HOW TO DEFINE A BONUS-MALUS SYSTEM WITH AN EXPONENTIAL UTILITY FUNCTION *

Jean Lemaire

We compute a merit-rating system for automobile third party liability insurance by two different ways, both with the help of an exponential utility function.

(i) We apply the principle of zero utility to exponential utilities.

(ii) We break the symmetry between the overcharges and the undercharges by weighting them differently through the introduction of a utility function, in order to penalize the overcharges.

The results are applied to the portfolio of a Belgian company and compared to the premium system provided by the expected value principle.

Deux méthodes différentes, basées sur l'emploi de fonctions d'utilité exponentielles nous permettent de définir un système bonus-malus en assurance automobile:

(i) le principe de l'utilité nulle;

(ii) la pénalisation des injustices de la compagnie, obtenue en pondérant les erreurs de prime au moyen d'une fonction d'utilité de manière à briser la symétrie entre les primes trop élevées et les primes trop basses.

Les résultats théoriques sont appliqués au portefeuille d'une compagnie belge et comparés aux primes fournies par le principe de l'espérance mathématique.

1. THE EXPECTED VALUE PRINCIPLE

In automobile third party liability rate-making, the policy-holders are usually differentiated by two kinds of discriminating variables: a set of a priori factors (like power and use of the car, age and sex of the driver, territory, ...) and an a posteriori classification scheme or merit-rating system.

Let us consider a given risk class, with all the a priori factors fixed and suppose that the introduction of more variables is either practically impossible or would not improve the homogeneity of the risk class. The problem is then to define an "optimal" bonus-malus system, where the meaning of the word "optimal" is to be specified.

A first possibility is to charge each policy an amount proportional to its expected number of claims: this is the expected value principle, developed by Bühlmann (1964) and Bichsel (1964) among others. For our numerical examples, we shall classically assume that the number of claims of each policy is Poisson distributed

$$p_k(\lambda) = \frac{e^{-\lambda}\lambda^k}{k!} \qquad (\lambda > 0),$$

with a Γ-structure function

$$dU(\lambda) = \frac{\tau^a e^{-\tau\lambda}\lambda^{a-1}}{\Gamma(a)} \, d\lambda \qquad (a, \tau > 0).$$

---

* An earlier version of this paper was presented at the 14th ASTIN Colloquium, Taormina, October 1978.

The Γ-distribution has a mean $m = \dfrac{a}{\tau}$, a variance $\sigma^2 = \dfrac{a}{\tau^2}$ and a moment-generating function $M(t) = \left(1 - \dfrac{t}{\tau}\right)^{-a}$ $(t < \tau)$.

It is well-known that in this case—see for instance DERRON (1962)—the distribution of the number of claims in the portfolio is a negative binomial

$$p_k = \binom{k+a-1}{k} \left(\frac{\tau}{1+\tau}\right)^a \left(\frac{1}{1+\tau}\right)^k.$$

Applied to the following observed distribution—see LEMAIRE (1977)—this model provides a fairly good fit, accepted by the χ^2-test of goodness of fit

TABLE 1

Number of claims	Absolute frequencies	
	Observed	Adjusted
0	96,978	96,895.5
1	9,240	9,222.5
2	704	711.7
3	43	50.7
4	9	3.6
More than 4	0	0
Total	106,974	106,974

Mean = .1011, Variance = .1074, a = 1.6049, τ = 15.8778.

Suppose the risk class has been observed for t years, and let k_l $(l = 1, \ldots, t)$ be the number of claims declared during year l. Those k_l's are realizations of random variables K_l, assumed to be independent and equidistributed. To each set of observations $(k_1, \ldots, k_t)$, we must associate a premium $P_{t+1} = P_{t+1}(k_1, \ldots, k_t)$.

If $P(k_1, \ldots, k_t \mid \lambda)$ denotes the probability that a policy-holder with given parameter λ will produce a sequence $(k_1, \ldots, k_t)$, the a posteriori distribution of λ is

$$dU(\lambda \mid k_1, \ldots, k_t) = \frac{P(k_1, \ldots, k_t \mid \lambda)dU(\lambda)}{\int_0^\infty P(k_1, \ldots, k_t \mid \lambda)dU(\lambda)}.$$

The expected value principle defines the premium $P_{t+1}(k_1, \ldots, k_t)$ by

$$P_{t+1} = K(1 + \alpha) \int_0^\infty \lambda dU(\lambda \mid k_1, \ldots, k_t),$$

where K is a constant and α the safety loading.

It is easier to define a bonus-malus system by the relativities

$$100 \cdot \frac{\int_0^\infty \lambda \, dU(\lambda \mid k_1, \ldots, k_t)}{\int_0^\infty \lambda \, dU(\lambda)}$$

i.e. the premium the policy-holder has to pay if its initial premium ($t = 0$) is 100.

The negative binomial model has the interesting property that the a posteriori distribution of the claim frequency λ also admits a Γ-distribution

$$dU(\lambda \mid k_1, \ldots, k_t) = \frac{\tau'^{a'} \lambda^{a'-1} e^{-\tau'\lambda}}{\Gamma(a')} \, d\lambda,$$

with parameters $a' = a + k$ and $\tau' = \tau + t$, where $k = \sum_{l=1}^{t} k_l$ is the total number of claims.

The relativities are in this case

$$100 \cdot \frac{a + k}{\tau + t} \cdot \frac{\tau}{a}.$$

Applied to our example, they provide the following merit-rating system.

TABLE 2

				k			
t	0	1	2	3	4	5	6
0	100						
1	94.07	152.69	211.30	269.92			
2	88.81	144.15	199.48	254.82	310.16	365.51	
3	84.10	136.51	188.92	241.32	293.73	346.14	398.55
4	79.87	129.64	179.41	229.18	278.95	328.73	378.50

In the following section we shall use two different approaches to determine a bonus-malus system: the principle of zero utility and the penalization of overcharges. They have one thing in common: the use of utility functions. As GERBER (1974a, 1974b) has shown that exponential utility functions, of the form

$$u(x) = \frac{1}{c} \left(1 - e^{-cx} \right) \qquad (c > 0)$$

possess very desirable properties for the insurers, we shall only use this particular form in the sequel.

2. THE PRINCIPLE OF ZERO UTILITY

This is Gerber's work, and all we have to do is to apply formula (19) in GERBER (1974a) to a degenerate distribution (since a merit-rating system is based solely on the number of claims and not on their amount). We obtain

$$P_{t+1}(k_1, \ldots, k_t) = K \cdot \frac{a+k}{c} \left| \text{Log} \left(1 - \frac{e^c - 1}{t + \tau}\right) \right|$$

$$[c < \text{Log} (\tau + 1)]$$

This credibility premium is a non-decreasing continuous function of c. A choice a $c = .4$ yields a reasonable initial premium $P_1 = .1262$; since the pure premium is .1011, it corresponds to a safety loading of about 25%.

The table of relativities differ deceptively little from the preceding one.

TABLE 3

t	0	1	2	k 3	4	5	6
0	100						
1	93.99	152.55	211.11	269.67			
2	86.66	143.90	199.14	254.38	309.62	364.86	
3	83.90	136.17	188.45	240.72	293.00	345.28	397.55
4	79.62	129.23	178.85	228.50	278.07	327.68	377.30

The differences are small, even for very unreasonable values of c. The following table was computed with $c = 1.65$, a value that trebles the initial premium.

TABLE 4

t	0	1	2	k 3	4	5	6
0	100						
1	93.13	151.17	209.20	267.23			
2	87.16	141.46	295.77	250.08	304.38		
3	81.90	132.94	183.97	235.01	286.04	337.07	388.11
4	77.25	125.39	173.52	221.66	269.79	317.92	366.06

One notices that in both cases all the premiums are below the corresponding figures of table 2. This naturally implies a higher initial premium P_1.

3. PENALIZATION OF OVERCHARGES

In this section we develop an idea of FERREIRA (1977).

If we represent the a posteriori distributions of the claim frequencies [like in fig. 1 for $t = 3$, $k = 0$ (group 1) and $t = 3$, $k = 2$ (group 2)], we observe that those distributions substantially overlap. All the policy-holders of the second group must pay a premium 2.24 times higher than the members of group 1, and yet, many of them have an actual claim frequency (see shaded area in fig. 1) that place them below the average of group 1. Those people are thus strongly overcharged: they pay more than twice their fair premium. The problem is that, since no distinction amoung group 2 is available, we do not know which of the group 2 policies are those with the lowest claim frequencies. The problem increases with the value of k since the injustices due to overcharges grow in amount and do not become so scarce in quantity since the variance of $dU(\lambda \mid k_1, \ldots, k_t)$ grows (linearly in the negative binomial case) with the number of accidents, for given t.
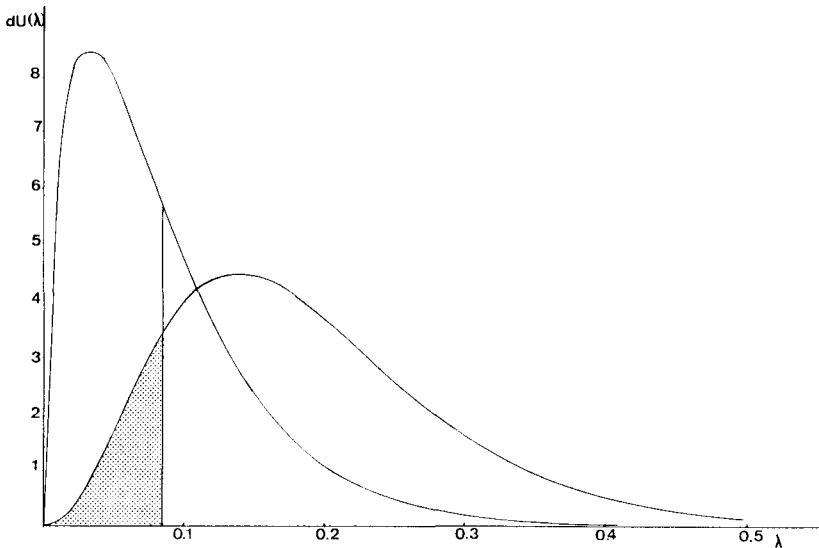

Fig. 1

The rates obtained by the expected value principle possess interesting properties, (see BÜHLMANN (1964)), they minimize the sum of the squared 'errors' (overcharges and undercharges) taken over all the policies of a given group, they ensure financial stability in the sense that the premiums will compensate for the expected losses for every value of t. However, it might seem unfair, from the policy-holder's point of view, that they treat overcharges and undercharges symmetrically: 'paying too much' and 'not paying enough'

are valued the same way; one might argue that the error which consists of asking too much to a policy-holder is more severe than the opposite one. Some sense of equity commands us to distinguish them; one should weight differently both kinds of errors, penalizing the overcharges.

Since all the members of group 2 must pay the same amount, this practically means that this premium must be reduced. Consequently the premium for the first group must be raised in order to restore the financial balance. However, as the highest risk classes are nearly always the least populated, the increase of the premium of group 1 will usually be quite small.

The fact that the population of the sub-groups decreases with k can be illustrated by the results of a simulation performed on the portfolio described in section 1. CORLIER, LEMAIRE and MUHOKOLO (1979) obtained the following populations.

TABLE 5

				k			
t	0	1	2	3	4	5	6
0	10,000						
1	9,059	877	58	6	0	0	0
2	8,297	1472	197	31	2	1	0
3	7,584	1947	381	73	12	2	1
4	6,991	2238	600	130	29	8	4

Consequently it is only necessary to raise the premium of group 1 by 1 B.F. in order to diminish the contributions of group 2 by 20 B.F.

One way to treat the problem asymmetrically is to index one's preferences over all overcharging and undercharging possibilities by a utility function, and to maximize the weighted expected utility, naturally with the condition that the system will be financially stable.

For a given value of t, let us denote by
- $m + 1$ the number of groups (m is the maximum value of k),
- N_k the population of those groups,
- $N = \sum_{k=0}^{m} N_k,$
- $p_k = P_{t+1}(k_1, \ldots, k_t),$
- $dU(\lambda \mid \bar{k}) = dU(\lambda \mid k_1, \ldots, k_t),$ and
- $\bar{\lambda} = \int_0^\infty \lambda dU(\lambda).$

Using exponential utility functions, with arguments equal to the difference between the premium p_k and the real value λ, we shall maximize

$$Z = \frac{1}{N} \sum_{k=0}^{m} N_k \int_0^{\infty} \frac{1}{c} [1 - e^{-c(\lambda - p_k)}] \, dU(\lambda \mid \bar{k})$$

subject to the condition

$$\frac{1}{N} \sum_{k=0}^{m} N_k p_k = \bar{\lambda},$$

or minimize the Lagrangian function

$$\psi = \frac{1}{c} \frac{1}{N} \sum_{k=0}^{m} N_k \int_0^{\infty} e^{-c(\lambda - p_k)} \, dU(\lambda \mid \bar{k}) - \alpha \left(\frac{1}{N} \sum_{k=0}^{m} N_k p_k - \bar{\lambda} \right).$$

(1) $$\frac{\partial \psi}{\partial \alpha} = 0 \rightarrow \frac{1}{N} \sum_{k=0}^{m} N_k p_k = \bar{\lambda},$$

(2) $$\frac{\partial \psi}{\partial p_k} = 0 \rightarrow \frac{1}{N} N_k \int_0^{\infty} e^{cp_k} e^{-c\lambda} \, dU(\lambda \mid \bar{k}) = \frac{\alpha}{N} N_k \qquad k = 0, \ldots, m.$$

If we denote by $M_k(x) = \int_0^{\infty} e^{x\lambda} \, dU(\lambda \mid \bar{k})$ the momentgenerating function of the a posteriori distribution of λ, equations (2) simplify to

$$e^{cp_k} M_k(-c) = \alpha \qquad k = 0, \ldots m$$

or

(3) $$p_k = \frac{1}{c} \log \alpha - \frac{1}{c} \log M_k(-c) \qquad k = 0, \ldots m.$$

Let $\beta = \frac{1}{c} \mathrm{Log}\, \alpha$. β can be obtained by multiplying (3) by N_k, summing up over all the values of k and dividing by N. We get

$$\frac{1}{N} \sum_{k=0}^{m} N_k p_k = \frac{1}{N} \sum_{k=0}^{m} N_k \beta - \frac{1}{N} \frac{1}{c} \sum_{k=0}^{m} N_k \mathrm{Log}\, M_k(-c)$$

and, by (1)

$$\beta = \bar{\lambda} + \frac{1}{Nc} \sum_{k=0}^{m} N_k \mathrm{Log}\, M_k(-c).$$

Finally

$$p_k = P_{t+1}(k_1, \ldots, k_t) = \bar{\lambda} + \frac{1}{c} \left[\frac{1}{N} \sum_{i=0}^{m} N_i \text{ Log } M_i(-c) - \text{ Log } M_k(-c) \right].$$

The value of c can be chosen in order to reflect one's preferences over the asymmetry of the errors. If we set $c = 11.5$, it means that it is necessary to undercharge two policies by .03 each in order to compensate for a single overcharge of .04. A choice of $c = 17.5$ summarizes a sense of fairness that requires two undercharges of .04 to balance one overcharge of .04.

Using the Γ-structure function and the populations of table 5, we obtain the following relativities for $c = 11.5$

TABLE 6

				k			
t	0	1	2	3	4	5	6
0	100						
1	95.48	140.17	184.62	229.55			
2	91.58	134.28	177.02	219.74	262.45	305.17	
3	87.73	128.68	169.63	210.48	251.43	292.38	333.23
4	84.26	123.52	162.79	202.05	241.32	280.58	319.84

and for $c = 17.5$.

TABLE 7

				k			
t	0	1	2	3	4	5	6
0	100						
1	95.93	136.14	176.36	216.56			
2	92.39	130.97	169.54	208.13	246.69	285.28	
3	88.91	125.98	163.06	200.14	237.21	274.29	311.36
4	85.69	121.39	157.08	192.77	228.46	264.15	299.86

4. CONCLUSION

Comparing the different tables, we notice that the merit-rating systems defined by tables 2 to 4 require very high surcharges for the bad risks. Although perfectly justified, it seems very difficult to enforce them practically, for political and commercial reasons. In fact, no system in operation in the world present such severe maluses. A consequence is that in many countries the bonus-malus systems are out of balance financially, in the sense that the maluses are too small to compensate for the bonuses.

On the other hand, the systems constructed by the technique developed in 3 are less severe (and the higher the aversion to injustice c, the smaller the ratio between the extreme premiums). If it is impossible for practical reasons to charge the necessary maluses, one could think of applying this technique, since it produces smaller maluses and yet preserves financial stability.

Université Libre de Bruxelles

REFERENCES

BICHSEL, F. (1964). Erfahrungs-Tarifierung in der Motorfahrzeughaftpflicht-Versicherung; *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker*, **64**, 119-129.

BÜHLMANN, H. (1964). Optimale Prämienstufensysteme; *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker*, **64**, 193-213.

CORLIER, F., LEMAIRE, J. and D. MUHOKOLO. (1979). Simulation of an automobile portfolio, The Geneva papers on risk and insurance, **12**, 40-46.

DERRON, M. (1962). Mathematische Probleme der Automobilversicherung; *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker*, **62**, 103-123.

FERREIRA, J. (1977). Identifying Equitable Insurance Premiums for Risk Classes: An Alternative to the Classical Approach; presented at the XXIIIth International Meeting of the Institute of Management Sciences. Athens.

GERBER, H. (1974a). On additive premium calculation principles; *ASTIN Bulletin*, **7**, 215-222.

GERBER, H. (1974b). On iterative premium calculation principles; *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker*, **74**, 163-172.

LEMAIRE, J. (1977). Selection procedures of regression analysis applied to automobile insurance; *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker*, **77**, 65-72.