A GENERALIZED BONUS-MALUS SCALES MODEL FOR INSUREDS OF DIFFERENT SIZES

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

July 24, 2022

ABSTRACT

A Bonus-Malus Scales (BMS) model 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, insureds usually move down by a level if they do not claim during their contract, and move up a specific number of levels for each claim. The insured's level at the end of the year is then used to compute the next annual premium. Some recent papers generalized the BMS models theory using the newly available granular insurance data, which has only recently become available. Even if it has been shown that these new BMS approaches using panel data often offer better fit statistics and predictive measures than those obtained with many advanced panel data models, BMS models still have problems. One is that the rating system may appear unfair to many insureds because it does not recognize the initial risk of the insured. While some authors proposed creating different BMS for each type of insureds, we proposed a unique and general approach to that problem. Based on an improved technique for estimating the parameters of the BMS, we show that the new Generalized BMS model generates surcharges and discounts that depend on the size of the insured, or on the *a priori* risk. We apply this new generalized BMS model to real data from a major Canadian insurance company for their farm insurance products, where the size of each insured differs significantly.

***K*eywords** Claim Count, Ratemaking, Bonus-Malus Systems, Generalized Additive Models

1 Introduction

The general idea of any experience-based ratemaking model is quite simple: the insurer calculates the premiums of each insured based on their past claims experience. A Bonus-Malus Scales (BMS) model is this type of rating model, and is comprised of a class system with a finite number of levels where a relativity is assigned to each level. Depending on the transition rules of the BMS, insureds usually move down by a level if they do not claim during their contract, and move 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.

Classic BMS theory is described in detail in Lemaire (2012) or Denuit et al. (2007). Recently, Tan et al. (2015), Gyetvai and Ágoston (2018) and Ágoston and Gyetvai (2020) worked on finding optimal transition rules for classic BMS theory, via integer programming. The classic BMS theory has the problem of not taking advantage of the data structure that is now available from insurers. Indeed, to find the relativity of each BMS level, the classic theory used aggregate data and transition matrices based on assumptions about the heterogeneity distribution of a specific count distribution. Boucher and Inoussa (2014) first explained how BMS models theory could be generalized for granular data, where each insured can be observed for several contracts. Verschuren (2021) generalizes the approach for multi-products insurance, and adds a more flexible estimation methods, using generalized additive models (GAM) theory. Finally, Boucher (2022a) shows how BMS models are linked with simple GLM models that have covariates associated with the past claims experience.

Even if it were shown that these new BMS approaches often offer better fit statistics and predictive measures than those obtained using many advanced panel data models (as shown in Boucher and Pigeon (2019) in the case of automobile insurance), the rating system of the BMS models may appear unfair to many insureds as it does not recognize the initial risk of the insured, sometimes referred as the *a priori* risk. As explained in Denuit et al. (2007), Section 4.5.3, the BMS model generates the same surcharges and the same discounts for all insureds. Risky policyholders are expected to claim more than less risky policyholders, so they should normally also be expected to be less penalized for a claim. Similarly, risky insureds should be rewarded much more than lower risk insureds if they do not claim. Not considering the *a priori* risk in an experience-rating systems may appear to penalize risky insureds twice.

The problem of different *a priori* risks in experience-rating models has been known in the actuarial community for a long time. Indeed, even classic past claim rating models normalize the past experience of each insured i before applying claim penalties. For example, by designating $\lambda_{i,t}$ a measure of the *a priori* risk, for contract t of insured i, a normalized past experience $\frac{\sum_t n_{i,t}}{\sum_t \lambda_{i,t}}$ is used in the Buhlmann-Straub credibility model instead of $n_{i,\bullet} = \sum_t n_{i,t}$, which is used in the Bühlmann (1967) credibility model. More advanced models based on longitudinal data also include a weighted past claim experience to compute future premiums; see for example Bolancé et al. (2007), Abdallah et al. (2016) or Pechon et al. (2019).

To account for the difference between the *a priori* risk, many authors proposed using distinct BMS models, according to the *a priori* characteristics. Indeed, using the classic BMS theory based on aggregate data on an automobile insurance portfolio, Denuit et al. (2007) uses a BMS model for young drivers and another BMS model for older drivers. Using the BMS theory for granular data, Boucher (2022b) develops an iterative grouping procedure to partition a portolio from a farm insurance product to create five different BMS models based on the size of each farm. Even if it results in an interesting solution that generates a good log-likelihood and an interesting predictive score, using distinct BMS models is far from perfect. Indeed, dividing the portfolio and estimating all parameters of each group independently means that we consider each group to be fundamentally different, which is not the case. Moreover, creating separate BMS models can be a good idea when the BMS used depends on the age of the driver. Indeed, because the insurer knows when an insured will transfer to the other BMS, the insurer can propose simple solutions to limit the impact of the transition. However, transitioning between BMSs is not that simple for BMS models based on the size of the insured. For example, the size of the farm, which is defined by the number of insured items, can change at any time. Without an approach that considers all possible transitions from one BMS to another, the experience-rating model could lead to illogical results.

Situations where the insured could receive a large surcharge for past claims, even if no new claim was reported, could occur. An extensive study of all transition rules between BMS structures should be then be considered, which highly complicates the general rating system and eliminates the biggest advantage of the BMS: its understandability.

Instead of working on separate BMS models with special transition rules, we propose the creation of a unique and general approach where a single BMS structure is created, but where surcharges and discounts are based on the size of the insured or on the *a priori* risk. We apply this new Generalized BMS model to real data from a major Canadian insurance company for their farm insurance product where the size of each insured can differ significantly.

The paper has the following structure. In Section 2, we will carefully explain BMS models and apply the model to the farm insurance product. Compared to Boucher (2022a), an improved estimation technique for estimating the parameters of the BMS based on an recursive algorithm, is proposed. Following the estimation, results are discussed, and the problem of the size of the farm is highlighted. As expected, we show that the higher BMS levels are filled with bigger farms, and conversely bigger farms have an average BMS level much higher than smaller farms. In Section 3, a flexible approach to the BMS is proposed, where generalized additive theory is used. The new recursive algorithm is then generalized to find the parameters of the Generalized BMS model. This model is then applied to the same farm insurance portfolio. The last section concludes the paper.

2 Review of the BMS Model

2.1 Summary of Past Claims Rating Model

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 contract of period T that will consider all of the insured's past insurance contracts, or ther insured's past claims experience from $t = 1, \dots, T-1$. A new insured with $T = 1$ is simply someone without any past experience.

Concerning experience rating, Boucher (2022a) introduces two kinds of variables:

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

Then, formally, if we want to model the frequency part of the premium (i.e. the target variable) based only on the past number of claims (i.e. the scope variable), it means that we are looking to compute the conditional expected value of $N_{i,T}$, the number of claims of insured i for contract of period T, defined as :

$$E[N_{i,T}|\boldsymbol{n}_{i,(1:T-1)}, \boldsymbol{X}_{i,T}] \tag{1}$$

where $\boldsymbol{n}_{i,(1:T-1)}$ is a vector of all past number of claims between time 1 and time $T-1$ for insured i, and $\boldsymbol{X}_{i,T}$ is a vector containing the covariates used in the ratemaking, for contract T. This usually corresponds to information about the age of the insured, the marital status of the insured, etc. For the remainder of paper, we will continue to use the number of claims $N_{i,T}$ as the target, with the past number of claims $\boldsymbol{n}_{i,(1:T-1)}$ as the scope variables, but generalizations can be easily made.

2.1.1 The Bonus-Malus Scale Models

To model the number of claims of insured i, for contract T, the Poisson distribution of mean $\lambda_{i,T}$ is usually the starting point. It has a probability mass function defined as:

$$\Pr\big(N_{i,T}=n|\boldsymbol{n}_{i,(1:T-1)},\boldsymbol{X}_{i,T}\big)=\frac{\lambda_{i,T}^{n}\exp\left(-\lambda_{i,T}\right)}{n!}.$$

For experience rating, to differentiate between new insureds and insureds with experience, past claims rating models using the number of past claims, and using $\kappa_{i,t}=I(n_{i,t}=0)$, the number of contracts without claims, can be used. In such situations, with minor transformations, the parameter of the Poisson distribution can be expressed as:

$$\lambda_{i,T}=\exp(\boldsymbol{X}_{i,T}'\boldsymbol{\beta}+\gamma_0(100-\kappa_{i,\bullet})+\gamma_1 n_{i,\bullet}) \tag{2}$$

where for insured i, $n_{i,\bullet}=\sum_{t=1}^{T-1}n_{i,t}$ corresponds to the insured's total number of past claims, and $\kappa_{i,\bullet}=\sum_{t=1}^{T-1}\kappa_{i,t}$ is the sum of policy periods without claims. Another way of understanding the mean parameter of the model with $\kappa_{i,\bullet}$ and $n_{i,\bullet}$ is to rewrite $\lambda_{i,T}$ as follows:

$$\lambda_{i,T} \quad = \quad \exp(\boldsymbol{X}_{i,T}'\boldsymbol{\beta}+\gamma_0(100-\kappa_{i,\bullet}+\frac{\gamma_1}{\gamma_0}n_{i,\bullet})=\exp(\boldsymbol{X}_{i,T}'\boldsymbol{\beta}+\gamma_0\ell_{i,T}), \tag{3}$$

with:

$$\ell_{i,T} \quad = \quad 100-\kappa_{i,\bullet}+\frac{\gamma_1}{\gamma_0}n_{i,\bullet}$$

where γ_0 is the **Relativity** parameter, $\Psi=\frac{\gamma_1}{\gamma_0}$ is the **Jump** parameter. The new variable $\ell_{i,T}$, based on $\kappa_{i,\bullet}$ and $n_{i,\bullet}$, summarizes all past claim experience and is called a **claim score**.

Despite its simplicity, this regression model, which is called the Kappa-N model in Boucher (2022a), has several desirable qualities regarding the implied ratemaking structure:

- For an insured i without insurance experience, we would have $n_{i,\bullet}=0$, and $\kappa_{i,\bullet}=0$, which means an **entry level** of 100. In other words, a new insured without experience has 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 Ψ.
- The impact of a single claim on the premium is then roughly equal to Ψ years without claims.
- 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 $(1-\exp(-\gamma_0))\%$.

Compared with standard count models that do not use any covariates that are linked to experience rating, Boucher (2022a) showed that the Kappa-N model's quality of prediction is significantly better. However, one obvious problem with the Kappa-N model is the spectrum of possible values for the claim score $\ell_{i,T}$. Indeed, the Kappa-N model does not limit the claim score to minimum or maximum values. For example, with the database used in our numerical application, we saw that the some insureds had claimed 15 or even 20 times in the past. Even if those insureds also had 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, because 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 that deals with extreme situations that arises in the Kappa-N model would be to limit the values of all claim scores, but also to apply this limit to all past insurance contracts. By adding **maximum** and **minimum** values of the claim score to all past contract, Boucher (2022a) shows that the Kappa-N approach can be seen as a **Bonus-Malus Scales** (BMS) system (see Lemaire (2012) or Denuit et al. (2007) for a historial review). Instead of having a claim score, $\ell_{i,T}$ can be seen as the BMS level of insured i, or the BMS score.

Formally, BMS models can be defined with four structural parameters: the entry level ℓ_0, the jump parameter Ψ, the maximum level of the system ℓ_{max} and the minimum level of the system ℓ_{min}. For a specific insured, we then have the BMS level defined as:

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

where the level $\ell_{i,t}$ is limited to always be between ℓ_{min} and ℓ_{max} for all $t = 1, \ldots, T$.

2.2 Estimation Algorithm

For the Kappa-N model with a Poisson distribution, GLM packages, such as those already programmed in R, could be used. However, for BMS models, Boucher (2022a) mentioned that finding the best values for structural parameters Ψ, ℓ_{min} and ℓ_{max} is not direct. Indeed, 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 to their current contract.

To estimate the structural parameters, Boucher (2022a) proposes trying all possibilities for all structural parameters and select the BMS model with the biggest log-likelihood. An iterative technique is proposed that works by first initiating $\ell_{min}^{(0)} = 0$, and defining $\Psi^{(0)}$ from a standard Kappa-N model. Because $\Psi = \gamma_1/\gamma_0$, estimating a Kappa-N model means that we can obtain a first estimate of the jump parameter. Then, for step k:

1. With $\Psi^{(k-1)}$ and $\ell_{min}^{(k-1)}$, find $\ell_{max}^{(k)}$, the value from the best BMS model for all models with ℓ_{max} between 100 and a reasonnable maximum value;
2. With $\ell_{max}^{(k)}$ and $\ell_{min}^{(k-1)}$, find $\Psi^{(k)}$, the value from the best BMS model for all models with Ψ between 1 and a reasonnable maximum value;
3. With $\ell_{max}^{(k)}$ and $\Psi^{(k)}$, similarly, find $\ell_{min}^{(k)}$.

We repeat those steps until we reach 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.

2.2.1 An Improved Algorithm

It is possible to develop a better estimation algorithm to estimate the parameters of a BMS model. However, to better explain the functioning of this new estimation algorithm, it is easier to take a simple example in order to detail the different steps. Thus, let us assume three insureds have been observed for 10 years, as illustrated in Table 1, which describes the claims history of these insureds between years one and 10. We want to predict the number of claims for year 11, i.e. $N_{i,11}$ for $i = 1, 2, 3$. First, the values of $\kappa_{i,\bullet}$ and $n_{i,\bullet}$ are computed for each insured i.

Let us initially assume a BMS model with a value of $\Psi = 4$. The left graph of Figure 1 shows the evolution of the BMS level for these three insureds when no value of ℓ_{max} and ℓ_{min} is applied to the calculation of the levels (which corresponds to a Kappa-N model). The graph on the right of the same figure shows the evolution of the score through the years with $\ell_{max} = 115$ and $\ell_{min} = 85$.

The numerical example allows us to see that the most important thing to consider in the computation of the BMS score $\ell_{i,11}$ is not $\kappa_{i,\bullet}$ and $n_{i,\bullet}$, but rather the total number of jumps and the total number of drops, noted respectively as $j_{i,\bullet}$ and $d_{i,\bullet}$. More formally, the two variables are defined as:

$$d_{i,\bullet} = \sum_{t=1}^{T-1} (\ell_{i,t} - \ell_{i,t+1}) I(\ell_{i,t+1} < \ell_{i,t}) \quad , \quad j_{i,\bullet} = \sum_{t=1}^{T-1} (\ell_{i,t+1} - \ell_{i,t}) I(\ell_{i,t+1} > \ell_{i,t}) \tag{4}$$

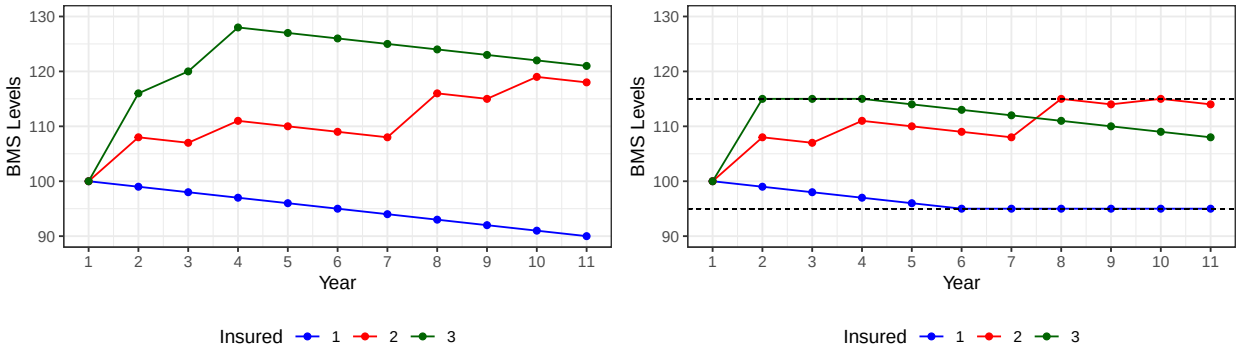

Figure 1: *Insureds with claim experience, with and without limits*

Insured	Years												Kappa-N model			BMS model		
i	1	2	3	4	5	6	7	8	9	10	$\kappa_{i,\bullet}$	$n_{i,\bullet}$	$d_{i,\bullet}$	$j_{i,\bullet}$	$\ell_{i,11}$	$d_{i,\bullet}$	$j_{i,\bullet}$	$\ell_{i,11}$
1	0	0	0	0	0	0	0	0	0	0	10	0	10	0	90	5	0	95
2	2	0	1	0	0	0	2	0	1	0	6	0	6	24	118	6	20	114
3	4	1	2	0	0	0	0	0	0	0	7	0	7	28	121	7	15	108

Table 1: *Insureds with claims experience, with $\Psi = 4$ for the BMS model*

Thus, for any BMS model, regardless of the values of ℓ_{max} and ℓ_{min}, the BMS score at time T of insured i is always calculated as:

$$\begin{aligned}
\ell_{i,T} &= \ell_{i,1} + \sum_{t=1}^{T-1} (\ell_{i,t+1} - \ell_{i,t}) \\
&= \ell_{i,1} - \sum_{t=1}^{T-1} (\ell_{i,t} - \ell_{i,t+1}) I(\ell_{i,t+1} < \ell_{i,t}) + \sum_{t=1}^{T-1} (\ell_{i,t+1} - \ell_{i,t}) I(\ell_{i,t+1} > \ell_{i,t}) \\
&= \ell_{i,1} - d_{i,\bullet} + j_{i,\bullet} = 100 - d_{i,\bullet} + \Psi \tilde{n}_{i,\bullet} \text{ , with } \tilde{n}_{i,\bullet} = \frac{j_{i,\bullet}}{\Psi}
\end{aligned} \tag{5}$$

For an unbounded BMS model, meaning no value has been set for ℓ_{max} and ℓ_{min} (i.e. the Kappa-N model), we obviously have $d_{i,\bullet} = \kappa_{i,\bullet}$ and $\tilde{n}_{i,\bullet} = n_{i,\bullet}$. Those two identities do not always hold for a BMS model using ℓ_{max} and ℓ_{min}, because in certain situations the BMS score could have been limited over time. For our example with three insureds, the last columns of Table 1 summarize the values of the jump and drop parameters of each insured at time 11 for each of the two BMS models. We can see the difference in values between $\kappa_{i,\bullet}, n_{i,\bullet}$ and $d_{i,\bullet}, j_{i,\bullet}$ of all insureds i, for the Kappa-N and the BMS models. Furthermore, we can verify for each insured that equation (5) can easily be used to compute the score at time 11, $\ell_{i,11}$, for both the Kappa-N and BMS models.

2.2.2 A Recursive Algorithm using d and $\tilde{n}$

The general idea of the proposed estimation method is to use Equation (2), from the original Kappa-N model, and substitute $\kappa_{i,\bullet}$ with $d_{i,\bullet}$, and $n_{i,\bullet}$ with $\tilde{n}_{i,\bullet}$:

$$\lambda_{i,T} = \exp(\boldsymbol{X}'_{i,T}\boldsymbol{\beta} + \gamma_0(100 - d_{i,\bullet}) + \gamma_1 \tilde{n}_{i,\bullet}). \tag{6}$$

This model, designated Kappa-N*, can be seen as a Kappa-N model where the number of observed claims would be $\tilde{n}_{i,\bullet}$, with $d_{i,\bullet}$ periods without claims. In other words, even if we are working with a BMS model where all BMS levels from time $t = 1, \ldots, T$ are limited by ℓ_{max} and ℓ_{min}, we will show that it ends up estimating a Kappa-N model.

The problem is that $j_{i,\bullet}$ is not necessarily the ultimate real number of jumps to be used to compute the final claim-score $\ell_{i,T}$, and $d_{i,\bullet}$ is not the correct number of drops. Indeed, both were computed with equations (4), which are based on a determined value of Ψ, for known values of ℓ_{max} and ℓ_{min}. Thus, if the estimated value $\hat{\Psi} = \hat{\gamma}_1/\hat{\gamma}_0$ from that Kappa-N* model turns out to be different from the selected value Ψ, the computed values of $\tilde{n}_{i,\bullet}$ and $d_{i,\bullet}$ are no longer valid and the results generated by this model are meaningless. This is why we instead consider $\tilde{n}_{i,\bullet}$ and $d_{i,\bullet}$ to be partially observed variables, and use an expectation-maximization algorithm to estimate the parameters instead.

We base our algorithm conditionally on $\{\gamma_0, \gamma_1\}$, the parameters associated with the jump parameter $\Psi = \gamma_1/\gamma_0$. The log-likelihood function to be maximized is based on equation (6), where $E[\tilde{n}_{i,\bullet}|\Psi^{(0)}]$ and $E[d_{i,\bullet}|\Psi^{(0)}]$ are needed. Equation (4), which depends on Ψ, is used to estimate those values at each iteration.

In summary, for a selected pair (ℓ_{max}, ℓ_{min}), we initiate $\Psi^{(0)} = \dfrac{\hat{\gamma}_1}{\hat{\gamma}_0}$ from the estimation of the Kappa-N model. Then, for step k, we have:

- **Computation**: The values of $d_{i,\bullet}^{(k)}$ and $\tilde{n}_{i,\bullet}^{(k)}$ are computed using Equation (4) and $\Psi^{(k-1)}$;
- **Maximization**: The log-likelihood of the Kappa-N* model, using $\tilde{n}_{i,\bullet}^{(k)}$ and $d_{i,\bullet}^{(k)}$, is maximized to estimate all parameters. An updated value of $\Psi^{(k)} = \dfrac{\hat{\gamma}_1}{\hat{\gamma}_0}$ is computed.

Both the computation and the maximization steps are repeated until convergence is reached. To obtain the best BMS model, the estimation algorithm must be used for all possible values of the pair (ℓ_{max}, ℓ_{min}). In other words, we have to use a grid of all possible values of (ℓ_{max}, ℓ_{min}), estimate all parameters with the algorithm, and finally choose the BMS model that generates the best log-likelihood.

Compared to the previous estimation procedure, the EM approach is only slightly faster. However, as we will see it in Section 3, there are other important uses of this new algorithm.

2.3 Summary of the Numerical Illustration

2.3.1 Data Used

To illustrate the BMS model, Boucher (2022a) used farm insurance data from a major insurance company in Canada. The same dataset is used here. The general form of the data is like the sample shown in Table 2, where each line of the database corresponds to a specific coverage from an annual contract. For each observation, we have information about the insured, the contract, the items covered, but we also see the date of the first insurance contract with the insurer. Information about claims that happened during that period of time is also available.

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 around 60 [1]. The maximum available number of years of past claims experience for all insureds is 15 years, and only insurance experience with the same insurer is available. That means that we considered the first year of insurance of any insured to be his 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 have been a new insured without any prior experience in 2003.

[1] Farms are sometimes passed from generation to generation. Insurance experience would not be reset in such a case.

Policy Number	Number of Items	Effective Date	First Insurance	Coverage	...	Province	Number of Claims	Costs of Claims
...	...	...	...	...	...	...	...	...
125721	2	2017-01-15	1995-01-15	MACHINERY	...	Ontario	2	186,592
125722	15	2017-03-22	2013-03-22	MACHINERY	...	Quebec	0	0
125723	1	2016-01-11	1993-11-05	MACHINERY	...	Manitoba	1	18,889
125724	27	2018-02-17	2018-02-17	MACHINERY	...	Nova Scotia	1	7,444
...	...	...	...	...	...	...	...	...

Table 2: *Fictive Data Sample - Contract Level*

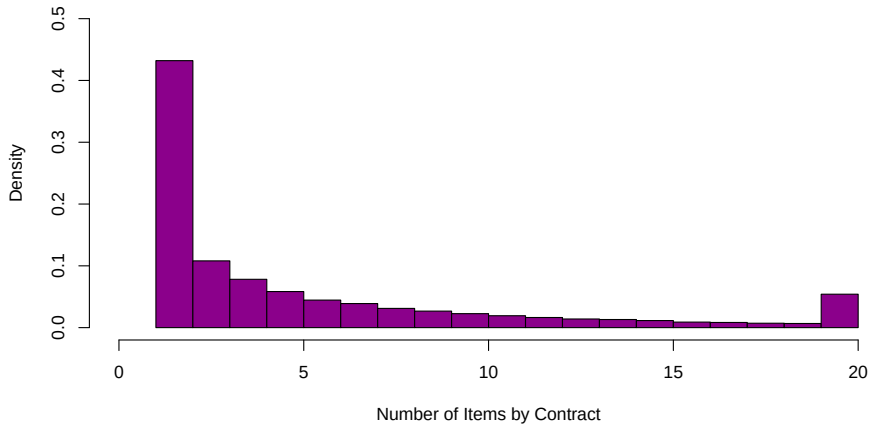

Figure 2: *Distribution of the number of years of experience with the insurer (left), distribution of the number of items by contract (right)*

For the farm insurance product, an item corresponds to a specific tractor or combine which specific information is available. With a total of approximately $700,000$ insured items insured for more than $120,000$ contracts, the average number of items insured per contract is around 6. The distribution of the number of items insured per contract can be seen in Figure 2. 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 more than 200 insured items for a single contract. As we will see in the next sections, the difference between small farms and larger farms is important for BMS models.

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, we decided to analyze the loss experience of each insured at the contract level. That means grouping all items from a single contract into a single observation. This will also correct the situation where a single event resulted in damage to multiple items.

2.3.2 Estimated Parameters of the BMS Model

The estimated parameters of the Poisson BMS model are shown in Table 3. The log-likelihood value is shown. 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 descriptions of other scores) to define the prediction quality. To better understand the results obtained for the Poisson

Distributions	BMS Parameters				log-likelihood (train)	Log. Score (test)
	ℓ_{max}	ℓ_{min}	$\hat{\Psi}$	$\hat{\gamma}_0$		
Poisson	116	85	6	0.0312	-8,490.026	2,857.029

Table 3: *Results of the Poisson BMS model*

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 6 years without a claim to return to the original premium.
- The value of γ_0 is 0.0312. That means that the penalty for a claim is equal to $\exp(0.0312 \times 6) - 1 = 20.6\%$, and each year without a claim decreases the premium by $1 - \exp(-0.0312) = 3.07\%$.
- The maximum BMS level is $\ell_{max} = 116$, meaning that the maximum surcharge, compared to level 100, is $\exp(0.0312 \times 16) - 1 = 64.7\%$;
- The minimum BMS level is $\ell_{min} = 85$, meaning that the minimum surcharge, compared to level 100, is $1 - \exp(-0.0312 \times 15) = 37.3\%$.

As we can see, these basic results are found and computed easily. This method of computing the surcharges and discounts would clearly be useful to any insureds, brokers or administrators. It is simple to explain to insureds how large their penalties for a claim will be, and how long they will be penalized for that claim. Another interesting result of the BMS model is that all insureds will have a premium located between 0.627 and 1.647 times the basic premium for a new insured, at level 100. This narrowly limits the range of premiums.

2.3.3 Problems With the Size of Farms

By comparing the predicted and the observed claims frequency on the training and the test datasets, Boucher (2022a) showed that the BMS model seems to fit the data well. We see that classifying insureds by their claim score (or BMS level) works well as the insureds with higher levels have worse claims experience than insureds with lower levels.

However, the size of each farm in the insurance portfolio is different and size has a direct impact of the past rating model. Figure 2 showed the distribution of the number of insured piece of machinery (called items) per farm. The BMS model used here generates the same surcharges and the same discounts for all insureds. However, because large farms are expected to have more claims than smaller ones, they should normally also be expected to be penalized less for a claim. Similarly, a large farm should be rewarded much more for a year without a claim. An experience-rating system that does not recognize this type of situations may appear to penalize larger farms twice.

That means that the connection between the BMS levels and the size of the farm is noteworthy. To more clearly see the impact of the number of insured items on each farm, using the BMS model from Table 3, we compute the BMS level of each contract in the database. For each BMS level, we compute the average number of insured items from each farm for that level. Similarly, we computed the average BMS level based on the number of insured items. Figure 3 illustrates the result. As we expect, the higher BMS levels are filled with bigger farms, and bigger farms have an average BMS level much higher than smaller farms. Despite the prediction quality of the BMS model, it is clear that the BMS model could seem unfair to many insureds and regulators because it does not recognize the initial risk of the insured [2]. To correct this situation and promote the use of the BMS in practice, the BMS model should be generalized.

[2]It may be paradoxical, however, that regulators might prefer to use pricing models that are less accurate than a BMS model (in both the fit statistics and the prediction quality), simply because the penalty structure appears unfair. If the fit and prediction of a BMS model is better than another rating model, it most likely means that the BMS model is more accurate and thus, in some sense, more fair.

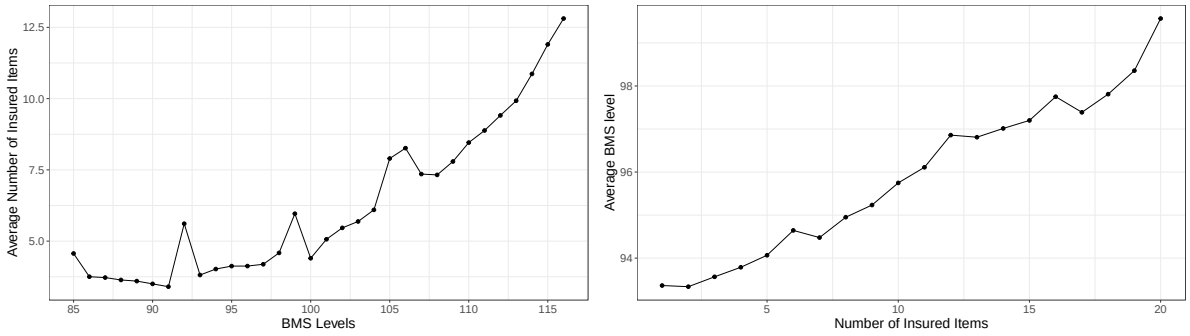

Figure 3: *Average number of insured items by BMS level (left) and average BMS level by number of insured items*

3 A Flexible Approach to Past Claims Models

To define the *a priori* risk of a farm, we used its number of insured items, noted $\omega_{i,T} \in \{1, \ldots, W\}$. We refer to Figure 2 for the distribution of the number of items per contract. Because $\omega_{i,T}$ is a discrete variable, this variable was appropriate for the recursive division algorithm, as used in Boucher (2022b). However, instead of dividing the portfolio into defined groups, we propose the development of a much more flexible approach.

3.1 Generalized Kappa-N model

To generalize the BMS model by accounting the size of the risk in the penalty structure, we will use the same method as the one summarized quickly in Section 2.1.1. The idea is to develop the BMS model by generalizing a Kappa-N model. Indeed, we used the Kappa-N model to clearly isolate and identify a claim score.

First, we use the mean function defined by equation (2) that we generalize to add two interactions with the size of the insured $\omega_{i,T}$. Because we cannot clearly identify the link between past claims experience and $\omega_{i,T}$, functions $f_1(\omega_{i,T})$ and $f_2(\omega_{i,T})$ are introduced. It leads to the creation of a new model called Generalized Kappa-N that has a mean function expressed as:

$$\begin{aligned} \lambda_{i,T} &= \exp(\boldsymbol{X}_{i,T}'\boldsymbol{\beta} - \gamma_0\kappa_{i,\bullet}f_1(\omega_{i,T}) + \gamma_1 n_{i,\bullet}f_2(\omega_{i,T})) \\ &= \exp(\boldsymbol{X}_{i,T}'\boldsymbol{\beta} - \kappa_{i,\bullet}s_1(\omega_{i,T}) + n_{i,\bullet}s_2(\omega_{i,T})) \end{aligned} \tag{7}$$

where $s_1(\omega_{i,T})$ and $s_2(\omega_{i,T})$, two non-parametric functions, are used to replace the parametric functions $f_1(\omega_{i,T})$ and $f_2(\omega_{i,T})$. This more flexible approach allows us to better understand the connection between the size of the insured $\omega_{i,T}$ and the past claims experience. Using splines as non-parametric functions, which are associated with a Poisson distribution, allows us to use the generalized additive models (GAM) theory (see Wood (2006) for an overview).

As with the Kappa-N model in Section 2.1.1, the idea is to modify the form of the mean $\lambda_{i,T}$, expressed in equation (7), to create a generalized claim score $\ell_{i,T}$. For the generalized claim score, each claim increases the score by the jump parameter Ψ and each year without claim decreases the claim score by Υ, a new parameter called the **drop parameter**.

The generalized claim score can be expressed in at least two ways:

1. Generalized Claim-Score #1:

Insured (i)	ω_i	$\log(\omega_i)$	Generalized Claim-Score #1			Generalized Claim-Score #2		
			$\Psi_i^{(1)}$	$\Upsilon_i^{(1)}$	$\gamma_{0,i}^{(1)}$	$\Psi_i^{(2)}$	$\Upsilon_i^{(2)}$	$\gamma_{0,i}^{(2)}$
1 (blue)	1	0	4.15	1	0.0702	10.87	2.62	0.0268
2 (red)	12	2.48	6.86	1	0.0171	4.38	0.63	0.0268

Table 4: *Parameters for the Generalized Kappa-N model for both insureds*

$$\begin{aligned}\lambda_{i,T} &= \exp\left(\boldsymbol{X}_{i,T}'\boldsymbol{\beta} + s_1(\omega_{i,T}) \times \left(-\kappa_{i,\bullet} + \frac{s_2(\omega_{i,T})}{s_1(\omega_{i,T})} n_{i,\bullet}\right)\right) \\ &= \exp\left(\boldsymbol{X}_{i,T}'\boldsymbol{\beta} + \gamma_0(\omega_{i,T}) \times (-\Upsilon\kappa_{i,\bullet} + \Psi(\omega_{i,T}) n_{i,\bullet})\right) \\ &= \exp\left(\boldsymbol{X}_{i,T}'\boldsymbol{\beta} + \gamma_0(\omega_{i,T}) \times \ell_{i,T}^{(1)}(\omega_{i,T})\right),\end{aligned}$$

where $\ell^{(1)}(\omega_{i,T}) = -\Upsilon\kappa_{i,\bullet} + \Psi(\omega_{i,T}) n_{i,\bullet}$, the generalized claim-score depends on the size of the insured $\omega_{i,T}$. The other parameters are now defined as:

- $\gamma_0(\omega_{i,T}) = s_1(\omega_{i,T})$: the relativity parameter depends on the size of the insured $\omega_{i,T}$;
- $\Psi(\omega_{i,T}) = \frac{s_2(\omega_{i,T})}{s_1(\omega_{i,T})}$: the jump parameter depends on the size of the insured;
- $\Upsilon = 1$: the drop parameter does not depend on the size of the insured.

2. Generalized Claim-Score #2:

$$\begin{aligned}\lambda_{i,t} &= \exp(\boldsymbol{X}_{i,T}'\boldsymbol{\beta} + (-s_1(\omega_{i,t})\kappa_{i,\bullet} + s_2(\omega_{i,t}) n_{i,\bullet}) \\ &= \exp(\boldsymbol{X}_{i,T}'\boldsymbol{\beta} + \gamma_0^* \times (-\Upsilon(\omega_{i,t})\kappa_{i,\bullet} + \Psi(\omega_{i,t}) n_{i,\bullet}) \\ &= \exp\left(\boldsymbol{X}_{i,T}'\boldsymbol{\beta} + \gamma_0^* \times \ell_{i,t}^{(2)}(\omega_{i,t})\right),\end{aligned}$$

where $\ell^{(2)}(\omega_{i,T}) = -\Upsilon(\omega_{i,T})\kappa_{i,\bullet} + \Psi(\omega_{i,T}) n_{i,\bullet}$, the generalized claim-score, also depends on $\omega_{i,T}$. For this second generalized claim-score, the parameters are now defined as:

- γ_0^*: the relativity parameter, which comes from the Kappa-N model, is used to obtain normalized parameters that can easily be compared with the parameters of the Kappa-N model. The relativity parameter does not depend on the size of the insured $\omega_{i,T}$;
- $\Psi(\omega_{i,T}) = \frac{s_2(\omega_{i,T})}{\gamma_0^*}$: the jump parameter depends on the size of the insured.
- $\Upsilon(\omega_{i,T}) = \frac{s_1(\omega_{i,T})}{\gamma_0^*}$: the drop parameter depends on the size of the insured;

A numerical application can be used to show the differences between the two generalized claim-scores. Using the same dataset as in Section 2.3.1, the Generalized Kappa-N model is used. A Poisson distribution is again used to model the number of claims.

Figure 4 shows the functions $\Psi(\omega_{i,T})$, $\Upsilon(\omega_{i,T})$ and $\gamma_0(\omega_{i,T})$ for the two generalized claim scores, based on the smoothing functions $s_1(\omega_{i,T})$ and $s_2(\omega_{i,T})$ estimated by a Poisson GAM. The dashed line shows the estimates of Ψ, Υ and γ_0 from the Kappa-N model. It is interesting to see that the Generalized Kappa-N model allows us to develop a flexible relationship between the parameters Ψ, Υ and γ_0 and the risk size ω.

3.1.1 Example of Two Generalized Claim Scores

To better interpret the generalized claim score and to analyze the differences between the two generalized claim scores, we provide an example with two insureds, both summarized in Table 4 and illustrated with specific dots in all the graphs in Figure 4):

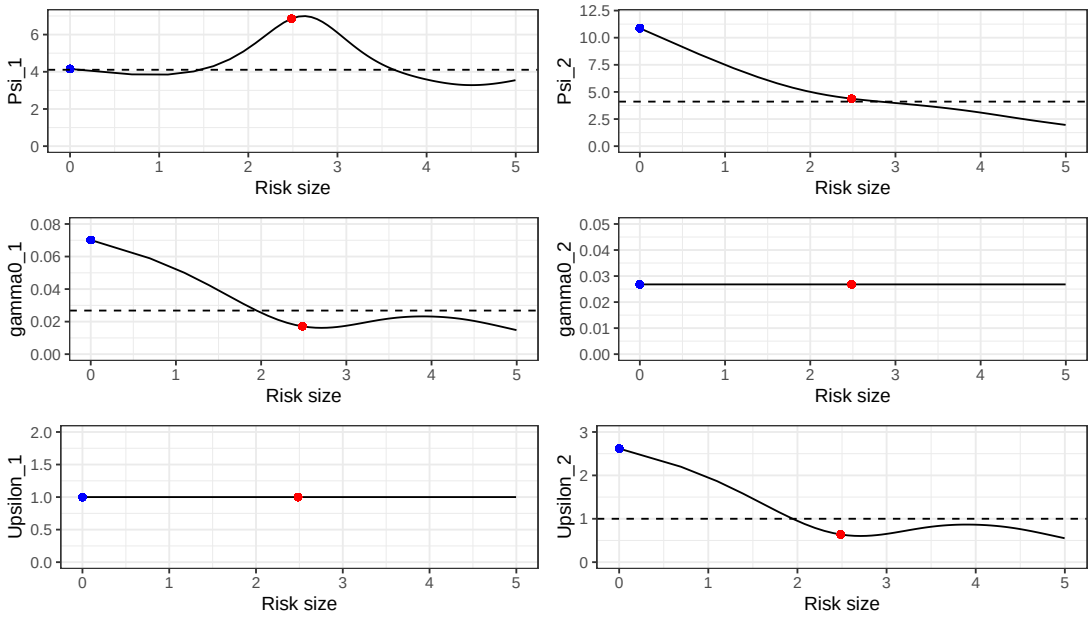

Figure 4: *Smoothed functions from the Generalized Kappa-N model (generalized claim-score #1: left, generalized claim score #2: right)*

1. An insured with only one insured item (i.e. $\omega_{i,T} = 1$), which is represented by the blue dots. If the insured claims, the value of the first score $\ell^{(1)}(\omega_{i,T})$ will increase by $\Psi^{(1)} = 4.15$ and the value of the second score $\ell^{(2)}(\omega_{i,T})$ will increase by $\Psi^{(2)} = 10.87$. Because $\gamma_0^{(1)} \neq \gamma_0^{(2)}$, we can show that the resulting surcharge will be the same.
2. An insured with twelve insured items ($\omega_{i,T} = 12$), which are represented by the red dots. If no claim is made, the insured will be rewarded by a drop of $\Upsilon^{(1)} = 1$ for the first score $\ell^{(1)}(\omega_{i,T})$, and a drop of $\Upsilon^{(2)} = 0.63$ for the second score $\ell^{(2)}(\omega_{i,T})$. Again, despite the difference between the scores, the resulting discount will be the same for each score.

Another way to compare the two scores is shown in Figure 5 with the same two insureds. Both insureds are at level 100. The figure shows how the two generalized scores can be interpreted. For the first score, illustrated by the top graph, we see two relativity curves ($\gamma_0(s_1)$ in blue and $\gamma_0(s_2)$ in red). Policyholder #1, in blue, will be rated by the value shown by the blue curve, and the second policyholder will be rated by the relativities associated with the red curve. We also see that the jump parameter is not the same for both insureds. Indeed, the first insured has a jump parameter of $\Psi(s_1) \approx 4$, and the second insured has a jump parameter of $\Psi(s_2) \approx 7$. For the first generalized score, the drop parameter Υ is the same regardless of the size of the insured (in purple), meaning that both insured moves down by one level if they do not claim.

The second generalized score is illustrated in the bottom graph of the same figure. This time, we see that only one relativity curve γ_0 applies to all insureds. The jump parameters $\Psi(s_1)$ and $\Psi(s_2)$ diffres depending on the size of the insured. The jump parameters from generalized claim score #2 are also very different than the jump parameters of generalized claim score #1. Indeed, we have $\Psi(s_1) < \Psi(s_2)$ for the first score, and $\Psi(s_1) > \Psi(s_2)$ for the other score. Drop parameters ($\Upsilon(s_1)$ and $\Upsilon(s_2)$) also depend on the size of the insured.

3.1.2 Practical Considerations

Table 5 shows the log-likelihood for the Generalized Kappa-N model ($L^{(gKN)} = -8486.277$) and the log-likelihood obtained for the Kappa-N model ($L^{(KN)} = -8506.237$). Adding an interaction between the size of the insureds

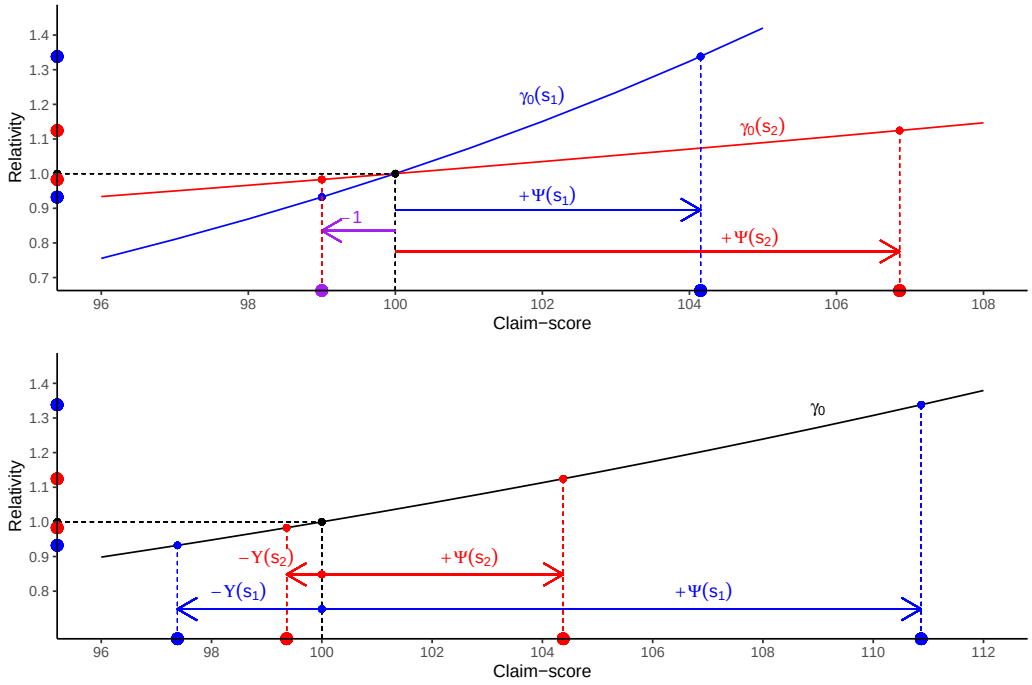

Figure 5: *Examples of the penalty structure for two insureds for the Generalized Kappa-N model*

and the total number of claims, and between the size of the insured and the total number of years without a claim significantly improves the log-likelihood.

The Kappa-N model supposes that each year without claims is rewarded by $\Upsilon = 1$. Maximum likelihood estimation of the joint density of the Kappa-N shows a value of $\hat{\Psi} = 4.11$. For practical purposes and ease of use, Ψ can be rounded to obtain $\Psi = 4$. Consequently, the rating system can be understood more easily by the legislative authorities that regulate pricing, by the various administrators of insurance companies and maybe more importantly, by policyholders. The same transformation should be done with the Generalized Kappa-N model. In other words, Ψ and Υ could also be rounded. However, $\Upsilon^{(2)}$ is more problematic because its lowest value is around 0.2 for large insureds. Rounding $\Upsilon^{(2)}$ to 0 in that case will not be useful approximation, because it would mean that a year without a claim would not lead to as discount. It could be possible to round to the nearest 0.5 instead with the constraint that $\Upsilon^{(2)} \geq 0.5$.

Table 5 also shows the impact of rounding Ψ and Υ of the Generalized Kappa-N model. The impact of rounding the claim-score parameters was less notable for the Kappa-N model, as seen by the small difference in the log-likelihood between both models. For the Generalized Kappa-N model, the impact seems more significant. In Figure 4, we see that $\Psi^{(1)}$ and $\Psi^{(2)}$ can easily be rounded without loosing to much of precision. Figure 6 shows the distribution of the rounded values of Ψ and Υ by the number of insured items, where $\Upsilon^{(1)}$ is not shown because it is always equal to 1 (by design).

3.2 Generalized BMS model Using the Size of the Insured

The problems present in the Kappa-N model also exist in the Generalized Kappa-N model. First, by using a generalized claim score based on $\kappa_\bullet$ and $n_\bullet$, the rating model does not allow for forgiveness and a claim will always have an impact on the premium. Moreover, because the Generalized Kappa-N model has no minimum or maximum values for the claim-score ($\ell^{(1)}$ or $\ell^{(2)}$), it can lead to situations where the computed surcharges or discounts would be too large to be

Model	Rounded	Loglike. (train)	Log. Score (test)
Kappa-N	No	-8506.237	-2858.234
.	Yes	-8506.303	-2858.179
Generalized Kappa-N	No	-8486.277	-2854.272
.	Yes (score 1)	-8487.000	-2855.131
.	Yes (score 2)	-8489.187	-2854.993

Table 5: *Kappa-N and Generalized Kappa-N models with rounded claim-score parameters*

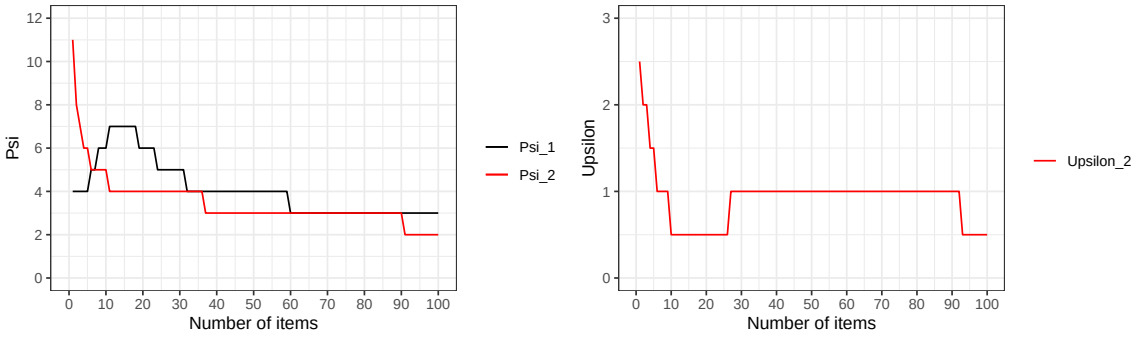

Figure 6: *Rounded values of Ψ and Υ (by 0.5) by the number of insured items*

applied in practice. Figure 7 shows the distribution of claim scores by the number of insured items, from which we observe that the spectrum of possible values for $\ell^{(1)}$ or $\ell^{(2)}$ is quite large, especially for small insureds.

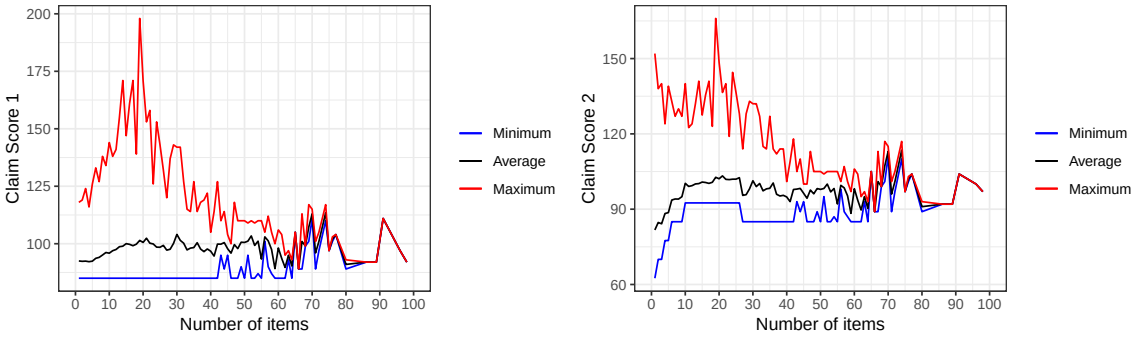

Figure 7: *Minimum, maximum and average values of claim-scores $\ell^{(1)}$ or $\ell^{(2)}$ by the number of insured items*

Compared to a Kappa-N model, the general idea of the BMS model is to limit the maximum surcharge and the maximum discount of the insured, for the current contract but also for all available past contracts. The same kind of modifications have to made for the Generalized Kappa-N model.

Like we did when adapting the BMS model from the Kappa-N model, we can adjust several models using all possible values of ℓ_{max} and ℓ_{min}. However, this approach is not ideal because it assumes that the optimal values of ℓ_{max} and ℓ_{min} are the same for all policyholders regardless of their ω size. The generalized score #1, illustrated in Figure 8, better illustrates the problem. Seeking an ideal value of ℓ_{max} impacts each relativity curves differently. For example, the graph on the left of Figure 8 illustrates a fictitious situation where $\ell_{max} = 132$. In this example, which can be understood with the help of Figure 5, we see that the impact for policyholders would be very different depending on the size of the insured, and thus would very differently impact their relativities. Indeed, this limit ℓ_{max} would mean

that the maximum surcharge for insureds in red would be only 17.4%, which is probably too low, but would be about 61.6% for insureds in green, which is probably too high. Consequently, we could take a different approach and assume that instead of trying to obtain an optimal ℓ_{max} to limit the insured's premiums, we could simply find the maximum relativity for all policyholders, regardless of their size. This situation is illustrated in the right-hand graph in the same figure. This time, a maximum relativity of 1.25 would imply a maximum BMS level of about 115 for the insureds in green and 144 for the insureds in red. Such a situation would only create more problems, where only one or even less than one claim can reach the maximum BMS level for the green group of insureds, while the limit becomes potentially unattainable for the insureds in the red group. In other words, regardless of whether the left or right approach in Figure 8 is used, it gives the impression that this kind of BMS will be well suited to the policyholders in the blue group, but will cause problems for all other policyholders.

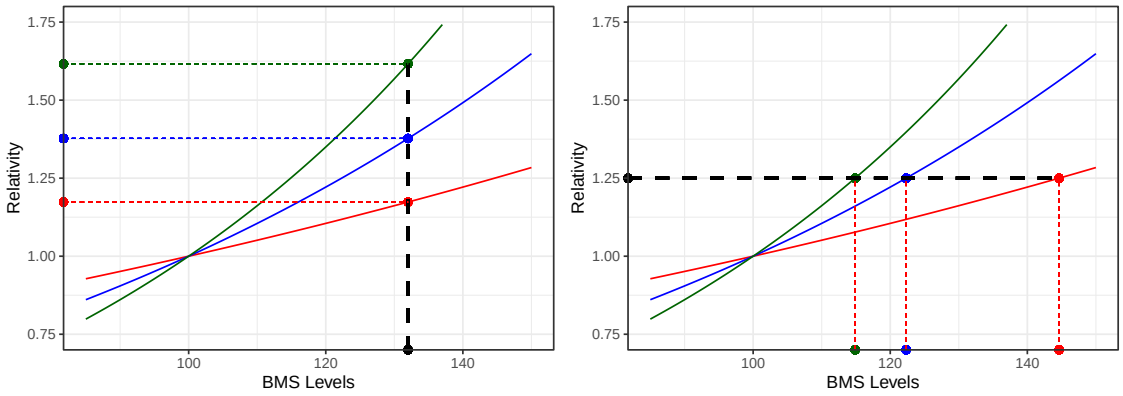

Figure 8: *Problems of limiting the claim score*

3.2.1 How to Find the Best Limits for the Generalized Claim Score

We have to develop an approach to limit the generalized claim score. To achieve this, we need to work in stages.

The first step is to fit a Generalized Kappa-N model that does not suppose any upper or lower bounds on the generalized claim score. That means that we have to work with the structural parameters, $\Psi^{(gKN)}$, $\Upsilon^{(gKN)}$ and $\gamma_0^{(gKN)}$, estimated by the Generalized Kappa-N model. We then use all possible combinations of $\ell_{max} \in \{100, 101, \ldots, 200\}$ and $\ell_{min} \in \{85, 86, \ldots, 100\}$ to compute the total log-likelihood $L(\ell_{min}, \ell_{max})$ of each partial-BMS model[3] for all possibilities:

$$L(\ell_{min}, \ell_{max}) = \sum_{i=1}^{n} \log(\Pr(N_{i,T}|\ell_{i,T}^{(gKN)})) \tag{8}$$

where $\ell_{i,T}^{(gKN)}$ is the generalized claim score, limited by ℓ_{min} and ℓ_{max}, but computed with $\Psi^{(gKN)}$, $\Upsilon^{(gKN)}$ and $\gamma_0^{(gKN)}$. Figure 9 shows the evolution of the total log-likelihood $L(\ell_{min}, \ell_{max})$ for different values of ℓ_{max} and ℓ_{min}. A partial-BMS model with $\ell_{max}^* = 125$ and $\ell_{min}^* = 85$ allows us to obtain a log-likelihood of $L^* = L(85, 125) = -8,482.025$ (identified by the red dot on the graph).

To obtain ℓ_{min} and ℓ_{max} values that could depend on ω, a simple solution would be to have a unique optimal value of ℓ_{max} and ℓ_{min} for each value of ω from the portfolio. That means that we take the values of ℓ_{max} and ℓ_{min} that maximize the log-likelihood for each ω:

[3]We call it "partial" because the parameters Ψ, Υ and γ_0 are not yet estimated with the BMS structure.

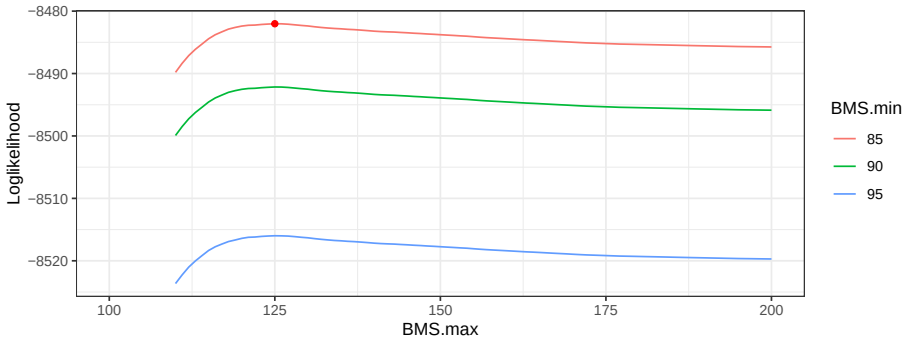

Figure 9: *Total log-likelihood of the model by using various values of ℓ_{min} and ℓ_{max}*

$$L^*_\omega = \max_{\ell_{min},\ell_{max}} L_\omega(\ell_{min},\ell_{max}), \text{ for } \omega = 1,2,\ldots,70. \tag{9}$$

where the maximum number of insured items, W, has been limited to 70, meaning that all insureds with 70 or more insured items are now grouped together. Figure 10 shows the result based on the number of insured items. Using this approach significantly improves the fit of the partial-BMS model because a total log-likelihood of $\sum_{\omega=1}^{W} L^*_\omega = -8,462.089$ is found.

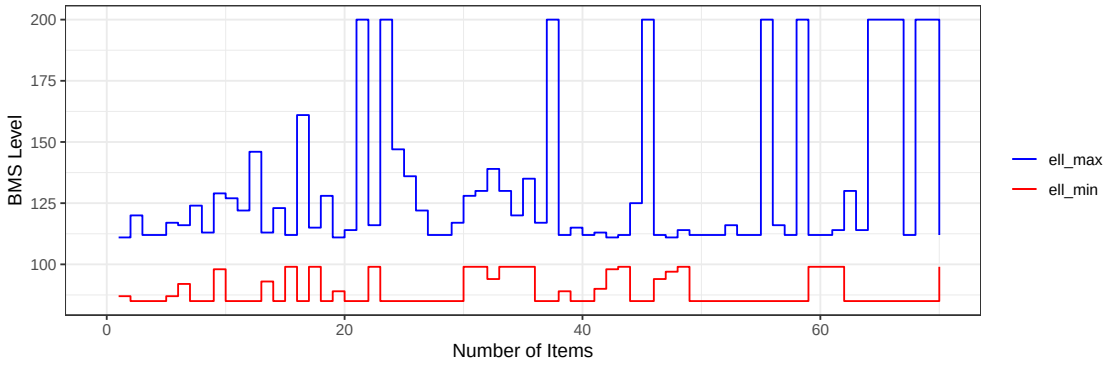

Figure 10: *Values of ℓ_{max} and ℓ_{min} generating the best log-likelihood for each ω*

For this type of solution, the possibility of overfitting should obviously be considered and cross-validation analyses, or comparisons of the different models with the test database, should be studied. Nevertheless, at this stage of construction of our model, overfitting is not yet our main concern. The problem with this solution, in which we use different values of ℓ_{max} and ℓ_{min} for each ω, is similar to our main criticism of the recursive partition model used in Boucher (2022b). Using different values of $\ell_{max}(\omega)$ and $\ell_{min}(\omega)$ could probably causes serious practical problems where an insured who could have large variations in premiums simply by adding or removing an item from their insurance policy. For example, insureds with 23, 24 or 25 insured items, as we can see in Figure 10, could have large variations in their premiums because ℓ_{max} alternates between 200, 123 and 200.

3.2.2 Finding the Best Path

In order to ensure that the rating structure we are trying to implement is consistent, we need to limit the variation of ℓ_{max} for insureds with approximately the same number of insured items ω. One solution is to only allow a variation of plus or minus 1 between the values of ℓ_{max} for successive values of ω. The objective can also be interpreted as trying to find the optimal *path*, from $\omega = 1$ to $\omega = 40$, with the constraint that $\left|\ell_{max}(\omega+1) - \ell_{max}(\omega)\right| \leq 1$

for all $\omega = 1, \ldots, W-1$. Even with such a constraint, the total number of paths to verify and compute is much too large.

Instead, we develop an algorithm to find the best path[4]. The pseudo-code of the **Best Path** algorithm in described in Algorithm 1. To apply the Best Path algorithm, we first need to compute all the log-likelihoods for $\ell_{min} \in \{85, \ldots, 100\}$, $\ell_{max} \in \{100, \ldots, 200\}$, and $\omega \in \{1, 2, \ldots, 70\}$. Those values are then stored in $\boldsymbol{L}$, and we suppose that the log-likelihood does not exist for other values of ℓ_{min} or ℓ_{max}. The general idea of this algorithm is to work backward to update $\boldsymbol{L}$ at each step, as explained in Algorithm 1.

Algorithm 1 Best Path

Require: $\boldsymbol{L}$, the log-likelihoods obtained for all values of ℓ_{min}, ℓ_{max} and ω
  for k in $\{69, 68, \ldots, 1\}$ **do**
    for l_1 in $\{85, 86, \ldots, 100\}$ **do**
      for l_2 in $\{100, 101, \ldots, 200\}$ **do**
        $(l_1^*, l_2^*) \leftarrow \underset{\ell_{min}, \ell_{max}}{\arg\max} \{\boldsymbol{L}_{k+1}(\ell_{min}, \ell_{max} | \ell_{min} \in (l_1 - 1, l_1 + 1), \ell_{max} \in (l_2 - 1, l_2 + 1)\}$
        $\boldsymbol{L}_k(l_1, l_2) \leftarrow \boldsymbol{L}_k(l_1, l_2) + \boldsymbol{L}_{k+1}(l_1^*, l_2^*)$
        $\boldsymbol{p}_1(k) \leftarrow l_1^*$
        $\boldsymbol{p}_2(k) \leftarrow l_2^*$
  $(\ell_{min}^*, \ell_{max}^*) \leftarrow \underset{\ell_{min}, \ell_{max}}{\arg\max} \{\boldsymbol{L}_1(\ell_{min}, \ell_{max})\}$
  $LL \leftarrow \boldsymbol{L}_1(\ell_{min}^*, \ell_{max}^*)$
  for k in $\{1, 2, \ldots, 70\}$ **do**
    $\boldsymbol{\ell}_1(k) \leftarrow \ell_{min}^*$
    $\boldsymbol{\ell}_2(k) \leftarrow \ell_{max}^*$
    $\ell_{min}^* \leftarrow \boldsymbol{p}_1(k)$
    $\ell_{max}^* \leftarrow \boldsymbol{p}_2(k)$
  return $(LL, \boldsymbol{\ell}_1, \boldsymbol{\ell}_2)$

The Best Path algorithm has been applied to all observations of the training set. The result obtained is illustrated in Figure 11. This figure can be compared to Figure 10. Obviously, adding a constraint to the variation of ℓ_{max} and ℓ_{min} causes the maximum likelihood statistic to change from $-8,462.089$ (unconstrained) to $-8,478.756$. However, this fit statistic is still much better than the one obtained using the Generalized Kappa-N model.

This flexible approach to determining $\ell_{max}(\omega)$ and $\ell_{min}(\omega)$, which depend on ω, shares similarities with non-parametric approaches, where the Best Path algorithm allows us to let the data speak for itself without strictly replicating what has been observed. Indeed, allowing a change of at most $-1/+1$ on ℓ_{max} and on ℓ_{min} allows us to limit the variation, and potentially to limit the overfitting. Other variations of the criteria could be studied, for example $-y_1/+y_2$ on ℓ_{max} and/or ℓ_{min} for different values of y_1 and y_2. Because we know that models that do not limit the variations can probably generate better log-likelihoods, several cross-validation analyses could allow us to see which maximum variations between groups of neighboring ω could be optimal in terms of predictive quality on a series of validation samples. Transition rules for insureds who add or remove items from their insurance policy could also be analyzed.

3.3 Using the Recursive Algorithm for Ψ, Υ and γ_0

Because it limits the variations for similar ω, the algorithm for selecting $\ell_{max}(\omega)$ and $\ell_{min}(\omega)$ yields a flexible pricing structure for defined values of $\Psi(\omega), \Upsilon(\omega)$ and $\gamma_0(\omega)$. However, when we applied the Best Path algorithm to obtain all the limits of the claim score, the values of Ψ, Υ and γ_0, as well as the $\boldsymbol{\beta}$ parameters or any overdispersion parameters

[4]It should be noted that this kind of optimization problem is a textbook dynamic programming applications. We program a small algorithm to find the best path, but simpler and faster solutions certainly exist (see Needleman and Wunsch (1970) who first used this solution for protein sequence analysis)

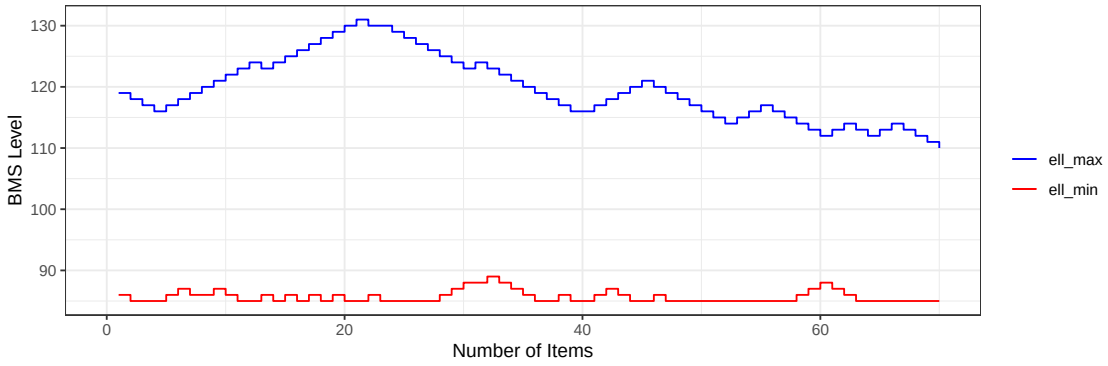

Figure 11: *Values of ℓ_{max} and ℓ_{min} from the best path*

used in the count distribution, were all estimated with the Generalized Kappa-N model defined in equation (7). By adding ℓ_{max} and ℓ_{min} to the model to limit the variation of the claim score for each farm, it becomes clear that many estimates of the underlying count distribution, particularly the values of $\Psi(\omega)$, $\Upsilon(\omega)$ and $\gamma_0(\omega)$, are no longer correct.

The GAM approach, which was described in Section 3.1, was interesting because it allowed the different parameters of any BMS model to remain flexible (see Figure 4). However, the estimation procedure was based on the Generalized Kappa-N model and at no time were the parameters ℓ_{max} and ℓ_{min} involved in the computation of those flexible estimates. That means that we must develop another method to estimate the parameters of a BMS model that takes into account the size of the insured.

Like in the estimation approach developed in Section 2.2.2, we can, however, use the described algorithm to estimate all the parameters $\Psi(\omega)$, $\Upsilon(\omega)$ and $\gamma_0(\omega)$, while using the Best Path algorithm to find $\ell_{max}(\omega)$ and $\ell_{min}(\omega)$. The general idea of the new estimation method is like to one we used to express equation (6): we have to substitute $\kappa_{i,\bullet}$ with $d_{i,\bullet}$, and $n_{i,\bullet}$ by $\tilde{n}_{i,\bullet}$ from the original Generalized Kappa-N model, to develop a Generalized Kappa-N* model; its mean is expressed by:

$$\lambda_{i,T} \quad = \quad \exp(X'_{i,T}\beta - d_{i,\bullet}s_1(\omega_{i,T}) + \tilde{n}_{i,\bullet}s_2(\omega_{i,T})) \tag{10}$$

If we do not have the limits $\ell_{max}(\omega)$ and $\ell_{min}(\omega)$ to compute the generalized claim score, we can use $d_{i,\bullet} = \kappa_{i,\bullet}$ and $\tilde{n}_{i,\bullet} = n_{i,\bullet}$. However, when limits are used, $\tilde{n}_{i,\bullet}$ is not necessarily equal to $n_{i,\bullet}$ for all insureds i, and $d_{i,\bullet}$ can also be different from $\kappa_{i,\bullet}$.

The different steps of the algorithm of ection 3.1 that has to be used to estimate all the parameters of the generalized BMS model are similar to what was described in Section 2.2.2, but modifications are included to allow the use of the Best Path algorithm, which is needed to find $\ell_{max}(\omega)$ and $\ell_{min}(\omega)$. More precisely, the general estimation algorithm is as follow:

- **Initiation**: A Generalized Kappa-N model using $n_{i,\bullet}$ and $\kappa_{i,\bullet}$ is estimated to obtain $\Psi^{(0)}(\omega)$, $\Upsilon^{(0)}(\omega)$ and $\gamma_0^{(0)}(\omega)$. The Best Path algorithm is used to obtain $\ell_{max}^{(0)}(\omega)$ and $\ell_{min}^{(0)}(\omega)$.

  Then, for each step k:

- **Computation**: The values of $d_{i,\bullet}^{(k)}$ and $\tilde{n}_{i,\bullet}^{(k)}$ are computed for each insured i using Equation (4), as are the parameters $\Psi^{(k-1)}(\omega)$, $\Upsilon^{(k-1)}(\omega)$, $\ell_{max}^{(k-1)}(\omega)$ and $\ell_{min}^{(k-1)}(\omega)$.

Model	Rounded	Loglike. (train)	Log. Score (test)
Generalized BMS	No	-8471.739	-2851.839
.	Yes (score 1)	-8474.552	-2852.022

Table 6: *Summary of Generalized BMS models, with and without rounded $\Psi(\omega)$ parameters*

- **Maximization**: The log-likelihood of a Generalized Kappa-N* model, using $\tilde{n}_{i,\bullet}^{(k)}$ and $d_{i,\bullet}^{(k)}$, is maximized to estimate all parameters. Updated values for $\Psi^{(k)}(\omega)$, $\Upsilon^{(k)}(\omega)$ and $\gamma_0^{(k)}(\omega)$ are found and the Best Path algorithm is used to obtain an updated values of $\ell_{max}^{(k)}(\omega)$ and $\ell_{min}^{(k)}(\omega)$.

The initiation step is performed once, but the computation and the maximization steps are repeated until convergence is reached. As opposed to the algorithm described Section 2.2.2, the estimation algorithm does not have to be used for all possible values of the pair (ℓ_{max}, ℓ_{min}). Indeed, because the Best Path algorithm is used in the estimation procedure, optimal values of $\ell_{max}(\omega)$ and $\ell_{min}(\omega)$ are found directly.

3.4 Final BMS model

The estimation procedure described earlier was applied to the farm insurance data. We show and analyze the final results using only generalized score #1, because it is simpler to explain and apply. Indeed, for the BMS score #1, the drop parameter Υ is always one, meaning that all insureds are rewarded by a drop of one level for each year without claim, no matter their size.

The estimation results are shown in Table 6. A model with integer values for $\Psi(\omega)$ is also shown, and the result can be seen in the same table. We can compare the log-likelihood value and logarithmic score of the two Generalized BMS models with what we had obtained for the Kappa-N model and the Generalized Kappa-N model (see Table 6); and for the standard BMS model (see Table 3). It shows that both Generalized BMS models generate the best log-likelihood using the training database, and the best logarithmic score using the test database. Rounding the value of $Psi(\omega)$ slightly decreases the quality of the fit and the quality of the prediction, but a BMS model with an integer jump parameter $\Psi(\omega)$ is much easier to use.

Figure 12 shows the final values of $\ell_{max}(\omega)$ and $\ell_{min}(\omega)$ for the Generalized BMS model. We see a major difference from what was obtained in the first iteration of the Best Path algorithm that was used in the initiation step of the estimation procedure, i.e. for $k = 0$, which was already shown in Figure 11. The values of $\ell_{max}(\omega)$ are fairly stable for all ω; they all fall between 110 and 120. However, significant variations in $\ell_{min}(\omega)$ can be seen. As a new insured enters level $\ell_0 = 100$, $\ell_{min}(\omega)$ can be used to compare the best insureds with several years of experience with the new insureds. We see that the difference is larger for small farms.

Figure 13 compares the parameters $\Psi(\omega)$ and $\gamma_0(\omega)$ of the generalized BMS model with what was found with the Generalized Kappa-N model. We can also see the impact of the rounding of the jump parameter Ψ in the left graph of the same figure. The differences observed between the parameters obtained using the Generalized Kappa-N model and the Generalized BMS model are significant, especially for the largest farms.

The results obtained using the Generalized BMS model cannot be analyzed by looking independently at each of the various estimates of the parameters. Indeed, it is often the combination of several parameters that gives a better understanding of the rating structure of a BMS model.

Figures 14 and 15 probably give a better overview of the rating structure constructed by the Generalized BMS model. The graph on the left of Figure 14 shows the surcharge for a claim, and the claim-free discount, for each value of ω. We

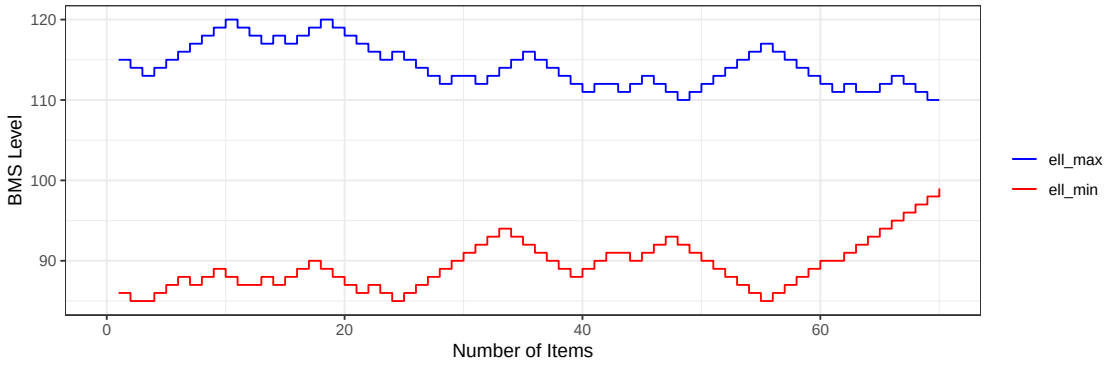

Figure 12: *Values of $\ell_{max}(\omega)$ and $\ell_{min}(\omega)$ for the Generalized BMS model*

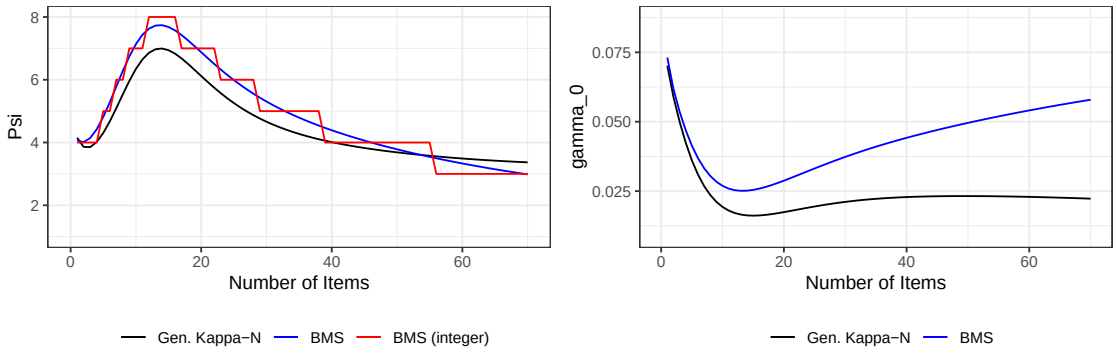

Figure 13: *Functions $\Psi(\omega)$ and $\gamma_0(\omega)$ for the Generalized Kappa-N and the Generalized BMS models*

see that the group of insureds that differs the most is small farms. Indeed, for these farms, the surcharge for a single claim is much higher than for any other sizes of farms. The discount for a claim-free year is also higher for small farms. A similar conclusion can be drawn from the right-hand graph of Figure 14, which shows the maximum surcharges and discounts. Indeed, we see that the suggested rating structure for small farms is to have a much larger surcharge than others, and a higher maximum discount. This seems to suggest a strong heterogeneity between all small farms, compared to medium and large farms. Finally, the results displayed in Figure 15 show the theoretical number of claims needed to reach ℓ_{max}, for an insured at the initial ℓ_0 level. The curve is relatively stable, but increases for farms with more than 50 insured items. In practice, because the number of claims is an integer, the curve would be more stable, varying between 2 and 5 claims.

Previously, when analyzing the curves in Figure 13, it would be tempting to conclude that medium-sized farms, with 10 to 50 insured items, had a different loss experience than other farms. This is not quite the case, as our analysis in Figure 14 shows. Indeed, the high value of Ψ for the medium-sized farms is offset by the decrease in γ_0. In the end, the surcharges and discounts for these farms are not particularly strange. However, the value of Ψ can also be interpreted differently. Indeed, the impact of a single claim on the premium is roughly equal to Ψ years without claims. In other words, if an insured claims, it would take Ψ years without a claim to return to the premium the insured had prior to the claim. Given the $\Psi(\omega)$ function in Figure 14, this could perhaps mean that the rating structure proposed in the Generalized BMS is less forgiving for medium-sized farms than for small and large farms, as it would take more time for a medium-sized farm to return to the premium they had before claiming. This observation likely lead to further modelling work that could be done on other BMS transition rules. Indeed, instead of using a basic BMS model that only rewards insureds when they do not claim in a given year, or penalizes each claim using the same jump parameter Ψ, we can study if other rewards or other penalties should be given. For example, three, five or ten consecutive years

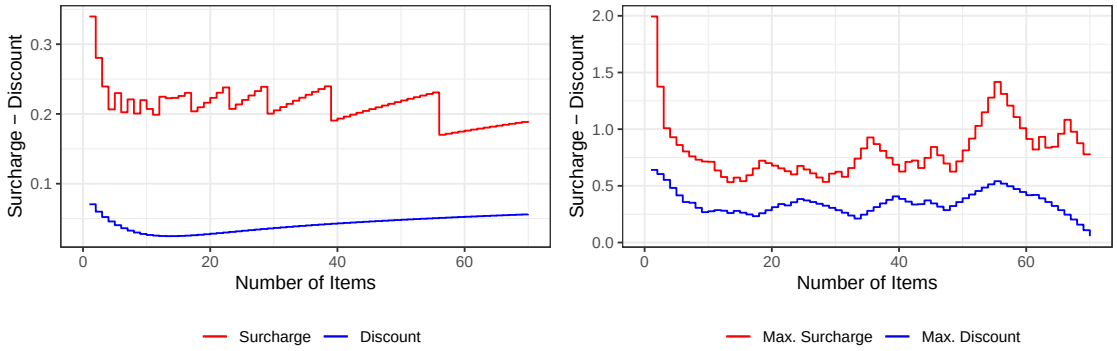

Figure 14: *Maximum surcharges and maximum discounts for the Generalized BMS*

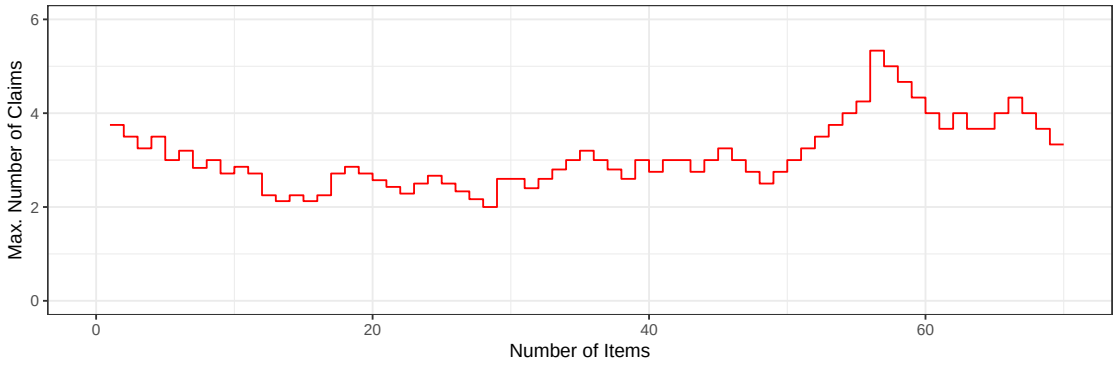

Figure 15: *Number of claims needed to reach ℓ_{max}, for an insured at the initial ℓ_0 level*

without a claim could be rewarded by a greater decrease of levels. Penalties could be different for a second or third claim in a single year. Lemaire (2012) provides an impressive list of bonus-malus systems, with many different possible transition rules. This should be considered in future work.

3.4.1 Practical Considerations

To conclude our paper, and to show the advantages of the BMS approach, we will give a simple example that summarizes the important elements of an experience rating structure that uses the Generalized BMS. Table 7 shows four insureds of different sizes. For each insured, we see all the structural parameters that come from the Generalized BMS. The table also shows some characteristics of the BMS model, such as the surcharge per claim, the claim-free discount, the maximum surcharge and the maximum discount, and finally the number of claims needed to reach ℓ_{max}, for an insured at the initial level ℓ_0. When the BMS model has been adjusted, even though they are easy to compute, all

Insured	ω	Ψ	γ_0	ℓ_{max}	ℓ_{min}	Surcharge by claim	Claim-free Discount	Maximum Surcharge	Maximum Discount	Maximum nb. of claims
1	1	4	0.0731	115	86	33.9%	7.1%	199.4%	64.1%	3.750
2	12	8	0.0253	118	87	22.4%	2.5%	57.8%	28.1%	2.250
3	33	5	0.0396	114	94	21.9%	3.9%	74.1%	21.1%	2.800
4	70	3	0.0579	110	99	19.0%	5.6%	78.4%	5.6%	3.333

Table 7: *Parameters for the Generalized Kappa-N model for both insureds*

of those characteristics of the BMS model deeply explain the way experience rating works, and how it can impact each insured.

For the purpose of practically applying the BMS to rate policyholders, we see that the values of Ψ, ℓ_{max} and ℓ_{min} are readily available and can be given to policyholders, brokers or regulators. With this level of transparency, policyholder # 1, for example, will know that he/she would move up 4 levels for each claim he/she makes, and will know that he/she would move down one level for each year without a claim. Insured #1 also knows that the maximum score is 115, and the lowest score is 86. All of this information can easily be given and explained to all insureds, as Table 7 shows. The relativity parameter γ_0, which also depends on the size of the insured, is a bit more difficult to explain and share with the insureds. Most of the $\boldsymbol{\beta}$ relativities, which are associated with risk characteristics such as the age or the territory of the insured, are also difficult to explain. If the insurer wanted to be completely transparent, a table of relativities arranged by BMS level for each ω, could also be a solution to this problem.

Finally, knowing that it is very possible that an insured with a certain number of insured items ω will add or remove an item from his or her policy, it might be worthwhile to measure the potential impact of such a change on the experience rating structure. The Best Path algorithm, which limits the variation of $\ell_{min}(\omega)$ and $\ell_{min}(\omega)$ for similar ω, combined with the use of a GAM to smooth the values of $\Psi(\omega)$ and $\gamma_0(\omega)$ suggests that the impact will be limited. However, more rigorous analyses would be interesting.

4 Conclusion

Recent BMS models have proven to be both accurate and easy to use. However, BMS models could appear unfair to many insureds and regulators because they do not seem to recognize the initial risk of the insured. Historically, to avoid double penalizing high frequency insureds, many authors proposed the creation of several separate BMS systems for each group of insureds. As shown by Boucher (2022b), this approach cannot be used in all circumstances. Moreover, it has been shown that it creates serious problems when there are insureds that can move from one BMS system to another at any moment.

In this paper, we have thus developed a new BMS model that takes into account the *a priori* risk of the insured. In our numerical application, which uses data from the machinery insurance product for farms, we used the size of the insured as an *a priori* measure of risk, defined as a function of the number of insured items. By using a Generalized Kappa-N model that includes an interaction between the size of the insured and the past claims history, we were able to use the GAM theory to determine smooth functions for all parameters $\Psi(\omega)$, $\Upsilon(\omega)$ and $\gamma_0(\omega)$ so that they depend directly on ω. Then, with the objective of limiting the variations between policyholders of similar size, we develop an algorithm to identify the best values of $\ell_{max}(\omega)$ and $\ell_{min}(\omega)$. Finally, by generalizing the previously introduced estimation algorithm to efficiently estimate the parameters of a BMS model, we explained how a Generalized BMS model could be estimated. The final BMS model produces both the best log-likelihood for the training databaset, and the best logarithmic score for the test database, compared to the other models used in the paper.

We believe that the development of a generalized BMS model could greatly facilitate experience-based pricing, but it can be generalized in several ways:

1. Only the Poisson distribution has been used as a count distribution for all models in this paper. As in Boucher (2022a), other count distributions could easily be implemented with the BMS model. Various forms of negative binomial distributions or any zero-inflated distributions could yield interesting results.
2. As shown in Lemaire (2012), many different transition rules could be used with a BMS.
3. Multiple scope variables and multiple target variables (see Section 2.1 for definitions of those variables) could be tried in order to create a multivariate BMS model, or a BMS model where the BMS levels are calculated

based on the cost of past claims. Using BMS to model the loss cost or to model the severity of a claim could also be studied.

Acknowledgements

The author would like to thank Mathieu J. Dupont for his advice. The author gratefully acknowledges The Co-operators through the Co-operators Chair in Actuarial Risk Analysis.

References

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.

K. C. Ágoston and M. Gyetvai. Joint optimization of transition rules and the premium scale in a bonus-malus system. *ASTIN Bulletin: The Journal of the IAA*, 50(3):743–776, 2020.

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. Bonus-malus scale models: Creating artificial past claims history. *To be published in Annals of Actuarial Science*, 2022a.

J.-P. Boucher. Bonus-malus scales models with insureds of different sizes: Recursive partitioning. *To be published in Risk*, 2022b.

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. *Canadian Institute of Actuaries*, 2019.

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

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. Gyetvai and K. C. Ágoston. Optimization of transition rules in a bonus-malus system. *Electronic Notes in Discrete Mathematics*, 69:5–12, 2018.

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

S. B. Needleman and C. D. Wunsch. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of molecular biology*, 48(3):443–453, 1970.

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, 2019.

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

C. I. Tan, J. Li, J. S.-H. Li, and U. Balasooriya. Optimal relativities and transition rules of a bonus–malus system. *Insurance: Mathematics and Economics*, 61:255–263, 2015.

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.

S. Wood. *Generalized additive models: an introduction with R*. Chapman & Hall/CRC., 2006.