts would clearly be useful to anybody involved in the insurance industry. It is quite simple to explain to insureds the penalties of claiming, and how long the penalties for that claim will last.

Insured	Years											
i	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	$-\kappa_{i,\bullet}$	$n_{i,\bullet}$
1	0	0	0	0	0	0	0	0	0	0	-10	0
2	2	0	1	0	0	0	2	0	1	0	-6	6
3	4	1	2	0	0	0	0	0	0	0	-7	7

Table 1: *Insureds with claims experience*

2.3.1 Limits for the Claim Score

One obvious problem with the Kappa-N model is the possible values of the claim score $\ell_{i,t}$ for the whole portfolio. Indeed, the Kappa-N model has no minimum or maximum values. With the database used in Section 4, for example, we can see that the some insureds may have a past claims experience of up to 10,15 or 20 claims. Even if there are many years without claims, premiums for these insureds would include an extreme surcharge of $\exp(20\Psi\gamma_0) - 1$ times the basic premium. Similarly, as there is no discount limit for insureds who did not claim in the last 10, 15 or 30 years, the Kappa-N model can generate large discounts.

A solution could be to limit the value of the actual claim score $\ell_{i,t}$ in the modeling. For example, we can limit the claim score to be between ℓ_{min} and ℓ_{max}, such as having

$$\mu_{i,t} \quad = \quad g\left(X_{i,t}'\beta^* + \gamma_0\ell_{i,t}^*\right), \text{ with } \ell_{min} \leq \ell_{i,t}^* \leq \ell_{max}.$$

The problem with this solution is that it only limits the actual claim score. Suppose, for example, that we have insureds from Table 1, where the insurance experience is shown for three insureds. Suppose a decrease of one for each year without a claim and a jump-parameter Ψ of 4, meaning that each claim penalizes the insured with an increase of 4 to the claim score. Iw we start at level 100 in 2011, that means insureds would respectively be at levels 90, 118 and 121 in 2021. If we suppose limits to the claim score, for example $\ell_{min} = 95$ and $\ell_{max} = 115$, insureds would be rated in 2021 with corrected levels ℓ^* of 95, 115 and 115. Even if this approach limits the spectrum of possible premiums, it has many undesirable consequences:

1. Insured #1 would not receive any premium surcharge if he claims one time in 2021. Indeed, by being at level $\ell_{1,2021} = 90$, a claim in 2021 would mean he would have a claim score of $\ell_{1,2022} = 90 + 4 = 94$, meaning $\ell_{1,2022}^* = \max(94, \ell_{min} = 95) = 95$. Insured #1 was already rated with $\ell_{min} = 95$ in 2020, meaning he will stay at the same level even if he claims once.
2. Insureds #2 and #3 would not receive any premium reduction if they do not claim in 2021. They are respectively at levels $\ell_{2,2021} = 118$ and $\ell_{3,2021} = 121$, resulting in a corrected claim score ℓ^* of 115 for both of them. By having a claim-free year in 2021, they will be at levels $\ell_{2,2022} = 117$ and $\ell_{3,202e} = 120$, meaning that they will both in fact be rated at the corrected levels $\ell_{max} = 115$ in 2022.
3. Similarly, insureds #2 and #3 won't have any premium surcharges if they claim in the next four or seven years, respectively. It would also take four and seven consecutive years without claims for these two insureds to obtain a premium reduction.

This kind of past claims rating structure is not desirable for an insurer because it greatly reduces incentives for insureds to not claim.

2.3.2 Discussion

In recent actuarial literature, many models using machine learning, statistical learning or neural network techniques have been used for ratemaking (see Denuit et al. (2020a) or Denuit et al. (2020b) for examples). In most cases, these approaches aim to find the best covariates, or functions of covariates, to model random variables, such as the number of

claims. With the possibility of using $n_{i,T-1}, n_{i,T-2}, \ldots, n_{i,1}$, as well as $\kappa_{i,T-1}, \kappa_{i,T-2}, \ldots, \kappa_{i,1}$, one might argue that this ratemaking problem comes down to simply finding the best relations between those covariates and the random variable we want to predict. It is not that simple because as we just saw, we must consistently limit the spectrum of possible premiums to avoid extremes, but we also want a consistent rating structure that, for example, aims to give an incentive for insureds to not claim.

Indeed, pricing is not just a one-time, one-year prediction. It should be seen as something that will be repeated each year. Because the same insured will be priced each year, it is necessary to ensure temporal consistency in computing annual premiums. More precisely:

- Premium increases and decreases based on loss experience must be logical: it is not expected that the premium will decrease following a claim, or that the premium will increase following a year without a claim. Indeed, an approach that tries to explain and model a phenomenon must not only be predictive but must also offer an explanation for this phenomenon. Empirical studies and analyses of policyholders' behavior indicate an increase in future claims frequency following a claim (see Abbring et al. (2003) for an exhaustive analysis of occurrence dependence and moral hazard in insurance), so the pricing model should reflect that behavior.
- Additionally, insurers want to promote insureds' good behavior by rewarding them as much as possible for each year without a claim, or conversely by penalizing them for each claim;
- If possible, the experience-based pricing structure should be easily understood, so that the system can be explained to the legislative authorities that regulate pricing, as well as to the various administrators of insurance companies and policyholders.

2.4 Bonus-Malus System

To satisfy these pricing objectives, a better way to deal with extreme situations that result from a Kappa-N model would be to again limit the value of all claim scores, but to limit them for all past insurance contracts. We can use the insureds in Table 1 again as an example.

To compute premiums in 2021, instead of applying maximum and minimum values to the 2021 claim scores, we apply the same limits but to all claim scores the insureds had in the past. Figure 1 shows how the limits applied on past contracts impact the actual claim score for all three insureds. For example, we can see that the claim score of insured #3 is limited as early as 2012. With this approach, we see that many of the problems mentioned earlier are solved. Indeed:

1. Even if insured #1 has the minimum claim score, i.e. $\ell_{1,2021} = 95$, he will receive a surcharge if he claims in 2021. However, as in the Kappa-N approach, he will not receive any other discounts for a claim-free year, but because there are only two possibilities: he either claims or does not claim, so there is a clear incentive to not claim in 2021.
2. Insureds #2 and #3 would receive premium reductions if they do not claim in 2021.
3. Insured #3 greatly improved his claiming behavior in the last 6-7 years, and has been rewarded with premium reductions in recent years.

This way of dealing with claim scores appears to more closely resemble with how insurers want to deal with their insureds: discouraging claims by giving surcharges, rewarding insureds who do not claim, forgiving old claims, etc. By adding maximum and minimum claim-score values for past contracts, the Kappa-N approach developed so far has been transformed into what is usually called a **Bonus-Malus Scale** system, or simply BMS, where the claim-score can be seen as a **BMS level**, or simply a level. The BMS can now be seen as a class system with a finite number of levels (when the jump parameter Ψ is an integer), where a relativity is assigned to each level. For this first BMS model presented so far, the transition rule is simple: an insured moves down by one level if he does not claim during his contract, and moves up by Ψ levels for each claim.

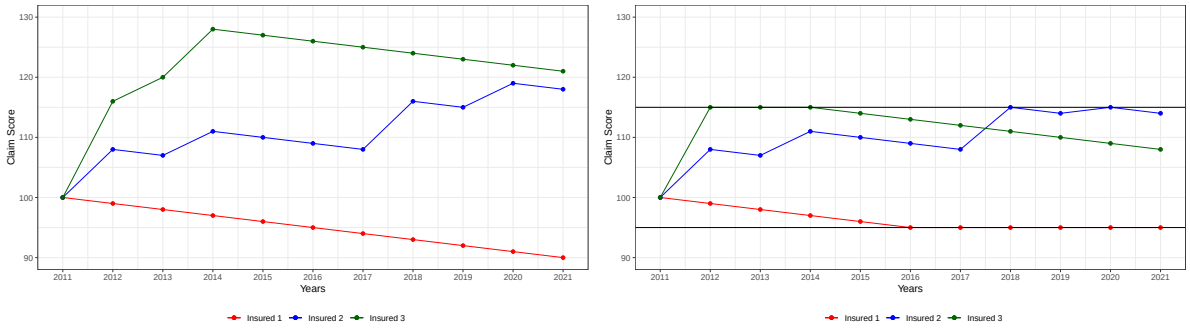

Figure 1: *Insureds with claim experience, with and without limits*

That means that, at minimum, BMS models can be defined with the following structural parameters:

1. The entry level, ℓ_0;
2. The jump parameter, Ψ;
3. The minimum level of the system, ℓ_{min};
4. The maximum level of the system, ℓ_{max}.

The BMS model presented in the previous example, in Figure 1, could then be defined as $(\ell_0; \Psi; \ell_{min}; \ell_{max}) = (100; 4, 95; 115)$. This way of presenting BMS models is slightly different than usual. Indeed, in the BMS literature, the lower level of the BMS is usually fixed at 0 (or 1), while the entry level must be estimated with data. By fixing the entry level at 100, however, we think that the BMS approach becomes more intuitive because the link between BMS and classic GLM with covariates of equation (2) is clearer.

2.4.1 Visualization

For a specific insured, as defined by Boucher and Pigeon (2019), a more formal way to define BMS is to define the claim score, or the BMS level, of contract t as:

$$\ell_{i,t+1} = \ell_{i,t} - \kappa_{i,t} + \Psi \times n_{i,t}, \quad \text{with } \ell_{min} \leq \ell_{i,t+1} \leq \ell_{max}.$$

A way to better visualize the BMS models is shown in Figure 2. This figure shows how the past claims relativities depend on level ℓ, and allows us to understand the impact of the jump parameter $\Psi = 4$ on the premium calculation. The blue curve, related to the value of $\exp(\gamma_0 \times \ell)$, mainly indicates the discount for a year without a claim, while the combination of the blue curve and the jump parameter indicates the surcharge for a claim. Starting at level 100, we can see how an insured would move across all the BMS levels and what relativities he will have depending on whether he claims ($+\Psi$ in red) or not (-1 in green).

The same figure also shows the spectrum of all relativities for a BMS model, as limits $\ell_{min} = 66$ and $\ell_{max} = 115$ are shown. With those limits, the maximum relativity is then limited at around 1.65, while the lower relativity is approximately 0.35. A figure illustrating the Kappa-N model, which corresponds to a BMS model without any limit, would be similar, with the difference that no maximum or minimum values of the relativities would be shown.

3 Numerical Application

3.1 Practical Considerations

The Kappa-N and BMS models presented in the last section are now completely defined. However, that does not mean that they can easily be applied with real insurance data. Indeed, data have limitations and must be transformed to be

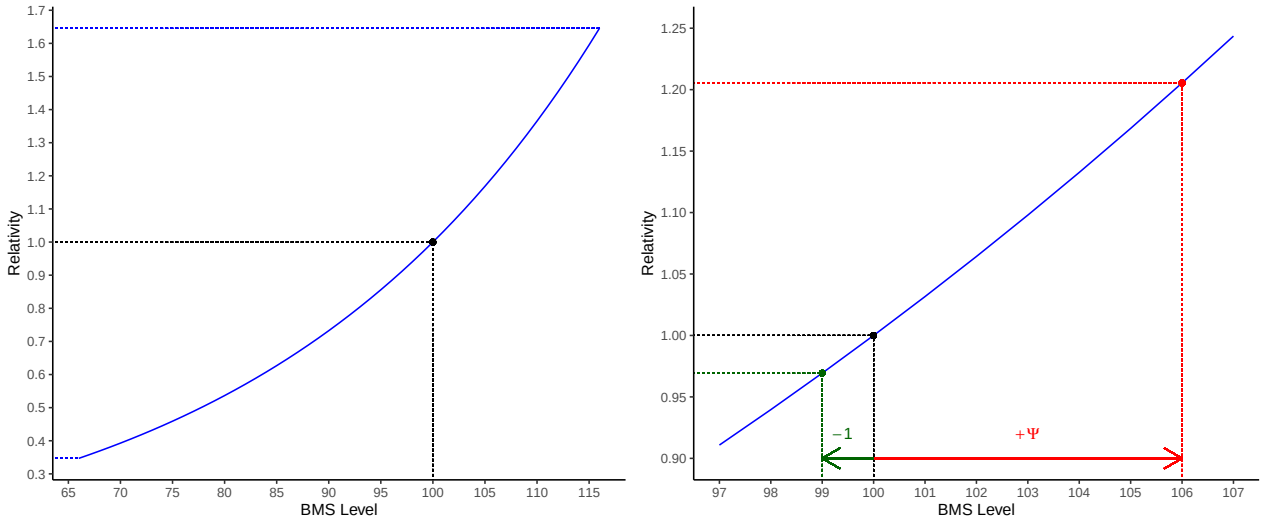

Figure 2: *Example of BMS Relativities (left: all levels, right: zoom on levels around 100)*

suitable for any rating models. Consequently, in this section of the paper, before explaining how to estimate the models with insurance data, we first explain several practical elements.

3.1.1 Data Transformation

To better understand the practical elements to consider before using experience rating models, we use a specific insurance product as an example. We use farm insurance data from a major insurance company in Canada. We were able to use contracts from 2014 to 2019. The general form of the data used in experience rating looks like the sample shown in Table 2. Each line of the database corresponds to the specific coverage of an annual contract. On each line, we find information about the insured, the contract, the items covered, but we also see the date of the first insurance contract with the insurer. This date is very important in constructing Kappa-N and BMS models. Information about claims over that period of time is also available. A meta-variable ξ has also been computed to summarize the information at the item-level. Details about how this specific database was constructed can be found in the Appendix.

Policy Number	Number of Items	Effective Date	First Insurance	Coverage	...	Province	ξ	Number of Claims	Costs Costs
1	2	2017-01-15	1995-01-15	MACHINERY	...	Ontario	0.015	2	186,592
1	2	2018-01-15	1995-01-15	MACHINERY	...	Ontario	0.003	0	0
1	2	2019-01-15	1995-01-15	MACHINERY	...	Ontario	0.041	1	87,112
2	5	2017-03-22	2003-03-22	MACHINERY	...	Quebec	0.045	0	0
2	4	2018-03-22	2003-03-22	MACHINERY	...	Quebec	0.023	1	18,889
2	7	2019-03-22	2003-03-22	MACHINERY	...	Quebec	0.081	1	7,444

Table 2: *Fictive Data Sample - Contract Level*

By supposing independence between each line of the database, standard GLM approaches can be used to model the number of claims, the severity of the claims or the loss costs to estimate the impact of specific covariates in the model.

In our case, to illustrate Kappa-N and BMS models, we decided to model the number of claims. This is done to avoid dealing with open claims where the final cost of the claims is unknown. Indeed, with the farm data, approximately 25% of the claims that occurred in last three years are still open. Severity modeling or loss costs modeling would then imply,

for example, the use of an algorithm that would project the final cost of each claim. This is an interesting solution, but this was not our focus for this project.

3.1.2 Availability of Past Information

Even though we model the number of claims of a specific coverage, we also decide to use all types of claims to define the claim score for the Kappa-N and the BMS model. This allows us to be more precise and to define two other aspects of any past claims rating:

1. The variable to model and to predict, named the **target** variable;
2. The information used to define what we consider as past claim experience, named the **scope** variable.

In our case, the target is the number of claims for the machinery coverage of the farm insurance product, while the scope is the claims from any coverages of the farm insurance product. Based on the reason why past claims rating exists, which was mentioned in the introduction, we think that a scope variable that includes all types of coverage, or insurance products, seems more coherent. Indeed, claim-prone individuals will claim on all coverages of their insurance contract. Similarly, the reason why an insured is no longer afraid of the claims process may be due to a claim made on another coverage. That means that, ideally, the scope variable should not only be all coverages of the farm insurance product but probably all insurance coverages: farm, automobile, home, etc. Multi-product pricing models are thus strongly justified (see Pechon et al. (2019a) for an application of multivariate credibility models between insurance products).

If all past contract information and past claims were available for all insureds from the portfolio, the Kappa-N and BMS models could easily be directly applied. However, and this has been mentioned many times in the literature, one major problem with past claims rating refers to the availability of past information. Often, insurers cannot access all past information about their insureds. In many cases, not only are insurers not able to obtain past information from other insurers, but they are also often unable to use information from their own old contracts. Indeed, insurers often modify their operating systems, and past databases are simply erased or are no longer useful.

Classic past claim rating models often normalized the past experience of each insured i by using $\frac{\sum_t n_{i,t}}{\sum_t \lambda_{i,t}}$ in the model instead of simply using $n_{i,\bullet}$. The Buhlmann-Straub credibility model is a well-known example of this kind of normalization. On the other hand, standard Kappa-N and BMS models shown in the previous section do not need to link past claims with past covariates [1]. That means that for BMS and Kappa-N models, our main interest in the information from the past is solely past claims, and not in past covariates:

- In automobile insurance, some jurisdictions have a central agency that collects claims from all insurers. In this situation, it is then possible to use (some) past claims from other insurers. However, for other insurance products, we are often unable to access past claims information from other insurers.
- In such cases, one possibility would be to rely on each insured's declaration concerning past claims experience. However, a quick analysis of the data available showed us that this kind of data is highly biased and unreliable. Indeed, the insured has a real incentive to lie to obtain a lower premium.

For our dataset, Figure 3 shows the distribution of the numbers of years of experience with the insurer. As opposed to automobile insurance in Canada, where insureds will frequently move from one insurer to another, we see that farm insurance has more stable insureds. Indeed, in our case, the average number of years with the insurer is 18.4, and the maximum observed years is 59 [2]. The maximum available number of past claims experience for all insureds is 15 years, and only insurance experience from within the insurer is available. That means that we considered that the first year of insurance for any insured corresponds to the insured's first year with the insurer. In other words, if a farm is first seen in the database in 2003, we will consider this farm to be a new insured without any prior experience in 2003.

[1] This generalization of the BMS is one of our teams' current areas of research

[2] Farms are sometimes passed from generation to generation. Insurance experience would not be reset in these cases.

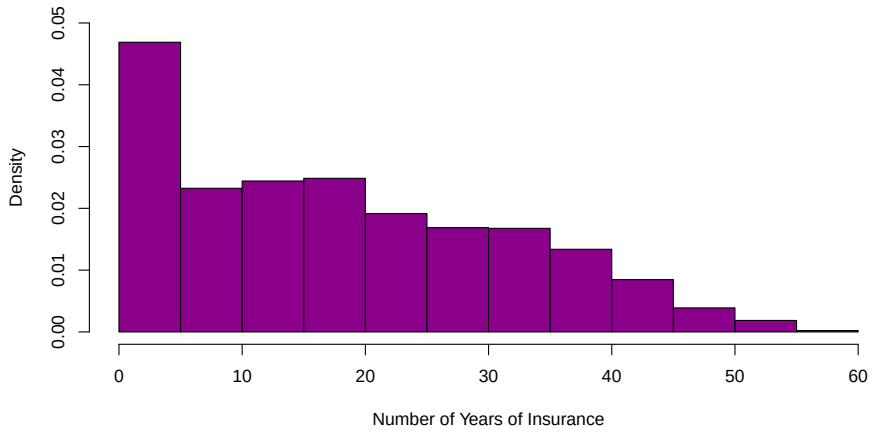

Figure 3: *Distribution of the number of years of experience with the insurer*

3.1.3 Evolution of Past Claims

One of the characteristics of a bonus-malus scale system is its Markovian property, which can be explained by saying that by knowing the actual BMS level at time t, ℓ_t, and the actual number of claims n_t, we are able to obtain the future BMS level at time $t+1$, ℓ_{t+1}. More formally, it means that:

$$\Pr(N_{i,t+1}|n_{i,1},\ldots,n_{i,t},\ell_{i,t}) = \Pr(N_{i,t+1}|n_{i,t},\ell_{i,t})$$

This is a well-known property of BMS (see Lemaire (2012) or Denuit et al. (2007)). BMS is even used occasionally as a simple example of a Markov process in introduction courses to stochastic processes. However, in reality, this property does not completely hold. Indeed:

- The premium for a renewal is calculated x weeks before the renewal, and is sent to the insured. That means that a claim that is made after the premium calculation cannot be considered in the ratemaking. In other words, for the premium calculation of insured i for contract T, we should use $n^*_{1,(1:T-1)}$, a subset of $n_{1,(1:T-1)}$, which excludes claims thatt are made between the premium renewal calculation and the end of the contract. However, to keep the Markovian property, a simple time transformation could probably be applied to solve the issue.
- Some past claims are forgiven or closed without any payments. Those claims are usually not considered by insurers in any past claims rating scheme (even if they probably should be). However, insurers might consider that any open claims with a positive case reserve should be penalized in the premium computation. This is logical because claims can take many years to close and it does not make any sense to not consider them in premiums calculation if the insurer thinks that they will ultimately generate positive payments.

  However, it means that it is also possible for a claim with a positive case reserve to finally close without payment. In such a case, we could therefore have a situation where a claim would first be used to penalize an insured in the calculation of the premium. Several years later, if this past claim finally closes at zero, the claim would no longer be considered in the premium calculation. For that reason, the Markovian property of BMS no longer holds. That means that the BMS rating algorithm of insured i for contract T should not be based only on $\ell_{i,T-1}$ and $n_{i,T-1}$, but should be programmed to always look at all past years of insurance in the computation of the premium.

Interestingly, in that specific fictive case, after the claim closes at zero, the insured will receive an important discount. Indeed, the claim score would decrease by Ψ (if the minimum value ℓ_{min} is not reached) because one claim will no longer be considered in the computation of the premium. It is up to the insurer to decide if past premiums ("incorrectly" computed with a claim) should be partially reimbursed or not. Analyzing this type of situation could be interesting.

3.2 Count Distributions and Estimation

To model the number of claims (our target variable), the first approach to try is to look at basic count distributions used without any covariates linked to experience rating. In this case, we have independence between all policyholders as well as between all contracts, meaning that we can write

$$\Pr(N_{i,t+1} = n|\boldsymbol{N}_{i,t}, \boldsymbol{X}_{i,1}, \ldots, \boldsymbol{X}_{i,t+1}) = \Pr(N_{i,t+1} = n|\boldsymbol{X}_{i,t+1})$$

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

$$\mathrm{E}[N_{i,t+1}|\boldsymbol{N}_{i,t}, \boldsymbol{X}_{i,1}, \ldots, \boldsymbol{X}_{i,t+1}] = \mathrm{E}[N_{i,t+1}|\boldsymbol{X}_{i,t+1}] = \lambda_{i,t+1} = \exp(\boldsymbol{X}'_{i,t}\beta).$$

The base model is usually the Poisson distribution, with a probability mass function defined as:

$$\Pr(N_{i,t} = n|\boldsymbol{X}_{i,t}) = (\lambda_{i,t})^n \exp\left(-\lambda_{i,t}\right)/n!,$$

and 0 elsewhere. To account for overdispersion, a standard alternative is the Negative Binomial distribution of type 2 (NB2) with parameters τ and $\lambda_{i,t}$ and probability mass function given by

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

and 0 elsewhere. A slightly different version of the Negative Binomial distribution (NB1) with parameters $\lambda_{i,t}$ and τ for which the probability mass function is

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

and 0 elsewhere, is often used.

All three count distributions are then be used with the following experience rating models:

1. **Standard Model**: This model refers to basic count distributions that don't use any covariates linked to experience rating;
2. **Kappa-N Model**: This approach corresponds to a generalization of the count distribution, where a claim score linked to the past claims experience has been added. In other words, two other covariates $-\kappa_{i,\bullet}$ and $n_{i,\bullet}$ are added to the base model.
3. **BMS Model**: This model is similar to the Kappa-N model, but minimum and maximum limits to the claims score (ℓ_{min} and ℓ_{max}) were added.

As this paper's proposed approach only deal with the form of the mean, straightforward generalizations can be made to use other count distributions than Poisson, NB2 or NB1. For example, various Poisson-inverse Gaussian, hurdle or zero-inflated distributions can be used. See Cameron and Trivedi (2013) or Winkelmann (2008) for a nice overview of count distributions, or Boucher et al. (2008) for a survey of claim counts for automobile insurance.

3.2.1 Estimation Algorithm

For the Standard and Kappa Models, simple GLM estimation techniques could be used for the Poisson distribution, such as the ones already programmed in R or SAS for example. For the NB2 or NB1 distributions, direct optimization of the loglikelihood can be performed. However, for the BMS model, finding the best values for structural parameters Ψ, ℓ_{min} and ℓ_{max} is not simple because they cannot be estimated directly. For the Kappa-N model, which does not have any limits, the jump parameter Ψ can be estimated directly, as we saw previously. However, limiting the claim score values by ℓ_{min} and ℓ_{max} for all contracts of each insured in the database means recomputing the claim score path of each insured from their first contract.

When Ψ, ℓ_{min} and ℓ_{max} are known, the model can be estimated easily. To find the structural parameters, as done by Boucher and Pigeon (2019), one obvious solution is to test all possibilities of the structural parameters and choose the best combinations[3]. This method obviously works, but it is time consuming. We used a simpler and faster iterative technique that works as:

1. Estimate a standard model with $\kappa_\bullet$ and $n_\bullet$. We already saw that it can be seen as a BMS without any limits.
   - We then have a first estimate of our jump parameter Ψ;
2. For this value of Ψ and $\ell_{min} = 0$, find the best value of ℓ_{max} by looking at all values between 100 and a reasonnable maximum value of ℓ_{max};
3. With this new value of ℓ_{max} and ℓ_{min}, find the best value of Ψ by looking at all values between 1 and a reasonnable maximum value of Ψ;
4. With this new value of ℓ_{max} and Ψ, find the best value of ℓ_{min} by looking at all values between 0 and 100;
5. Repeat lines 2, 3 and 4 until convergence.

Several models with structural parameters near the values found with this algorithm are finally checked to be sure that a local maximum has not been found. This simple algorithm allows us to estimate the BMS model in minutes using a simple laptop, when using the dataset used in this paper.

3.2.2 Results and Estimated Parameters

All models have been fit with the data, and the results are shown in Table 3. On the training dataset, the NB1 distribution always exhibits a better fit than the other count distributions for all three types of model. Indeed, the loglikelihood obtained for the NB1 distribution is better than the other ones. Based on the loglikelihood, comparison of all types of models shows that the Kappa-N model, using two more parameters, is always significantly better than the standard model. The loglikelihoods for the BMS models are also better than the Kappa-N models. That means that the BMS model NB1 with a BMS rating structure is the best model among those used. For the test dataset, the logarithmic score defined as $\sum_{i=1}^{n} -\log(\Pr(n_i; \widehat{\lambda}_i))$ has been used (see Roel et al. (2017) for details or description of other scores) to define the prediction quality. Results are similar regarding the ranking of the types of model, but for the underlying distribution, the NB2 distribution seems to always outperform the NB1.

Estimated parameters related to experience rating are shown in Table 4. With a lower jump parameter Ψ and a lower parameter γ_0, we see that the Kappa-N model penalizes less a claim than the BMS model, but also gives a smaller discount for a contract without claim. We also see that the three underlying distributions have approximately the same estimated parameters, where the only difference is that the NB2 distribution has $\ell_{max} = 115$ instead of 116 like the others. For privacy reasons, but also because this is not the focus on the paper, β parameters associated with covariates or with the meta-variable ξ are not shown in the table. However, as explained and shown clearly in Boucher and Inoussa (2014), compared with *a priori* rating, many β parameters might change significantly when past claims experience is included in the modeling.

[3]We restrict the BMS model to have integer values for structural parameters Ψ, ℓ_{min} and ℓ_{max}, because it is clearly easier to use and easier to explain to everyone involved in the insurance industry. The difference between a model with integer structural parameters and a BMS model with continuous parameters is negligible.

	Standard Model		Kappa-N Model		BMS Model	
Distributions	Train	Test	Train	Test	Train	Test
Poisson	-8,562.359	2,872.129	-8,506.149	2,858.352	-8,490.026	2,857.029
NB2	-8,555.539	2,869.348	-8,499.319	2,856.058	-8,485.290	**2,854.189**
NB1	-8,552.380	2,871.878	-8,497.791	2,857.056	**-8,482.994**	2,855.270

Table 3: *Results for all models*

	Kappa-N Model		BMS Model			
Distributions	$\hat{\Psi}$	$\hat{\gamma}_0$	ℓ_{max}	ℓ_{min}	$\hat{\Psi}$	$\hat{\gamma}_0$
Poisson	3.72	0.0254	116	85	6	0.0312
NB2	3.85	0.0259	115	85	6	0.0298
NB1	3.71	0.0254	116	85	6	0.0287

Table 4: *Some parameters for all Kappa-N and BMS models*

To add another comparison, we also show the result of a Poisson GAM model, where spline functions are used on $\kappa_\bullet$ and $n_\bullet$. Figure 4 shows the resulting smoothing functions. We note an almost linear function of $\kappa_\bullet$, meaning that the claims frequency always decreases when the number of years without claims increases. In relation to objectives of Section 2.3.1, the linear form of the smoothing function of $\kappa_\bullet$ is not a problem and simply indicates that the insurer should reward each year without a claim by a fixed percentage discount. This is logical and can easily be used in practice.

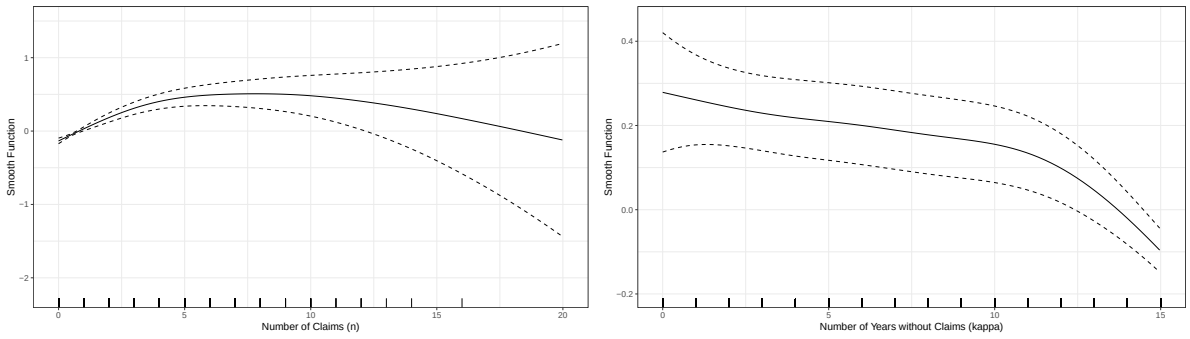

Figure 4: *Splines form the Poisson GAM model*

Conversely, the smoothing function of $n_{\ bullet}$ is more problematic. Indeed, if an insurer wants to directly use this result, he would have to use a ratemaking structure that would decrease the penalty of each additional claim. At some point, the model would even give discount for each new claim. Not only is it not logical, but it seems to be a bad way of explaining how and why insureds claim.

3.2.3 Discounts and Surcharges

Compared with Kappa-N models, a way to explain the success of the BMS model is related to how the model deals with old claims. Indeed, adding a maximum BMS level ℓ_{max} to a Kappa-N model means that the BMS model can allows insureds to improve over time, as older claims can be forgiven, and not always be considered in all future premiums.

This, in some way, solves one problem already mentioned by Young and De Vylder (2000), who used credibility theory to correct past claims rating models for "unlucky" insureds. The authors find that credibility models penalize insureds with claims too severely. In our case, with a BMS with $\ell_{max} = 116$ for the NB1 distribution, bad luck seems to be limited. With a jump parameter Ψ of 6, and $\ell_{max} = 116$, more than two claims in the first years of insurance will not

be severely penalized. More precisely, to better understand the results obtained for the BMS model, we can compute the discounts and surcharges of the model, based on the number of past claims. More concretely, we then have:

- The jump parameter Ψ is equal to 6, meaning that each claim increases the BMS level by 6. After a claim, an insured would need six years without a claim to return to the original premium.
- The value of γ_0 is now 0.0287. That means that the penalty for a claim is equal to $\exp(0.0287 \times 6) - 1 = 18.8\%$, and each year without a claim decreases the premium by $1 - \exp(-0.0287) = 2.83\%$.
- The maximum BMS level is $\ell_{max} = 116$, meaning that the maximum surcharge, compared to level 100, is $\exp(0.0287 \times 16) - 1 = 58.2\%$;
- The maximum BMS level is $\ell_{min} = 85$, meaning that the maximum surcharge, compared to level 100, is $1 - \exp(-0.0287 \times 15) = 35.0\%$.

As we can see, those basic results are found and computed easily. This way of computing the surcharges and discounts would clearly be useful to any insureds, brokers or administrators. It is quite simple to explain to insureds how large their penalties for a claim will be, and how long hte penalty for that claim will last. Another interesting conclusion about the BMS model is that all insureds will have a premium located between 0.650 and 1.582 times the basic premium for a new insured, at level 100. This narrowly limits the range of premiums.

3.2.4 Distribution over BMS levels

Figure 5 illustrates the predicted and the observed claims frequency for all levels of the BMS model on the training and the test datasets. The BMS model seems to fit the data quite well, and we can observe that classifying insureds by their claim score works well because the insureds with higher levels have worse claims experience than insureds with lower levels. On the test dataset, a similar conclusion can be made, even if we observe more variations. Insureds at level 100 or higher, who have reported claims at least once before, exhibit significantly higher claims frequency.

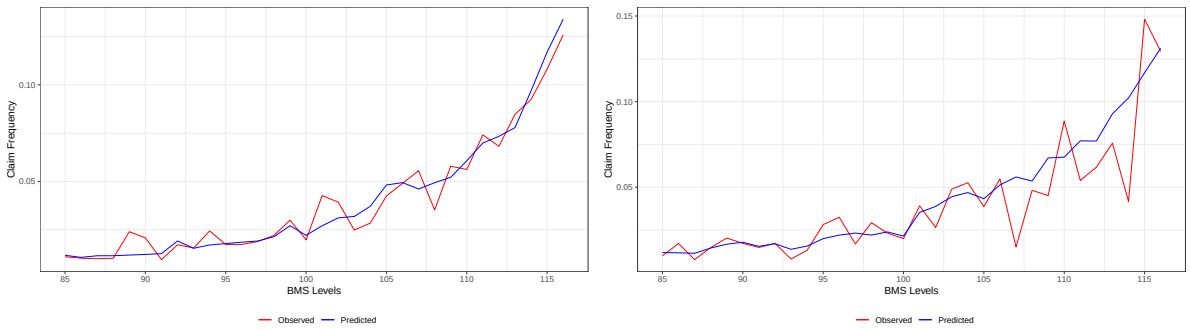

Figure 5: *Predicted versus observed results for each level of the BMS (training and test datasets)*

Distribution of insureds over all BMS levels for the test dataset is shown in Figure 6 (a similar graph is obtained for the training dataset). We see that most of the insureds are located at levels 100 or lower. A peak at level 85, corresponding to ℓ_{min} is also observed. Because the maximum number of past claims experience is 15 years, that means that a significant proportion of insureds in the portfolio did not report at all.

4 Past Experience

In our numerical applications, we were lucky because we were able to use many years of past claims experience in our experience rating models. However, in many situations, insurers are not able to get that much information. Section 3.1.2 detailed some situation regarding the availability of past information for many insurance products.

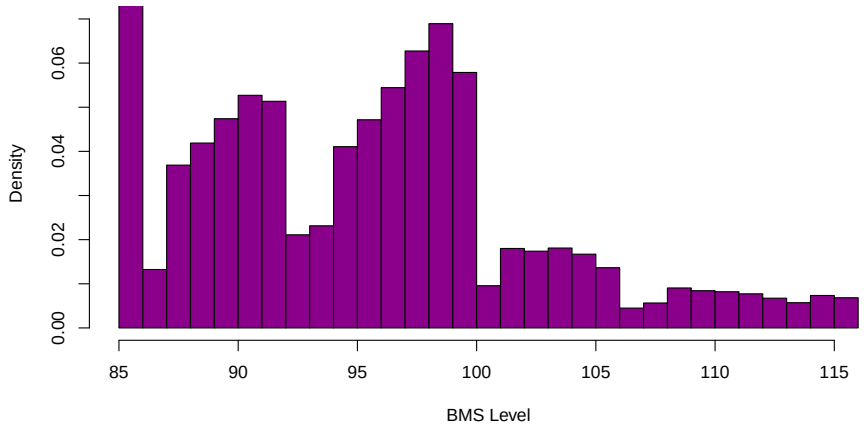

Figure 6: *Distribution over the levels*

Figure 7 illustrates the situation, where the timeline of insurance experience of a specific insured i is divided in two sections:

1. The Past Claims Information section, from $\tau_i^{(1)}$ to $\tau_i^{(2)}$. This corresponds to the time period where past claims information were available to compute $n_{i,\bullet}$ and $\kappa_{i,\bullet}$, or the Bonus-Malus level.
2. The Artificial Information section, from $\tau_i^{(0)}$ (the date of the first insurance contract of insured i) to $\tau_i^{(1)}$. When $\tau_i^{(2)} - \tau_i^{(1)}$ is small, experience rating models might be difficult to estimate because the amount of information needed to compute the bonus-malus level, for example, is too small. Moreover, if the Past Claims Information is small, it also means that the insurer will not be able to compute an adequate BMS level for new insureds, meaning that the rates between the insureds who only have a few years of experience and those who have been insured for a much longer time will be similar.

In other words, it could happened that many contracts from the rating database are censured when we want to compute $n_{i,\bullet}$ and $\kappa_{i,\bullet}$, or the Bonus-Malus level. A solution to this problem is to create artificial past claims experience. Artificial claims history allows us to use more years of past claims in the modeling. As we do not have limits in creating artificial past claims, we can take this idea even further by creating a full claims history for all insureds in the database, starting from $\tau_i^{(0)}$ to $\tau_i^{(1)}$, as long as we have access to the date any insured began to be insured.

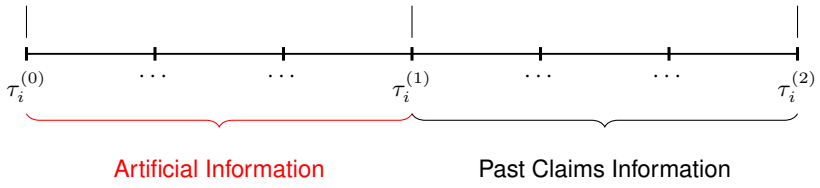

Figure 7: *Time Frame of Past Claims Information for insured i*

4.1 Methods for Creating Artificial Past Claims Experience

In the actuarial literature, two methods have been proposed to generate an artificial past claims history:

1. Boucher and Pigeon (2019) suppose that each unobserved year of experience would be considered a year without claims because the authors showed that this is the most probable outcome. Indeed, with a small claims

frequency, the vast majority of insureds will not claim in a single year. Under this construction, we would have an updated version of $n^*_{i,\bullet} = n_{i,\bullet}$ and $\kappa^*_{i,\bullet} = \kappa_{i,\bullet} + (\tau_i^{(1)} - \tau_i^{(0)})$, where $(\tau_i^{(1)} - \tau_i^{(0)})$ corresponds to the number of artificial years that have to be created. For the BMS model, that would mean that each artificial year would correspond to a decrease of one level. Even if this method of creating artificial past claims history is quick and efficient, it will, on average, underestimate the overall claim frequency because it is almost certain that some of those insureds claim in the past.

2. Another method is to suppose that all unobserved years of experience involve an average expected number of claims $\tilde{\mu}$, as Boucher and Inoussa (2014) have proposed. That would mean that a corrected version of $n_{i,\bullet}$ and $\kappa_{i,\bullet}$ could be computed as:

$$\begin{aligned} n^*_{i,\bullet} &= n_{i,\bullet} + \sum_{t=\tau_i^{(0)}}^{\tau_i^{(1)}-1} \tilde{\mu} = n_{i,\bullet} + (\tau_i^{(1)} - \tau_i^{(0)})\tilde{\mu} \\ \kappa^*_{i,\bullet} &= \kappa_{i,\bullet} + \sum_{t=\tau_i^{(0)}}^{\tau_i^{(1)}-1} \Pr(N=0;\tilde{\mu}) = \kappa_{i,\bullet} + (\tau_i^{(1)} - \tau_i^{(0)}) \Pr(N=0;\tilde{\mu}) \end{aligned}$$

where $\tau_1 - \tau_0$ represent the time period where an artificial past claim should be created. By supposing a Poisson distribution, $\Pr(N=0;\tilde{\mu}) = \exp(-\tilde{\mu})$, that could simply be seen as the probability of not claiming for a single year. Regarding the BMS model, using the corrected version of $n_{i,\bullet}$ and $\kappa_{i,\bullet}$ would also mean that, for $t \in (\tau_i^{(0)}, \tau_i^{(1)} - 1)$, the BMS levels of each artificial year correspond to:

$$\ell_{i,t} = \ell_{i,t-1} - \Pr(N=0;\tilde{\mu}) + \Psi\tilde{\mu}$$

We favor the second approach because it considers the possibility that past years of insurance include years with claims. There are many ways to estimate $\tilde{\mu}_i$ for insured i. Boucher and Inoussa (2014) estimate $\tilde{\mu}_i$ based on the first available covariates of insured i, i.e. at time $t^* = \tau_i^{(1)}$. That would mean using $\tilde{\mu}_i = \exp(X'_{i,t^*}\beta)$, with covariates X_{i,t^*}, corresponding to the time the first contract of insurance of insured i was observed on the rating database.

We propose to improve this approximation by using a panel data model to obtain a better estimate of each $\tilde{\mu}_i$, which will then help us to construct better artificial claims histories for all insureds in the portfolio. To improve our approximation of $\tilde{\mu}_i$ for all insureds $i = 1, \ldots, n$, we will model the joint distribution of available past claims history for each insured i, i.e. data from the Past Claims Information section of Figure 7. Formally, suppose that for each insured i, we have a vector containing all past numbers of claims $n_{i,t}$ for contracts beginning at time $t = \tau_i^{(1)}, \ldots, (\tau_i^{(2)} - 1)$. For an insured observed over all years for that period, we then want to compute the following joint distribution (subscript i is removed for simplicity):

$$\Pr[N_{\tau^{(1)}} = n_{\tau^{(1)}}, \ldots, N_{(\tau^{(2)}-1)} = n_{(\tau^{(2)}-1)}].$$

Many possible joint distributions can be used. In our case, we will suppose that it will be constructed by adding an unobserved individual random effect that affects all random variables $N_{i,t}$ of the same insured i. Conditionally on this random effect Θ_i, all those random variables $N_{i,t}$ of the same insured i are supposed to be independent. The joint probability mass function is then defined by:

$$\Pr[N_{\tau^{(1)}}, \ldots, N_{(\tau^{(2)}-1)}] = \int_{-\infty}^{\infty} \left(\prod_{k=\tau^{(1)}}^{\tau^{(2)}-1} \Pr[N_{i,k} = n_{i,k} | \theta_i] \right) dG(\theta_i).$$

where $G(\theta_i)$ is the cumulative distribution function of the random effect. Many possibilities exist for the conditional distribution and the random effect distribution. In our case, we will suppose:

$$(N_{i,k}|\Theta_i = \theta_i) \quad \sim \text{Poisson}\,(\theta_i \lambda_{i,k}) \text{ and } \Theta_i \sim \text{Gamma}(\alpha = \nu, \gamma = \nu).$$

This Poisson-gamma combination leads to what is called the multivariate binomial negative (MVNB) distribution, which can be expressed as:

$$\Pr[N_{\tau^{(1)}}, \ldots, N_{(\tau^{(2)}-1)}] = \left(\prod_{k=\tau^{(1)}}^{\tau^{(2)}-1} \frac{(\lambda_{i,k})^{n_{i,k}}}{n_{i,k}!} \right) \frac{\Gamma(n_{i,*} + \nu)}{\Gamma(\nu)} \left(\frac{\nu}{\lambda_{i,*} + \nu} \right)^{\nu} (\lambda_{i,*} + \nu)^{-n_{i,*}}, \tag{3}$$

with:

$$n_{i,*} \quad = \quad \sum_{k=\tau^{(1)}}^{\tau^{(2)}-1} n_{i,k} \quad \text{and} \quad \lambda_{i,*} = \sum_{k=\tau^{(1)}}^{\tau^{(2)}-1} \lambda_{i,k}$$

Usually, the panel count data distributions such as the MVNB are used to compute future premiums using the past number of claims. Based on the fact that collecting information on the $\boldsymbol{n}_{i,t},\ t = 1, \cdots, T-1$ improves the knowledge of the individual random effect Θ_i of insured i, we usually use the predictive mean to compute the expected number of claims for year T as:

$$E[N_{i,T}|n_{i,1}, \ldots, n_{i,T-1}] = \lambda_{i,T} \left(\frac{\nu + \sum_{k=1}^{T-1} n_{i,k}}{\nu + \sum_{k=1}^{T-1} \lambda_{i,k}} \right).$$

In our situation, we do not want to estimate the future number of claims of insured i, but want to use the same approach to create artificial past claims history. That means that in our situation, the MVNB distribution will not be used to predict the future number of claims, but will be used backward to approximate the number of past claims before $\tau^{(1)}$. Even used backward, the approach is still valid because we use all claims of insured i to improve our knowledge about the individual random effect Θ_i. The updated expected value of N would then be used as an approximation of past claims history. Formally, by reversing the time order, that means that we have:

$$\tilde{\mu_i} \quad = \quad E[N_{i,\tau^*}|n_{i,\tau^{(2)}-1}, \ldots, n_{i,\tau^{(1)}}] = \lambda_{i,\tau^{(1)}} \left(\frac{\nu + n_{i,*}}{\nu + \lambda_{i,*}} \right). \tag{4}$$

for any time $\tau^* < \tau^{(1)}$. Note that because past risk characteristics are not available before $\tau^{(2)}$, we use the oldest available risk characteristics of insured i, $X_{i,\tau^{(2)}}$ like Boucher and Inoussa (2014). We then have $\lambda_{i,k} = \lambda_{i,\tau^{(1)}}$ for all the Past Claims Information timeframe. This means that we can simplify $\lambda_{i,*} = (\tau^{(2)} - \tau^{(1)} + 1)\lambda_{i,\tau^{(1)}}$. This way of creating $\tilde{\mu_i}$ certainly improves the quality of the artificial past claims history because each insured i will have his own artificial history based on his observed claims experience.

Apart from the MVNB, other panel data count models can be used, such as those proposed in the recent literature on claim counts (see references mentioned in the introduction). An obvious alternative to the MVNB is the NB-beta, which has proven to be particularly suitable for insurance data (see Boucher et al. (2008) for example). Another interesting alternative would be to use advanced models that weight past claims by their age. Note, however, that if this kind of approach has to be used to create past claims history, we must be careful to use it backward. That would mean that more recent claims should have less impact than older claims in the panel data model because we want to create even older artificial claims.

4.1.1 A Warning from a Simple Example

To better understand how to construct an artificial claims history, a simple example is given. We suppose an insured observed for two contracts, in 2016 and in 2017. Those two contracts are used to estimate the parameters of the

experience-rating models, such as the Kappa-N or the BMS models. The values of $n_{i,\bullet}$ and $\kappa_{i,\bullet}$ come from the Past Claims Information period, which corresponds to years 2012-2015. We also suppose that the first insurance contract of that insured was in 2007, but we do not have information about his claim experience for his contracts of years 2007-2011. Figure 8 illustrates this example, and generalizes Figure 7 by adding a Rating Database section, that corresponds to the time period used to estimate the experience-rating models.

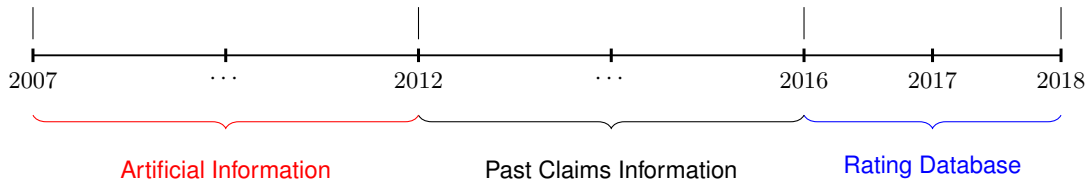

Figure 8: *Time Frame of Information for insured i*

The objective a creating an artificial past claims is to consider at least partially the time period 2007-2011 in the rating model, even if we do not know the real insurance experience for that time. As explained in the previous section, we use the Past Claims Information period, i.e. contracts 2012, 2013, 2014 and 2015 to compute $\tilde{\mu}$, from equation (). More precisely, we have:

$$\tilde{\mu} \quad = \quad \lambda_{2016}\left(\frac{\hat{\nu}+n_*}{\hat{\nu}+\lambda_*}\right) \quad \text{with} \quad n_* = n_{2012}+n_{2013}+n_{2014}+n_{2015} \ , \ \ \lambda_* = 4\times\lambda_{2016}.$$

The parameter $\hat{\nu}$ comes from the estimation of MVNB model (see equation [3]), that uses the Past Claims Information period for the whole porfolio. For the contract of 2017, the values of $n_{i,\bullet}$ and $\kappa_{i,\bullet}$ are again estimated from the Past Claims Information period, but can also be updated using the claims experience of the contract from the year 2016. Indeed, for example, we could now use n_{2016} to compute n_*.

Using Rating Database contracts to improve the artificial claims history can be interesting but would also create a lot of undesirable consequences. For example, this approach would mean that each additional contract experience will always modify the artificial claims history, which will in turn will modify $n_{i,\bullet}$, $\kappa_{i,\bullet}$ or the BMS levels that come from past contracts. In our example, if we are using a BMS model, a claim in 2016 can produce the following consequences:

- An increase[4] of the BMS level for the contract 2017, i.e. $\ell_{2017} = \ell_{2016} + \Psi$;
- An increase of $\tilde{\mu}$, which will potentially increase all past BMS levels, starting from year 2012 to year 2016.

On the other hand, if the insured does not claim in 2016, for the same reason, all BMS levels, starting from year 2012 to year 2016, will be modified retroactively because of the new computed value of $\tilde{\mu}$. Constant re-evaluation of the artificial claims history means that any surcharges and discounts following a contract will become extremely difficult to predict: in other words, we will lose one of the greatest advantages of BMS models. For this reason, creating an artificial claims history must only be made once for each insured i, based on the Past Claims Information time period, and never be modified by the claim experience observed in the Rating Database period.

4.2 Numerical Application

Using the same database as the one used in Section 3, we compared the three methods. We defined artificial history (AH) methods as follows:

- *no-AH*: no artifical claims history is created;

[4] If the BMS level is not limited by ℓ_{max}.

- *AH1* : each unobserved year of experience has a number of claims equal to the average expected number of claims computed with the available covariates (i.e. Poisson regression model);
- *AH2* : each unobserved year of experience has a number of claims equal to an updated average expected number of claims based on covariates and the past claims history (i.e. the MVNB regression model);
- *AH3* : each unobserved year of experience would be considered as a year without claims.

We also used different lengths of time from the Past Claims Information timeframe to compare AH methods. For all those time lengths, it means that the artificial claims history methods can be evaluated in at least two ways:

1. AH methods can be evaluated by comparing the artificial history with the observed claims history. For example, if we use Y years of past claims history to create an artificial claims history between Y and 15 years, this prediction can be compared with the real past claims history observed between Y and 15 years;
2. The artificial claims history can also be used to obtain BMS levels at time $\tau^{(1)}$ for each insured, which in turn will be used to obtain all BMS levels needed for the BMS models. AH methods can then be evaluated on their prediction quality for the model used for the Rating Database.

Because the objective of creating an artificial claims history is to improve the experience rating model, we will focus on the second type of evaluation.

4.2.1 Comparison Between Methods

We have up to 15 years of past information in the database used in Section 3. Our objective is to analyze the impact of using less than 15 years in BMS models. We divided the training dataset into five binds for cross-validation, where four of the five are used to estimate the parameters and one is used to check for prediction, and tried a BMS-Poisson model by varying the number of years used as past information. Except for the no-AH method which does not create any claims history, the other AH methods were used to complete the 15 years of past information needed for the model. Figure 9 shows the average logarithmic on the prediction fold. Several observations can be made:

1. As we should expect, the more years of past information we used, the more predictive the model is, for all AH methods.
2. The marginal improvement in the quality of prediction decreases with the number of years used. For example, for all AH methods, the logarithmic score improves a lot more between year 1 and year 7, than between year 7 and year 15.
3. Method AH1 is always the worst method to use. AH1 is even worse than the no-AH method, meaning that it seems better to not create an artifical claim history than to try to create it with the AH1 method.
4. AH2 and AH3 are the best methods, from a predictive point of view. Obtaining good results with AH2 was expected. However, the good result obtained using method AH3 is surprising because it supposes that each year of past experience is considered claim-free. This good performance can probably be explained by the fact that policyholders who do not claim stay with the same insurer for a longer time, which means that their past claims experience could be better than the average loss experience.

4.2.2 Estimated Parameters

Instead of selecting the best AH method based on their predictive quality, we can also compare the methods by analyzing the structural parameters of the BMS obtained for each model. Table 5 shows all the BMS parameters from a Poisson distribution, with 5, 7 and 9 years of experience used to construct the artificial claims history. Even if we saw in Figure 9 that the average logarithmic scores were close between all four AH methods for those years, we can see that the estimated parameters are clearly different. Once again, several observations can be made:

1. Deciding to not create artificial claim history, i.e. using method no-AH, means that the best BMS level an insured could have is $100 - X$ for $X = 5, 7, 9$ years, representing X consecutive years without claim. The

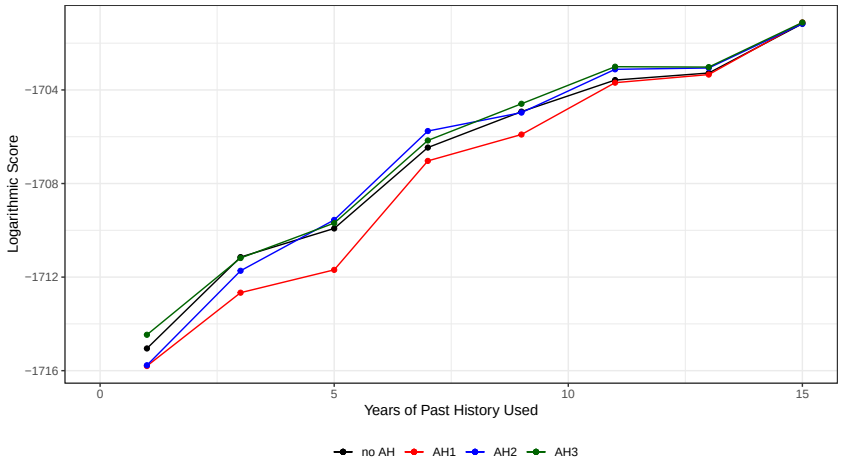

Figure 9: *Cross-validation results for the BMS model with artifical claims history*

	5 years				7 years				9 years			
AH Type	ℓ_{max}	ℓ_{min}	$\hat{\Psi}$	$\hat{\gamma}_0$	ℓ_{max}	ℓ_{min}	$\hat{\Psi}$	$\hat{\gamma}_0$	ℓ_{max}	ℓ_{min}	$\hat{\Psi}$	$\hat{\gamma}_0$
no-AH	108	95	4	0.0456	112	93	5	0.0422	111	91	6	0.0400
AH1	107	96	5	0.0501	113	87	6	0.0349	113	85	6	0.0330
AH2	113	85	6	0.0286	115	87	7	0.0308	113	85	6	0.0322
AH3	124	85	14	0.0152	123	85	11	0.0210	120	85	10	0.0226

Table 5: *Parameters of the BMS model for years 5, 7 and 9*

other methods do not have this limit, and method AH3 using only 5 years, for example, has a minimum BMS level of 85, which represents 15 years without claims. If the objective of the insurer is to eventually use a ratemaking structure using 15 years of past claims experience, but they do not have yet this amount of information, the no-AH method cannot obtain BMS relativities for levels smaller than $100 - X$. That means that if the ratemaking structure is put in production, several insureds already located at level $100 - X$ will not receive another discount even if they do not claim in the future. Other methods, on the other hand, give a solution to the insurer in its rating structure, in waiting to have enough past years to estimate the model with real data.

2. The jump parameter Ψ is much higher for method AH3 than for the other methods. Because the AH3 method supposes that all unobserved past experience (up to 15 years) is rewarded by a decrease of one level, some bad insureds might be rewarded too much just because they have lot of experience. In consequence, to distinguish between good and bad insureds, the penalty for a single claim could be higher to compensate.

3. Structural parameters for the AH2 method are more stable over time, and are close to the ones obtained with the BMS model using all 15 years of experience (see Table 4 with the Poisson-BMS model). This is interesting because it means that creating an artificial past claim history can be a temporary solution while waiting to have enough past years to estimate the complete model.

As mentioned earlier, all AH methods are worse than using real insurance data, which means that insurers should consider systematically keeping their insureds data in a longitudinal form. However, if the insurers do not have the data they want to create a BMS model, based on our comparison and analysis, it seems clear that an artificial claims history is a solution they can use to compensate. Among the three methods used to create an artificial claims history, the AH2 method was shown to be the best as it predicts adequately well and it seems capable of generating structural BMS parameters that approach those finally obtained when we use all 15 years of data.

5 Conclusion

Recent papers on BMS models showed that they are a good alternative to credibility models and panel data models. Past research showed that BMS models are at least as predictive as classic Poisson-gamma models (MVNB), while being much easier to apply in practice and easier to generalize for smoothing or to include dependence between different types of claims. This paper proposed an exhaustive analysis of how BMS systems should now be used with the actual data insurers collect. Even if we can find literature from the 80s or 90s about BMS models, the techniques proposed to estimate and calibrate the BMS models needed to be improved, because more granular data are now available. Other recent papers generalized the BMS theory with panel data, but none of them took the time to go into details and explain what the foundations of the model were. By linking BMS models with a simple GLM model with covariates that are associated with the past claims experience (Kappa-N model), this paper helps to understand how BMS model works. Figure 2 is a good example of a nicer way to illustrate BMS as it better explains the link between all parameters used in BMS. We explained how to deal with practical constraints while allowing for maximum usage of all data available. Finally, we used real insurance data from farm insurance to fit the BMS.

Often, insurers do not have all the past claims experience of their insureds, so we proposed a new way to create artificial past claims history for all insureds. While allowing for maximum usage of all data available, this new generalization improves the fit of the model as well as its prediction capacity. Again, we used real insurance data from farm insurance to show the difference between several methods to create artificial claims history.

Highlighting in greater detail the ways BMS can be interpreted allows us to more clearly identify how the model should be improved in the future. Indeed, BMS models can be generalized in many ways:

1. Special transition rules: instead of only rewarding insureds when they do not claim in a single year, we can study if other rewards should be given. Three, five or ten consecutive years without claim, for example, could reward a decrease of additional levels. Lemaire (2012) provides an impressive list of bonus-malus systems, where many different transition rules are used.
2. Multiple scope variables and multiple target variables: we model the number of claims on the machinery coverage and used all types of claims to compute BMS levels. Different jump parameters, one for past machinery claims and another one for other types of claims, could be analyzed.
3. Bonus-malus scales applied to the claim severity.
4. Combining the BMS model with statistical learning approaches to improve risks segmentation in ratemaking.

Future research on BMS, or any other past claims rating system, should approach these challenges in the future.

Acknowledgements

The author would like to thank Andra Crainic, Alexandre LeBlanc, Victor Lauzon, Guillaume Lepage and Qi Ann from the Co-operators General Insurance Company for their advice. Finally, the author would like to thank the Co-operators Chair in Actuarial Risk Analysis, and the Natural Sciences and Engineering Research Council (NSERC) of Canada for their financial support.

Appendix: Farm Insurance and Structure of the Data

The dataset comes from farm insurance contracts that range from January 1st, 2014 to December 31th, 2018. We focus our analysis on farm insurance, and more specifically on machinery coverage. This coverage, as the name implies, covers damages to any farm machinery, which includes tractors, but also swathers, combines, etc.

Item and Contracts

Table 6 shows an example of the available data for the machinery coverage of the farm insurance product. We see that each observation of the database corresponds to information at the item level. An item, for machinery coverage, corresponds to a specific tractor, or combine, from which specific information is available. The table shows that some covariates are at the item-coverage level (such as the type of machinery), while other covariates are at the contract level meaning that they are similar for each item of the same contract (for example, the province).

Policy Number	Item	Eff. Date	Coverage	...	Province	Type	# Claims	Loss Costs
1	1	2017-01-15	MACHINERY	...	Ontario	HARVEST	2	174,147
1	2	2017-01-15	MACHINERY	...	Ontario	SWATHER	1	12,445
1	1	2018-01-15	MACHINERY	...	Ontario	HARVEST	0	0
1	2	2018-01-15	MACHINERY	...	Ontario	SWATHER	0	0
1	1	2019-01-15	MACHINERY	...	Ontario	HARVEST	0	0
1	2	2019-01-15	MACHINERY	...	Ontario	SWATHER	1	87,112

Table 6: *Fictive Data Sample - Item Level*

Table 7 shows summary statistics from the database by year at the item level, as well as at the contract level. The total number of claims for the machinery coverage is also shown. We can see that the total number of claims at the item level (3206) is not the same as the total at the contract level (2783). This is because a unique incident can affect many items at the same time. The purpose of experience pricing is to analyze past claims to predict future claims experience, so we should make sure that an insured would not be penalized more than once for the same event, simply because a single disaster affected more than one item.

	Item Level			Contract Level		
Year	Number	Claims	Freq. (%)	Number	Claims	Freq. (%)
2014	126,515	556	0.439	22,347	500	2.237
2015	131,275	522	0.398	22,782	457	2.006
2016	136,757	655	0.479	23,246	591	2.542
2017	144,057	771	0.535	24,010	642	2.674
2018	154,345	702	0.455	24,939	593	2.378
Total	692,949	3,206	0.463	117,324	2,783	2.372

Table 7: *Summary statistics at the item level*

With a total of $692,949$ items insured for $117,324$ contracts, the average number of items insured by contracts is around six. The distribution of the number of items insured by contract can be seen in Figure 10. It is interesting to note that almost 50% of all farms only have one insured items, while approximately 10% of farms have more than 20 insured items. More precisely, 40 farms have more than 100 insured items, with a maximum of 212 for a single contract in 2016. The difference between small farms and larger farms is important, and it should be a priority to analyze the consequences of this difference in the BMS models in the future.

Meta-variable for the Contract Level

We saw with equation (2) that covariates used for *a priori* ratemaking can be used with the claim score. That would mean that the *a priori* premium, based on the covariates of each item of a contract, could be computed. We then start at the item level and suppose that the number of claims from item j, for contract t of insured i, has the following distribution:

$$N_{i,t}^{(j)} \sim \text{Poisson}(\lambda_{i,t}^{(j)}), \text{ with: } \lambda_{i,t}^{(j)} = d_{i,t}^{(j)} \exp(\boldsymbol{X}_{i,t}^{(j)} \beta^{(j)}).$$

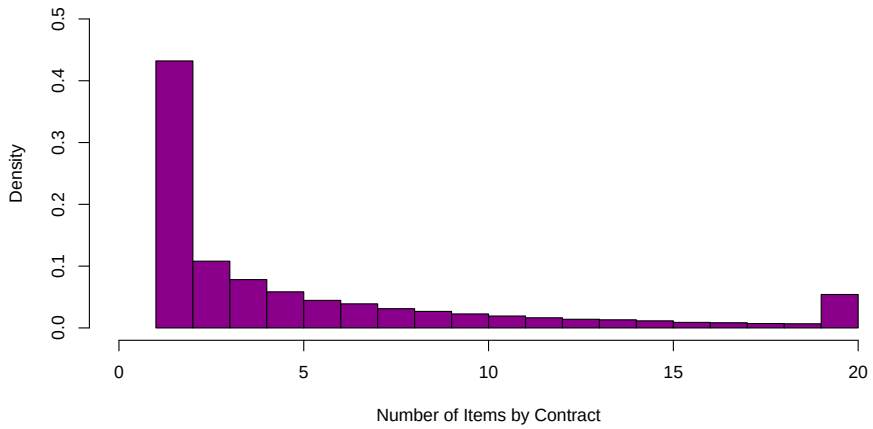

Figure 10: *Distribution of the number of items by contract*

The variable $d_{i,t}^{(j)}$ represents the risk exposure (in time) of item j, $\boldsymbol{X}_{i,t}^{(j)}$ represents risk characteristics of item j from contract t of insured i, and $\beta^{(j)}$ are parameters of the model to be estimated. Some covariates at the item level for the vector $\boldsymbol{X}_{i,t}^{(j)}$ were:

- Coverage Amount;
- Machinery Type;
- Machinery Make;
- Machinery Age;
- etc.

An estimation by maximum likelihood, using any GLM package, was performed to obtain $\widehat{\beta^{(j)}}$. With the estimated parameters, for each item of any contract, we were then able to have an estimation of $E[N_{i,t}^{(j)}]$, which can be seen as the *a priori* frequency part of the premium for each item.

Because the experience-rating algorithm is normally applied at the contract level and because we think that past claims will identify insureds that tend to claim more, for our illustration, we decided to analyze the loss experience of each insured at the contract level. That means grouping all items of a single contract into a single observation. This will also correct the situation mentioned previously where a single event resulted in damage to multiple items. The main problem with our goal of analyzing the number of claims at the contract level is that the majority of important covariates to be used in the rating only make sense at the item level. To correct this situation, we created a *meta-variable* ξ at the contract level representing the total risk of all items of the contract. More formally, we have:

$$\xi_{i,t} \quad = \quad \sum_{j}^{I} \widehat{\lambda_{i,t}^{(j)}} = \sum_{j}^{I} d_{i,t}^{(j)} \exp(\boldsymbol{X}_{i,t}^{(j)} \widehat{\beta^{(j)}})$$

with I being the number of items from contract t of insured i. The meta-variable $\xi_{i,t}$ represents the sum of expected claims frequency of each item of a contract. It can then be seen as an interesting risk measure using all the item information. With standard GLM modeling, the sum of prediction is equal to the sum of the observations. In other words, the $\sum_i \sum_t \sum_j N_{i,t}^{(j)}$ would be the same as the sum of all claims at the item level. Because we have fewer claims

at the policy level, we cannot use $\xi_{i,t}$ directly. Consequently, at the contract level, the number of claims can now be expressed as:

$$N_{i,t} \sim \text{Poisson}(\lambda_{i,t}), \text{ with: } \lambda_{i,t} = \exp(\boldsymbol{X}'_{i,t}\beta + \beta_{p+1}\log(\xi_{i,t})) \tag{5}$$

where $\boldsymbol{X}_{i,t}$ is the vector of dimension p that could include covariates at the contract level, such as the province, the civil status of the owner, the year of the contract, etc. We finally obtain a dataset at the contract level. Table 2 of Section 3.1 is an example. We then split the overall dataset into a training dataset (75%) and a test dataset (25%).

References

J. H. Abbring, P.-A. Chiappori, and J. Pinquet. Moral hazard and dynamic insurance data. *Journal of the European Economic Association*, 1(4):767–820, 2003.

A. Abdallah, J.-P. Boucher, and H. Cossette. Sarmanov family of multivariate distributions for bivariate dynamic claim counts model. *Insurance: Mathematics and Economics*, 68:120–133, 2016.

P. Albrecht. An evolutionary credibility model for claim numbers. *ASTIN Bulletin: The Journal of the IAA*, 15(1):1–17, 1985.

L. Bermúdez and D. Karlis. Multivariate inar (1) regression models based on the sarmanov distribution. *Mathematics*, 9 (5):505, 2021.

L. Bermúdez, M. Guillén, and D. Karlis. Allowing for time and cross dependence assumptions between claim counts in ratemaking models. *Insurance: Mathematics and Economics*, 83:161–169, 2018.

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

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

J.-P. Boucher and M. Pigeon. A claim score for dynamic claim counts modelling. 2019.

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

H. Bühlmann. Experience rating and credibility. *ASTIN Bulletin: The Journal of the IAA*, 4(3):199–207, 1967.

A. C. Cameron and P. K. Trivedi. *Regression analysis of count data*, volume 53. Cambridge university press, 2013.

M. Denuit, X. Maréchal, S. Pitrebois, and J.-F. Walhin. *Actuarial modelling of claim counts: Risk classification, credibility and bonus-malus systems*. John Wiley & Sons, 2007.

M. Denuit, D. Hainaut, and J. Trufin. *Effective Statistical Learning Methods for Actuaries II: Tree-Based Methods and Extensions*. Springer Nature, 2020a.

M. Denuit, D. Hainaut, and J. Trufin. *Effective Statistical Learning Methods for Actuaries III: Neural Networks and Extensions*. Springer Nature, 2020b.

E. W. Frees, R. A. Derrig, and G. Meyers. *Predictive modeling applications in actuarial science*, volume 1. Cambridge University Press, 2014.

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

J. Lemaire. *Bonus-malus systems in automobile insurance*, volume 19. Springer science & business media, 2012.

F. Pechon, M. Denuit, and J. Trufin. Home and motor insurance joined at a household level using multivariate credibility. *Annals of Actuarial Science*, pages 1–33, 2019a.

F. Pechon, M. Denuit, and J. Trufin. Multivariate modelling of multiple guarantees in motor insurance of a household. *European Actuarial Journal*, 9(2):575–602, 2019b.

V. Roel, K. Antonio, and G. Claeskens. Unraveling the predictive power of telematics data in car insurance pricing. *Available at SSRN 2872112*, 2017.

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

P. Shi and L. Yang. Pair copula constructions for insurance experience rating. *Journal of the American Statistical Association*, 113(521):122–133, 2018.

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

R. M. Verschuren. Predictive claim scores for dynamic multi-product risk classification in insurance. *ASTIN Bulletin: The Journal of the IAA*, 51(1):1–25, 2021.

R. Winkelmann. *Econometric analysis of count data*. Springer Science & Business Media, 2008.

V. R. Young and F. E. De Vylder. Credibility in favor of unlucky insureds. *North American Actuarial Journal*, 4(1): 107–113, 2000.