

THE TRUE CLAIM AMOUNT AND FREQUENCY DISTRIBUTIONS WITHIN A BONUS-MALUS SYSTEM

BY

JEAN FRANÇOIS WALHIN** AND JOSÉ PARIS*

* *Université Catholique de Louvain*
• *Secura Belgian Re*

ABSTRACT

We apply Lemaire's algorithm and a non-parametric mixed Poisson fit to a motor insurance portfolio in order to find the true claim frequency and claim amount distributions. The algorithm we develop accounts for the fact that observed distributions are distorted by bonus hunger, when a bonus-malus system is used by the insurer.

KEYWORDS

Mixed Poisson distribution, non-parametric fit, bonus-malus system, hunger for bonus, claim amount distribution, frequency distribution, censoring, maximum likelihood.

1. INTRODUCTION

When an insurer uses a bonus-malus system (BMS) independent of the claim amounts, it will notice a tendency with the insured not to report the smallest claims. Indeed it is in some cases more interesting for the policyholder to bear himself the cost of third party losses than to report the claim and to pay higher premiums in the future because of the malus. Lemaire (1977) called this fact the hunger for bonus. See also Lemaire (1995).

The hunger for bonus induces that the introduction of a (new) BMS creates a censored view of the claim amount and frequency distributions. Indeed some of the lowest claim amounts will not be reported to the insurance companies. Of course, for the policyholder, the natural question is: "up to which level of claim amount is it interesting for me to bear the cost myself?"

Lemaire (1977) answered this question by using an algorithm related to dynamic programming.

In the present paper, we will apply Lemaire's (1977) algorithm and the non-parametric mixed Poisson fit to a motor portfolio (see Walhin and Paris (1999))

in order to redefine the true claim amount and frequency distributions. This problem was already implicitly posed by Lemaire (1977) in his paper where he stated that he had to use old claim amount data because of recent data being influenced by the introduction of the BMS.

Throughout the paper we will use a numerical example. The data associated with this example are

a) observed claim frequency distribution

TABLE 1
OBSERVED FREQUENCY DISTRIBUTION

<i>Number of accidents</i>	<i>Number of policyholders</i>
0	103704
1	14075
2	1766
3	255
4	45
5	6
6	2

This reference portfolio has already been used in Walhin and Paris (1999) where we were looking for parametric and non-parametric mixed Poisson fits.

b) observed claim amount distribution

TABLE 2
OBSERVED CLAIM AMOUNT DISTRIBUTION

6	6	10	11	17	18	20	26	27	34
42	44	47	54	59	60	61	61	61	61
64	64	65	66	67	68	71	71	73	75
76	81	85	87	93	94	101	103	105	109
110	110	113	116	116	129	134	134	141	141
151	154	156	159	167	171	172	173	174	179
181	183	185	187	195	195	203	226	235	240
251	255	273	340						

This is a small hypothetical data set we use for pedagogical purpose. The whole numerical application of the paper can be performed with the data sets given in the introduction.

- c) BMS used by the company (or by the market as is still the case in Belgium). The BMS we use is the one derived in Walhin and Paris (1999):
- $s = 9$: 9 classes numbered 0, 1, ..., 8. 0 is the minimum class. 8 is the maximum class.
 - Entry of the system is in class 4.
 - In case of a claims free year, the policyholder comes down one class.
 - In case of claim(s), the policyholder climbs up 3 classes per claim.
 - The bonuses and maluses (in percentage) are given in the following table:

TABLE 3
PERCENTAGE PREMIUMS

s	0	1	2	3	4	5	6	7	8
C_s	75	80	90	95	100	150	170	185	250

The paper is structured as follows: section 2 briefly describes Lemaire's algorithm. Section 3 recalls the non-parametric mixed Poisson fit while section 4 recalls an efficient way to find the stationary distribution of the policyholders within a BMS. In section 5 we formulate our problem and a solution is given in section 6 in the form of an algorithm. Section 7 is devoted to the numerical solution of the problem with the data sets presented in the introduction. The conclusion is given in section 8.

2. LEMAIRE'S ALGORITHM

Lemaire's (1977) algorithm needs the following hypotheses:

Let a BMS be with s classes: $i = 0, \dots, s - 1$;

the claims frequency of a policyholder be Poisson distributed with mean λ ;

the claim amount distribution be X , with cumulative density function (cdf) $F_X(x)$;

β be the actualisation rate forecast for the future;

P be the total premium, i.e. the base premium at level 100%, including security loading, administration expenses, brokerage and taxes;

$1 - t$ with $0 \leq t < 1$ be the time remaining until the next premium payment;

m be the number of claims reported to the Company in $[0, t)$.

With these hypotheses an iterative algorithm can be performed in order to find the optimal policy of the driver as a function of his bonus-malus level. The optimal policy is simply the optimal retention of the driver as a function of his bonus-malus level. It is the level of claim amount up to which it is interesting for the policyholder to bear the cost himself and not to report the claim to the company. Of course, the optimal policy is also a function of t , the time at which the claim occurs and m , the number of claims reported before t unless one assumes $t = 0$. Optimal frequencies of the driver are also given by the algorithm.

The solution of the algorithm is shown to be unique if $\beta < 1$, which is always the case if the interest rate is positive.

In short, the algorithm of Lemaire gives the optimal frequency and the optimal retention of a driver based on the true claim amount and frequency distributions of the driver.

3. NON-PARAMETRIC MIXED POISSON FIT

The mixed Poisson distribution is often used to model the number of claims in a motor portfolio.

Let $\Pi(k, t)$ be the probability that a risk causes k accidents in t years. We have

$$\Pi(k, t) = \int_0^\infty e^{-\lambda t} \frac{(\lambda t)^k}{k!} dU(\lambda), \quad k \geq 0$$

Classical distributions are the Negative Binomial and the Poisson Inverse Gaussian.

In Walhin and Paris (1999), a non-parametric fit for $\Pi(k, t)$ is discussed. We recall that the maximum likelihood of the non-parametric fit is attained by a mixture of some Poisson distributions, depending on the form of the portfolio.

We have

$$\Pi(k, t) = \sum_{j=1}^r p_j e^{-\lambda_j} \frac{\lambda_j^k}{k!}, \quad k \geq 0$$

$$0 \leq \lambda_1 < \lambda_2 < \dots < \lambda_r$$

If we define u as the maximum number of claims per risk and v as the number of classes for which the observation is different from 0 then the maximum likelihood is unique under the following conditions:

$$r \leq \min\left(v, \left\lfloor \frac{u+2}{2} \right\rfloor\right) \quad \text{if } \lambda_1 = 0$$

$$r \leq \min\left(v, \left\lfloor \frac{u+1}{2} \right\rfloor\right) \quad \text{if } \lambda_1 > 0$$

For the example described in the introduction, we find

TABLE 4
MAXIMUM LIKELIHOOD ESTIMATION
FOR THE REFERENCE PORTFOLIO

$\lambda_1 = 0.05461$	$p_1 = 0.56189$
$\lambda_2 = 0.24599$	$p_2 = 0.41463$
$\lambda_3 = 0.95618$	$p_3 = 0.02348$

This fit gives an interesting interpretation of the portfolio: there are 3 types of risks $\lambda_j, j = 1, 2, 3$ with probabilities $p_j, j = 1, 2, 3$.

4. THE STATIONARY DISTRIBUTION OF THE POLICYHOLDERS WITHIN A BMS

In order to perform our calculations we will assume that the BMS has existed for a long time and that it has reached its stationary distribution (see Lemaire (1995), for instance).

Let $p(x)$ be the probability that a driver with average claim frequency λ causes x claims during a given year. The transition probability matrix (Q) of this driver within the BMS described in the introduction is thus

TABLE 5
TRANSITION PROBABILITY MATRIX

s	0	1	2	3	4	5	6	7	8
0	$p(0)$	0	0	$p(1)$	0	0	$p(2)$	0	$1 - p(0) - p(1) - p(2)$
1	$p(0)$	0	0	0	$p(1)$	0	0	$p(2)$	$1 - p(0) - p(1) - p(2)$
2	0	$p(0)$	0	0	0	$p(1)$	0	0	$1 - p(0) - p(1)$
3	0	0	$p(0)$	0	0	0	$p(1)$	0	$1 - p(0) - p(1)$
4	0	0	0	$p(0)$	0	0	0	$p(1)$	$1 - p(0) - p(1)$
5	0	0	0	0	$p(0)$	0	0	0	$1 - p(0)$
6	0	0	0	0	0	$p(0)$	0	0	$1 - p(0)$
7	0	0	0	0	0	0	$p(0)$	0	$1 - p(0)$
8	0	0	0	0	0	0	0	$p(0)$	$1 - p(0)$

As is the case for each BMS, we have an irreducible (there are no cycles) Markov chain where all states are ergodic (each state can be attained from another state). Under those conditions, there is a stationary probability distribution that is given by:

$$e_{\infty}(\lambda) = \lim_{n \rightarrow \infty} Q^n e_0(\lambda)$$

where $e_0(\lambda)$ denotes any initial distribution of the drivers in the BMS. The stationary probability distribution is independent of $e_0(\lambda)$.

The stationary probability distribution is also given by solving

$$e_{\infty}(\lambda) = e_{\infty}(\lambda)Q$$

with the normalizing condition

$$\sum_{i=0}^{s-1} e_{\infty}(i; \lambda) = 1$$

where $e_{\infty}(i; \lambda)$ is the i^{th} component of the vector $e_{\infty}(\lambda)$.

If one is interested in the stationary distribution of the portfolio (e_{∞}), we only have to take the weighted average of the stationary distributions for the

different types of policyholders (see Walhin and Paris (1999) for details). With our non-parametric mixed Poisson fit, we have

$$\mathbf{e}_\infty = \sum_{i=1}^r p_i \mathbf{e}_\infty(\lambda_i)$$

For our numerical example we find

TABLE 6
STATIONARY DISTRIBUTION OF THE DRIVERS

	$\lambda = 0, 05461$	$\lambda = 0, 24600$	$\lambda = 0, 95619$	<i>Portfolio</i>
0	0.8278	0.2598	0.0005	0.5728
1	0.0464	0.0724	0.0008	0.0561
2	0.0490	0.0926	0.0022	0.0660
3	0.0518	0.1185	0.0057	0.0783
4	0.0095	0.0876	0.0145	0.0420
5	0.0075	0.0942	0.0369	0.0441
6	0.0052	0.0977	0.0939	0.0457
7	0.0014	0.0880	0.2386	0.0429
8	0.0009	0.0888	0.6066	0.0516

Note that if the stationary distribution has not yet been reached, it is not a problem to work with the transient probabilities. The distribution of the drivers within the BMS after T years is given by

$$e_T(\lambda) = Q^T e_0(\lambda)$$

5. FORMULATION OF THE PROBLEM

When collecting data on a market where a BMS is in use, we do not observe the true claim amount and frequency distributions. Indeed they are influenced by the hunger for bonus.

The true claim amount distribution should have a lower mean whereas the true claims frequency distribution should have a higher mean.

Let us assume that there is a proportion p of the driving population that reports all accidents whereas $(1-p)\%$ only reports the claims exceeding the optimal retention given by Lemaire's algorithm.

Let us assume that a non-parametric fit for the claim frequency distribution had been performed on the observed portfolio (i.e. the reported claims). It reveals r types of risks λ_j each with probability p_j ($j = 1, \dots, r$). This distribution (N) is not the distribution of the number of accidents but the distribution of the number of accidents reported to the Company.

The density function (df) of N writes

$$\Pi(k, 1) = \sum_{j=1}^r p_j e^{-\lambda_j} \frac{\lambda_j^k}{k!}, \quad k = 0, 1, 2, \dots$$

Let N' be the true distribution of the number of claims. Its df writes:

$$\Pi'(k, 1) = \sum_{j=1}^r p'_j e^{-\lambda'_j} \frac{(\lambda'_j)^k}{k!}, \quad k = 0, 1, 2, \dots$$

We will assume that $p'_j = p_j \quad \forall j$ i.e. that the proportions of different risks for both distributions are the same.

Let X be the random variable representing the reported claim amounts.

Let Z be the random variable representing the true claim amounts. This random variable is unknown whereas X is the observed one.

The df of X is a function of the df of Z and writes

$$f_X(x) = p f_Z(x) + (1 - p) \frac{f_Z(x)}{1 - F_Z(c)} \mathbb{1}_{\{x \geq c\}}, \quad x \geq 0$$

where c is the average retention limit of the portfolio.

Our aim is to find the distribution of Z and N' .

6. AN ALGORITHM TO SOLVE THE PROBLEM

A solution to the problem described in section 5 will be given by means of an iterative algorithm using the non-parametric fit of the portfolio and an inversion of the algorithm of Lemaire (1977).

Step 0: Initializing step

Have an initial guess for the parameter c . Choose a parametric distribution to fit the random variable Z .

Step 1: Correction of the claim amount distribution

We use the average optimal retention (c) as a censor in order to find a new estimate for the vector of parameters of the distribution of Z . Therefore we maximize the likelihood:

$$(1) \quad L(\theta, p|c) = \prod_{i=1}^n \left[p f_Z(x_i; \theta) + (1 - p) \frac{f_Z(x_i; \theta)}{1 - F_Z(c; \theta)} \mathbb{1}_{\{x_i \geq c\}} \right]$$

With the new estimate $(\hat{\theta}, \hat{p}) = \operatorname{argmax} L(\theta, p|c)$ we move to step 2.

Step 2: Correction of the frequency distribution

In view of the non-parametric fit, there are r types of policyholders. For each of them we repeat the following:

Let λ_j $j = 1, \dots, r$ be the observed frequency (found by the non-parametric fit of the reference portfolio).

Let λ'_j $j = 1, \dots, r$ be the true frequency.

Let λ''_j $j = 1, \dots, r$ be the optimal frequency given by Lemaire's algorithm.

We then have:

$$(2) \quad \lambda_j = p\lambda'_j + (1 - p)\lambda''_j$$

λ_j is our observation. λ'_j is in fact the entry of the algorithm of Lemaire (1977). λ''_j is a by-product of this algorithm.

We apply a trial-error scheme on the entry λ'_j in order to match the observation λ_j in connection with equation (2).

For this λ'_j , the algorithm of Lemaire (1977) gives the optimal policy c_j .

The average optimal retention is given by

$$c = \sum_{j=1}^r c_j p_j$$

where p_j is the weight associated to λ_j .

With this new average optimal retention c we go back to step 1.

Stopping rule:

We stop the process when convergence occurs. We cannot prove this convergence but in practice it is bound to happen.

Let us note that the following intuitive result is easily shown by maximum likelihood:

$$\hat{p} = \frac{F_n(c)}{F_Z(c)}$$

Proof

Let $x_{(1)}, x_{(2)}, \dots, x_{(k)}, c, x_{(k+1)}, \dots, x_{(n)}$ be the order statistic of our observation.

The loglikelihood of (1) writes

$$\prod_{i=1}^k p f_Z(x_{(i)}) \prod_{i=k+1}^n \left[p f_Z(x_{(i)}) + (1 - p) \frac{f_Z(x_{(i)})}{1 - F_Z(c)} \right]$$

The normal equation for p gives

$$\hat{p} = \frac{\frac{k}{n}}{F_Z(c)}$$

□

Thus clearly the estimate of p , the proportion of insured reporting all accidents will depend on the parametric distribution chosen for Z .

In short the algorithm writes

Step 0: initialization

Do

Step 1: correction of the claim amount distribution

Maximize (1)

Step 2: correction of the frequency distribution

For $j = 1$ to r

Do

Try a value for λ'_j and apply Lemaire's algorithm

Until (2) is verified

Next j

Find the average optimal retention

Until convergence

7. NUMERICAL EXAMPLE

For the numerical example, we use the observed data set of section 1 as well as the BMS described in section 1.

In order to use Lemaire's algorithm, we set up the following hypotheses:

1. $\beta = \frac{1}{1 + 0.06}$
2. $t = \frac{1}{2}$
3. $P = 35$

The total premium may seem very high but in fact it is not. Indeed it includes tax, brokerage, administration expenses and a fluctuation loading.

The initializing step of the algorithm is chosen as

- $c = 30$
- Z is exponentially distributed with mean μ :

$$f_Z(x; \mu) = \frac{1}{\mu} e^{-\frac{x}{\mu}}, \quad x \geq 0$$

We will now describe in detail the first iteration of the algorithm.

Step 1: by maximizing (1) we get

$$\begin{aligned} \hat{\mu} &= 97.137 \\ p &= 0.4577 \end{aligned}$$

Step 2: for $j = 1, 2, 3$ we have to match equation (2). We describe in detail the trial-error scheme for $j = 1$.

Let us try a true claim frequency $\lambda'_1 = 0.075$.

The application of Lemaire's algorithm gives the following optimal retentions and frequencies:

TABLE 7
OPTIMAL RETENTIONS

s	θ	1	2	3	4	5	6	7	8
$m = 0$	18	28	52	76	102	145	119	87	53
$m = 1$	83	108	128	91	54	0	0	0	0
$m = 2$	89	53	0	0	0	0	0	0	0

TABLE 8
OPTIMAL FREQUENCIES

s	θ	1	2	3	4	5	6	7	8
$m = 0$	0.062	0.056	0.043	0.034	0.026	0.016	0.022	0.030	0.043
$m = 1$	0.031	0.024	0.020	0.029	0.043	0.075	0.075	0.075	0.075
$m = 2$	0.029	0.043	0.075	0.075	0.075	0.075	0.075	0.075	0.075

We assume that the stationary distribution is attained within our BMS. This stationary distribution was obtained in column 1 of table 6. The parameter of the Poisson distribution in the transition probability matrix is of course 0.05461 because the drivers move in the BMS according to the frequency of reported claims.

Average values for the optimal frequencies and retentions are then easily given by the scalar product between the stationary probability vector and the optimal frequency or retention vector. We find

TABLE 9
AVERAGE RETENTION LIMITS AND FREQUENCIES IN FUNCTION OF m

	<i>Retention limit</i>	<i>Frequency</i>
$m = 0$	25	0.0587
$m = 1$	85	0.0310
$m = 2$	76	0.0362
$m \geq 3$	0	0.075

We now want figures that are independent of m . Therefore we look for an average value of the optimal retention and frequency by applying the formulae:

$$c_j = \sum_{m=0}^{\infty} e^{-\lambda_j t} \frac{(\lambda_j t)^m}{m!} c_j(m) \quad j = 1, \dots, r$$

$$\lambda_j'' = \sum_{m=0}^{\infty} e^{-\lambda_j t} \frac{(\lambda_j t)^m}{m!} \lambda_j''(m) \quad j = 1, \dots, r$$

TABLE 10
AVERAGE RETENTION LIMIT AND FREQUENCY

<i>Retention limit: c_1</i>	<i>Frequency: λ_1''</i>
27	0.0587

Equation (2) writes:

$$0.05461 \neq (0.4577)0.075 + (1 - 0.4577)0.0587 = 0.0661$$

We then proceed by trial-error until equation (2) is matched. This happens with $\lambda_1' = 0.062$.

For $j = 2$ (resp. $j = 3$) we find $\lambda_2' = 0.3392$ (resp. $\lambda_3' = 1.0745$).

The second and subsequent iterations may now be completed. We find

TABLE 11
ITERATIONS UNTIL CONVERGENCE

<i>Iteration</i>	<i>c</i>	<i>μ</i>	<i>p</i>	<i>λ_1'</i>	<i>λ_2'</i>	<i>λ_3'</i>
1	30	97.137	0.4577	0.062	0.3392	1.0745
2	47.7386	85.2208	0.4096	0.0637	0.3628	1.1112
3	47.6437	85.3233	0.4105	0.0637	0.3624	1.1107
4	47.6434	85.3239	0.4105	0.0637	0.3624	1.1107

As we see, convergence occurs. The true claim amount distribution is then exponentially distributed with mean $\mu = 85.32$. The model shows that 41% of the policyholders report all the claims while 59% use the optimal retention. The true claim frequency distribution is non-parametric mixed Poisson distributed with

TABLE 12
PARAMETERS OF THE TRUE CLAIM FREQUENCY DISTRIBUTION

$\lambda'_1 = 0.0637$	$p_1 = 0.56189$
$\lambda'_2 = 0.3628$	$p_2 = 0.41463$
$\lambda'_3 = 1.1107$	$p_3 = 0.02348$

With these true distributions, the pure premium should have been

$$\mathbb{E}N \times \mathbb{E}X = 0.2122 \times 85.32 = 18.11$$

although it was

$$\mathbb{E}N \times \mathbb{E}X = 0.155 \times 113.40 = 17.58$$

with the observed distributions. As expected, the pure premium is higher with the true distributions because in the case of the observed distribution, some claims are withheld by the policyholders which makes the aggregate claim amount distribution less important.

The frequencies now compare as

TABLE 13
COMPARISON OF THE FREQUENCIES

	<i>True frequency</i>	<i>Frequency with hm</i>	<i>Increase</i>
λ_1	0.0637	0.0546	17%
λ_2	0.3622	0.2459	47%
λ_3	1.1107	0.9561	16%

This is not surprising as the bad drivers remain in the higher classes of the BMS and are less interested by the hunger for bonus because of the maximal penalty.

8. CONCLUSION

Changing the BMS is a task that may affect every insurance company. In particular, Belgian companies will be obliged to use new BMS due to the European directive that forbids the use of an unique BMS for all the drivers of all the companies.

In this case it is necessary to have an idea of the true claim amount and frequency distributions because we know that they are influenced by the hunger for bonus.

This paper gives a solution by using non-parametric mixed Poisson fits and an inversion of Lemaire's algorithm related to the hunger for bonus. The proportion of policyholders using the optimal policy of Lemaire's algorithm is also derived.

ACKNOWLEDGMENTS AND COMMENT

We would like to thank the editor and the anonymous referees for their very constructive comments which led to a better presentation of the paper. A first version of this paper were presented at the International Congress on Industrial and Applied Mathematics 99 in Edinburgh, July 99.

REFERENCES

- LEMAIRE, J. (1977) La Soif du Bonus. *ASTIN Bulletin* **9**, 181-190.
LEMAIRE, J. (1995) *Bonus-Malus Systems in Automobile Insurance*. Kluwer, Boston.
WALHIN, J.F. and PARIS, J. (1999) Using Mixed Poisson Distributions in Connection with Bonus-Malus Systems. *ASTIN Bulletin* **29**, 81-99.

JEAN-FRANÇOIS WALHIN AND JOSÉ PARIS

Institut de Statistique
Voie du Roman Pays, 20
B-1348 Louvain-la-Neuve
Belgium
E-mail: walhin@stat.ucl.ac.be

JEAN-FRANÇOIS WALHIN

Secura Belgian Re
Rue Montoyer, 12
B-1000 Bruxelles
Belgium
E-mail: jfw@secura-re.com

