DESIGNING OPTIMAL BONUS-MALUS SYSTEMS
FROM DIFFERENT TYPES OF CLAIMS

By

Jean Pinquet [1]

THEMA, University Paris X, 92001 Nanterre, France

Abstract

This paper provides bonus-malus systems which rest on different types of claims. Consistent estimators are given for some moments of the mixing distribution of a multi equation Poisson model with random effects. Bonus-malus coefficients are then obtained with the expected value principle, and from linear credibility predictors. Empirical results are presented for two types of claims, namely claims at fault and not at fault with respect to a third party.

Keywords

Fixed and random effects models, mixing distributions, expected value principle, linear credibility predictors.

Introduction

Bonus-malus systems (later referred to as BMS) in use throughout the world rest on a single type of claim. An exception is Korea, where the severity of claims is allowed for (Lemaire, 1995). Usually, claims at fault for the policyholder are retained. Most of them involve a third party, but they can also stem from material damage caused by the driver to his own vehicle.

The trend towards deregulation in insurance business will make it more and more difficult for countries to maintain a compulsory BMS. For instance, insurance companies in Europe operate within a framework of Freedom to Provide Services, and are not forced to follow BMS rules in foreign countries.

This context gives more value to the design of optimal bonus-malus systems. They are derived from a statistical analysis of insurance contracts, which uses models with random effects. Hidden individual characteristics are first summarized by a fixed effect in the risk distributions of an a priori rating model. This fixed effect could be also referred to as a structure term, or as a heterogeneity component. Considering it as the outcome of a random variable, we obtain distributions for generic individuals in a model with random effects. These distributions happen to be mixtures of those of the a priori rating model, and the generic individuals represent a class of risks defined conditionally on observable rating factors.

Once estimated, the model with random effects allows the prediction of the fixed effects, and the design of optimal BMS. The basic motivation in the development of these systems is to use the history of the policyholder as thoroughly as possible in the prediction of risks.

Whereas the starting point for the history is the number of claims at fault, actuarial literature proposes extensions in different ways.

• The allowance for severity of claims through a dichotomy between claims with or without bodily injury (Picard (1976), Lemaire (1985)), or through their cost (Pinquet (1997a)).

• The allowance for the date of claims, through credibility models with geometric weights (Gerber and Jones (1975), Sundt (1981)).

This paper develops statistical models which lead to BMS from different types of claims. There is evidence of a positive correlation between claims at fault and not at fault (Lemaire (1985)), so you might think of using claims not at fault, in order to assess whether these minor claims are able to reveal hidden features in the number distributions of claims at fault. You could also consider events which are not claims, like infractions to the road safety code. Using moving traffic violations in an experience rating scheme is common practice in all states of the US, because US insurers have direct access to records of the Motor Vehicles Division. A speeding ticket related to more than fifteen mph above the speed limit entails the same penalty as an accident at fault, and so does failure to stop at a traffic light, or failure to respect a stop sign. Overtaking a school bus while its red lights are blinking is still more penalized. Since 1992 in Quebec, the public monopoly which provides automobile insurance for bodily injuries includes a history of "demerit points" in its rating structure. For empirical results on that kind of events, see Dionne and Vanasse (1997).

This issue has already been addressed by actuarial literature, mostly within a multi-guarantee approach. Let us quote for instance

• Larsen et al. (1991) (in a multi-product setting), Partrat (1992, 1993) for models with a degenerate distribution for the random effects.

• Boulanger (1994) for a linear credibility approach and a joint distribution for the random effects.

A multi equation Poisson model with a joint distribution for the random effects (i.e. neither degenerate, nor with independent components) must be the basic tool here. Unfortunately, the likelihood of such a model does not

admit a closed form. The author proposed an inference method for linear and Poisson models with random effects (Pinquet (1997b), (1998)), which stems from consistent estimation of some moments of the mixing distribution. In a parameterized framework, consistent estimators for the parameters of the mixing distribution are then obtained with a method of moments. These estimators are obtained from residuals derived from the a priori rating model. In Section 1.2.2, an example is provided for a multi equation Poisson model with a Gaussian distribution for the random effects. As compared to the preceding contributions, the estimation procedure presented here allows to assess whether it is possible to consider a joint distribution for the random effects, and provides consistent estimators from any a priori rating structure.

The two ways of relating a model with random effects to prediction on longitudinal data are used in Section 1.3.

• The *expected value principle* (Lemaire (1985), Dionne and Vanasse (1989), Pinquet (1997a)), which rests on a parameterized specification of the mixing distribution. It is applied here to a multivariate Gaussian distribution for the random effects.

• The *linear credibility approach* (see Bühlmann (1967), Boulanger (1994) for the issue addressed by the paper), which provides predictors from moments of the mixing distribution. Consistent estimators for these moments are given in section 1.2.2, regardless of a parameterized specification for the mixing distribution. When applied to our model, the bonus-malus coefficient for each type of claim can be seen as a linear combination of "loss to premium" ratios, with a first increasing, then time-vanishing credibility for the other types.

Empirical results are provided in Section 2. We retain here two types of events, namely claims at fault and not at fault with respect to a third party.

If compared to the case where only claims at fault are allowed for, the results obtained are the following.

• Not surprisingly, each claim at fault becomes less meaningful in the prediction, since more types of events are taken into account.

• For each type of claim, the revelation throughout time of hidden features in the number distributions is enhanced. This improvement increases with the frequency of the other types, and with the squared covariances between the random effects.

1. A MULTI EQUATION MODEL FOR NUMBER OF CLAIMS

1.1. Presentation

Suppose q different types of claims. The number of claims of type j $(j = 1, ..., q)$ reported by the policyholder i $(i = 1, ..., p)$ in period t $(t = 1, ..., T_i)$ is denoted as N_j^{it}. It follows a Poisson distribution with a parameter $\lambda_j^{it} = \exp(x_j^{it}\theta_{1j})$. In the last expression, x_j^{it} is a line-vector of covariates, and θ_{1j} a column-vector of parameters. The number variables are

supposed to be independent. This implies that the different types of claims do not overlap, an assumption which may lead to redefine them. For instance, if N_1 is the number of claims at fault with respect to a third party, with among them N_2 claims which entail bodily injury, a two equation model will explain $N_1 - N_2$ and N_2.

Allowing for hidden features in these distributions, we consider the following equations with fixed effects

$$N_j^{it} \sim P(\exp(x_j^{it}\theta_{1j} + u_j^i)). \tag{1}$$

We retain here a time-independent fixed effect for each policyholder and each type of claim.

For a generic individual, the fixed effect u_j^i is the outcome of a random variable U_j^i, whose distribution is supposed not to depend on i. If this distribution is that of U_j, the supplementary parameters are the variances and covariances of the $(U_j)_{1 \leq j \leq q}$. The parameters of the model with random effects are then

$$\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} ; \ \theta_1 = \underset{1 \leq j \leq q}{vec}(\theta_{1j}); \ \theta_2 = \underset{1 \leq j \leq k \leq q}{vec}(V_{jk}), \ V_{jk} = Cov(U_j, U_k). \tag{2}$$

In the preceding expression, the different components of θ_1 and θ_2 are stacked in a column-vector. The parameter space is $\Theta = \Theta_1 \times \Theta_2$, where Θ_2 is a cone of positive semidefinite matrices. This cone is embedded in the space of q-dimensional symmetric matrices, which can be identified with $R^{q(q+1)/2}$. The random effects are supposed to have a null expectation. If $\theta_2 = 0$, we have $U_j \equiv 0 \ \forall j$, and we obtain the a priori rating model as a particular case, since no mixing of distributions is performed.

1.2. Inference from the Lagrangian

1.2.1. Local expansion of the likelihood

An important point to notice is that the model with random effects does not necessarily outperform the a priori rating model on a likelihood criterion, since $\theta_2 = 0$ lies at the boundary of Θ_2. Now, this condition must be fulfilled if you want to design an optimal bonus-malus system from this model, with an estimation performed from likelihood maximization. Here, 0 is the vertex of the cone of positive semidefinite matrices. It is natural then to compute the Lagrangian with respect to the parameters of the mixing distribution. Differentiation with respect to θ_2 can be performed at the boundary of Θ_2, since Θ_2 spans $S_q(R)$, the space of symmetric matrices. For instance, the Lagrangian with respect to a covariance could not be defined as a partial derivative, but is obtained by the extension of a linear form from Θ_2 to $S_q(R)$. Let $\widehat{\theta_1}^0$ be the maximum likelihood estimator for the parameters of the a priori rating model. The separate estimation of the q equations leads to $\widehat{\theta_{1j}}^0$ $(j = 1, ..., q)$, then to $\widehat{\theta_1}^0$ by stacking these components. The Lagrangian

computed later is the score with respect to θ_2, computed for $\theta_1 = \widehat{\theta_1}^0$, $\theta_2 = 0$. Write a symmetric matrix V as $V = \sum_{1 \le j \le k \le q} V_{jk} e_{jk}$. Let $\mathcal{L}_{jk}$ be the Lagrangian with respect to e_{jk}. We have

$$\mathcal{L}_{jj} = \frac{1}{2} \sum_i \left[(res_j^i)^2 - s_j^i \right] ; \quad \mathcal{L}_{jk} = \sum_i res_j^i res_k^i \ (1 \le j < k \le q).$$

In this expression, res_j^i and s_j^i are obtained from the first and second derivatives, with respect to the fixed effects, of the log-likelihood (see Pinquet (1997a) for expressions in a single equation model). This type of derivation has been addressed for a long time by statistical literature, since the seminal papers on the subject date back to Neyman (1959, 1966).

For the model defined in (1) and (2), we obtain

$$res_j^i = N_j^i - \widehat{\lambda_j^i}; \ s_j^i = \widehat{\lambda_j^i} \ \forall i,j,$$

if we write $N_j^i = \sum_t N_j^{it}$; $\widehat{\lambda_j^{it}} = \exp(x_j^{it} \widehat{\theta_{1j}}^0)$; $\widehat{\lambda_j^i} = \sum_t \widehat{\lambda_j^{it}}$.

The model with random effects will outperform the a priori rating model on a likelihood criterion if the Lagrangian belongs to a set defined in Section 3.1.

1.2.2. Estimation of the mixing distribution from the Lagrangian

Rewrite equation (1) as

$$N_j^{it} \sim P(\lambda_j^{it} w_j^i), \ \text{with} \ w_j^i = \exp(u_j^i).$$

If the distribution of the W_j^i is that of W_j, straightforward computations in the model with random effects lead to

$$E(N_j^i) = \lambda_j^i E(W_j); \ E(N_j^i N_k^i) = \lambda_j^i \lambda_k^i E(W_j) E(W_k) \ (j \ne k);$$

$$V(N_j^i) - E(N_j^i) = E^2(N_j^i) CV^2(W_j). \tag{3}$$

We wrote $CV^2(W_j) = V(W_j)/E^2(W_j)$ (remember that $N_j^i = \sum_t N_j^{it}, \lambda_j^i = \sum_t \lambda_j^{it}$). Now it can be proved that

$$\widehat{\lambda_j^i} \to E(N_j^i) \forall i,j, \tag{4}$$

where $\widehat{\lambda_j^i}$ is the frequency-premium in the a priori rating model and where the expectation is taken in the model with random effects (see Pinquet (1997b)). Thus, the a priori premium of any individual converges towards the frequency-risk of the related generic individual, this whatever is the value of the rating factors and of the mixing distribution.

From equation (3), we obtain

$$\widehat{V_{jj}}^{1} = \frac{\sum\limits_{i} \left[(N_j^i - \widehat{\lambda}_j^i)^2 - \widehat{\lambda}_j^i \right]}{\sum\limits_{i} \widehat{\lambda}_j^{i2}} \rightarrow CV^2(W_j);$$

$$\widehat{V_{jk}}^{1} = \frac{\sum\limits_{i} (N_j^i - \widehat{\lambda}_j^i)(N_k^i - \widehat{\lambda}_k^i)}{\sum\limits_{i} \widehat{\lambda}_j^i \widehat{\lambda}_k^i} \rightarrow \frac{Cov(W_j, W_k)}{E(W_j)E(W_k)} \quad (1 \leq j < k \leq q),$$

if the number distributions belong to the model with random effects. Some moments of the mixing distribution are then consistently estimated regardless of a parameterized specification. Notice that the Lagrangian appears at the numerator of these estimators. The superscript "1" is used for the estimators $\widehat{V_{jj}}^{1}$ and $\widehat{V_{jk}}^{1}$ because they are obtained at the first step of the Newton-Raphson algorithm of likelihood maximization, if the initial value is $\theta_1 = \widehat{\theta_1}^0$, $\theta_2 = 0$, with the notations of the preceding section.

If we retain a multivariate Gaussian distribution for the random effects, i.e.

$$U \sim N_q(0, V), \quad \left(U = \underset{j}{vec}(U_j) \right), \quad \text{then}$$

$$V_{jk} = Cov(U_j, U_k) \Rightarrow \frac{Cov(W_j, W_k)}{E(W_j)E(W_k)} = \exp(V_{jk}) - 1.$$

Hence

$$\hat{V}_{jk} = \log(1 + \widehat{V_{jk}}^{1}) \; (1 \leq j \leq k \leq q) \tag{5}$$

are consistent estimators for the parameters of the mixing distribution.

Owing to the unconstrained approach in the estimation, the $\hat{V}_{jk}$ are not bound to belong to the parameter space. For instance, the $(\hat{V}_{jj})_{1 \leq j \leq q}$ are nonnegative if there is an overdispersion of residuals for each of the equations. An optimal BMS can be designed from the data if the matrix $\hat{V}$ derived from the $\hat{V}_{jk}$ is positive semidefinite.

1.3. Prediction with the expected value principle and with a linear credibility approach

1.3.1. Expected value principle

Let us specify a multivariate Gaussian distribution for the random effects. The bonus-malus coefficient for the frequency of claims in equation $j(j = 1, ..., q)$ is the ratio of estimated expectations of $W_j = \exp(U_j)$ with

respect to prior and posterior distributions. Before any estimation, this ratio is equal to

$$\frac{E\left[\exp(U_j + \sum_{k=1}^q (n_k^i U_k - \lambda_k^i \exp(U_k)))\right]}{E\left[\exp(U_j)\right] E\left[\exp(\sum_{k=1}^q (n_k^i U_k - \lambda_k^i \exp(U_k)))\right]}$$

for the policyholder i. From equation (4), we have

$$\widehat{\lambda_k^i} \to E(N_k^i) = \lambda_k^i E[\exp(U_k)] \; \forall i, k,$$

if $\widehat{\lambda_k^i}$ is the frequency-premium in the a priori rating model and if the expectation of N_k^i is taken in the model with random effects. To obtain a bonus-malus coefficient, replace expectations by estimations derived from equation (5), and replace λ_k^i by $\widehat{\lambda_k^i}/\hat{E}[\exp(U_k)]$. Here $\hat{E}[\exp(U_k)] = \exp(\widehat{V_{kk}}/2) = \sqrt{1 + \widehat{V_{kk}}}^1$. The other expectations do not admit a closed form with respect to the parameters, but they can be estimated from simulations. A Choleski decomposition of $\hat{V}$ must be performed first (see Pinquet (1997a) for an application to a two equation model on number and cost of claims).

1.3.2. Linear credibility predictors

A linear credibility predictor of w_j^i which rests on all the equations stems from the affine regression of W_j with respect to the $(N_k^i)_{k=1,...,q}$. The notations are those of Section 1.2.2. For convenience, we will regress $\Lambda_j^i = \lambda_j^i W_j$ on the $(N_k^i)_{k=1,...,q}$. Since a bonus-malus coefficient is a ratio, results are unchanged if the random effect is multiplied by a non-random value. We denote the predictor of $E(\Lambda_j^i)$ as $a_j^i + \sum_{k=1}^q b_{jk}^i n_k^i$, with

$$\left(\widehat{a_j^i}, \widehat{b_{jk}^i}\right)_{k=1,...,q} = \arg \min_{a_j^i, b_{jk}^i} E\left[\left(\Lambda_j^i - a_j^i - \sum_{k=1}^q b_{jk}^i N_k^i\right)^2\right].$$

The expectation is taken in the model with random effects. Writing

$$N^i = \underset{1 \le k \le q}{vec} (N_k^i); \; b_j^i = \underset{1 \le k \le q}{vec} (b_{jk}^i),$$

it is well known that

$$\widehat{b_j^i} = \left[V(N^i)\right]^{-1} Cov(N^i, \Lambda_j^i); \; \widehat{a_j^i} = E(N_j^i) - \sum_{k=1}^q \widehat{b_{jk}^i} E(N_k^i), \qquad (6)$$

since $E(\Lambda_j^i) = E(N_j^i)$. If the moments in the preceding equation are seen as individual parameters, consistent estimations are obtained from the limits almost everywhere

$$\widehat{\lambda}_j^i \to E(N_j^i); \ \widehat{V_{jk}}^1 \to \frac{Cov(W_j, W_k)}{E(W_j)E(W_k)} \ \forall j, \forall k$$

(see Section 1.2.2). Remember that $\widehat{\lambda}_j^i$ is the cumulated frequency-premium, computed in the a priori rating model. From equation (3), we obtain

$$\hat{E}(N_k^i) = \widehat{\lambda}_k^i; \ \hat{V}(N_k^i) = \widehat{\lambda}_k^i + \widehat{\lambda}_k^i{}^2 \widehat{V_{kk}}^1; \ \widehat{Cov}(N_k^i, N_{k'}^i) = \widehat{\lambda}_k^i \widehat{\lambda}_{k'}^i \widehat{V_{kk'}}^1 \ (k \neq k') \quad (7)$$

as consistent estimators for the moments of equation (6). Besides,

$$Cov(N_k^i, \Lambda_j^i) = E(N_k^i)E(N_j^i)\frac{Cov(W_j, W_k)}{E(W_j)E(W_k)} \ \forall k \Rightarrow \widehat{Cov}(N_k^i, \Lambda_j^i) = \widehat{\lambda}_k^i \widehat{\lambda}_j^i \widehat{V_{kj}}^1 \quad (8)$$

are consistent estimators for the other moments of equation (6).

Let a_j^i and b_{jk}^i be the coefficients derived from equation (6) and from the preceding estimators. The predictor of $E(\Lambda_j^i)$ is equal to

$$\widehat{a_j^i} + \sum_{k=1}^{q} \widehat{b_{jk}^i} n_k^i = \widehat{\lambda}_j^i + \sum_{k=1}^{q} \widehat{b_{jk}^i}(n_k^i - \widehat{\lambda}_k^i) = \widehat{\lambda}_j^i \left(1 + \sum_{k=1}^{q} \widehat{b_{jk}^i} \frac{\widehat{\lambda}_k^i}{\widehat{\lambda}_j^i} \left(\frac{n_k^i}{\widehat{\lambda}_k^i} - 1 \right) \right) . \quad (9)$$

From equations (6), (7) and (8), the $\left(\widehat{b_{jk}^i} \right)_{k=1,\dots,q}$ are the solutions of the linear system

$$\forall k = 1, \dots, q \ : \ (1 + \widehat{\lambda}_k^i \widehat{V_{kk}}^1)\widehat{b_{jk}^i} + \sum_{k' \neq k} \widehat{\lambda}_{k'}^i \widehat{V_{kk'}}^1 \widehat{b_{jk'}^i} = \widehat{\lambda}_j^i \widehat{V_{kj}}^1 \quad (10)$$

(both members were divided by $\widehat{\lambda}_k^i$).

The bonus-malus coefficient for equation j is the ratio of the predictor and of $\widehat{\lambda}_j^i$, the a priori estimation of $E(\Lambda_j^i)$. It is obtained in equation (9) as a linear function of "loss to premium" ratios. From equation (4), this entails a fairness property in the prediction. Notice that the preceding results do not rest on a parameterized specification for the mixing distribution.

Suppose that the duration of observation converges towards infinity. The frequency premiums will behave in the same way, and we have

$$\lim_{T_i \to +\infty} \widehat{b_{jj}^i} = 1; \ \lim_{T_i \to +\infty} \widehat{b_{jk}^i} = 0 \ (k \neq j),$$

where T_i denotes a number of periods, or a duration if we reason in continuous time. To obtain this result, denote the jth vector of the canonical basis of R^q as e_j. From equations (7) and (8), we obtain

$$\hat{V}(N^i)e_j = \widehat{Cov}(N^i, \Lambda_j^i) + \widehat{\lambda}_j^i e_j.$$

Equation (6) leads then to

$$\widehat{b}_j^i = \left(I_q - \widehat{\lambda}_j^i [\hat{V}(N^i)]^{-1}\right) e_j.$$

Hence $\lim_{T_i \to +\infty} \widehat{b}_j^i = e_j$, since each premium becomes negligible if compared to its square.

This result means that, for each type of claim, the contribution of the other types in the prediction vanishes with time. The credibility coefficient b_{jj}^i increases with time from 0 to 1, whereas the others coefficients (i.e. $(\widehat{\lambda}_k^i/\widehat{\lambda}_j^i)\widehat{b}_{jk}^i$, $k \neq j$) first increase from 0, then decrease towards 0.

Besides, this property entails the consistency of the bonus-malus coefficient, since its limit is $w_j^i/E(W_j)$, the ratio of expectations of N_j^i taken in the models with fixed and random effects.

2. EMPIRICAL RESULTS FOR CLAIMS AT FAULT AND NOT AT FAULT WITH RESPECT TO A THIRD PARTY

The motivation here is to improve the prediction of the third party liability risk. We compare here bonus-malus coefficients derived from claims at fault, and from all the claims that trigger the guarantee.

Let us recall briefly the compensation scheme in France, if two cars are involved in an accident. The fault is determined from reports of the drivers and of the police, if any. In order to avoid contestation of the repair costs by the insurer of the third party, direct compensation works in the first place. The insurer of the driver not at fault receives a lump sum from the other insurer, and pays for the material damage of his policyholder (at least below a certain level). On our data, the average costs of claims at fault and not at fault are respectively equal to 11000 FF and 1400 FF (the observations date back to several years). The last average is the sum of 800 FF, a difference between the repair costs and the lump sum, and 600 FF due to payments to other third parties and administrative costs. The motivation here is to assess whether minor claims (claims not at fault) are able to reveal hidden features in the number distributions of claims at fault.

We will first compute bonus-malus coefficients for the frequency of claims at fault. Applying these coefficients to the third party liability risk supposes that claims not at fault have a negligible cost. We will then introduce costs in an example.

The working sample is part of the automobile policyholders portfolio of a French insurance company. The rating factors are:
- The characteristics of the vehicle: group, age.
- The characteristics of the insurance contract: type of use, geographic zone.

Other rating factors are the policyholder's occupation, as well as the year when the period began (in order to allow for a generation effect). The

different levels of these six rating factors are represented by thirty-five indicators. The periods have not the same duration, and the parameters of the Poisson distributions are proportional to this duration.

The policyholders considered in the working sample are observed on one, two or three periods. More precisely, we have

	Number of policyholders observed on:
at least one period	85909
at least two periods	68344
three periods	44428

Hence, $p = 85909$; $\sum_{i=1}^{p} T_i = 198681$. The working sample is here a non-balanced panel data set, and the average duration of observation of a policyholder is equal to nineteen months.

The first equation will be related later to claims at fault, and the second to claims not at fault. The average of n_1 conditional on $n_2 = 0, 1$ or 2 is respectively equal to 0.092, 0.149 and 0.207, so there is evidence of a positive correlation between the two number distributions. From the estimation of two Poisson models with the aforementioned covariates, we obtain

$$\sum_{i=1}^{p} n_1^i = 8495; \quad \sum_i (n_1^i - \widehat{\lambda_1^i})^2 = 9374.6; \quad \sum_i \widehat{\lambda_1^i}^{\,2} = 1191.6.$$

$$\widehat{V_{11}}^1 = \frac{\sum_i \left[(n_1^i - \widehat{\lambda_1^i})^2 - \widehat{\lambda_1^i} \right]}{\sum_i \widehat{\lambda_1^i}^{\,2}} = 0.738; \quad \widehat{V_{11}} = \log\left(1 + \widehat{V_{11}}^1\right) = 0.553.$$

$$\sum_{i=1}^{p} n_2^i = 9968; \quad \sum_i (n_2^i - \widehat{\lambda_2^i})^2 = 10968.12; \quad \sum_i \widehat{\lambda_2^i}^{\,2} = 1592.18.$$

$$\widehat{V_{22}}^1 = 0.628; \quad \widehat{V_{22}} = \log(1 + \widehat{V_{22}}^1) = 0.487.$$

$$\sum_i \left[(n_1^i - \widehat{\lambda_1^i})(n_2^i - \widehat{\lambda_2^i}) \right] = 464.5; \quad \sum_i \widehat{\lambda_1^i} \widehat{\lambda_2^i} = 1267.6.$$

$$\widehat{V_{12}}^1 = \frac{\sum_i \left[(n_1^i - \widehat{\lambda_1^i})(n_2^i - \widehat{\lambda_2^i}) \right]}{\sum_i \widehat{\lambda_1^i} \widehat{\lambda_2^i}} = 0.366; \quad \widehat{V_{12}} = \log(1 + \widehat{V_{12}}^1) = 0.312.$$

Let us compute bonus-malus coefficients for the frequency of claims at fault. From equation (10) and the preceding estimations, the linear credibility predictor is obtained from the linear system

$$\left(1 + (0.738\widehat{\lambda_1})\right)\widehat{b_{11}} + (0.366\widehat{\lambda_2})\widehat{b_{12}} = 0.738\widehat{\lambda_1}$$
$$(0.366\widehat{\lambda_1})\widehat{b_{11}} + \left(1 + (0.628\widehat{\lambda_2})\right)\widehat{b_{12}} = 0.366\widehat{\lambda_1} \qquad (11)$$

if both types of claims are taken into account (we drop the individual index). Consider an insurance contract without claim of any type reported during the first year. Suppose that the frequency premiums related to claims at fault and not at fault are respectively equal to $\widehat{\lambda_1} = 6.5\%$ and $\widehat{\lambda_2} = 7.5\%$, which are roughly the average values for one year. The credibility coefficients related to both types of claims are then

$$\widehat{b_{11}} = 4.5\%; \; \frac{\widehat{\lambda_2}}{\widehat{\lambda_1}}\widehat{b_{12}} = 2.5\%.$$

From equation (9), they represent the contribution to a 7% bonus for this type of contract. The bonus is found equal to 6.7% with the expected value principle (see Pinquet (1997a) for a computation of coefficients with this approach). If only claims at fault are taken into account, the bonus is equal to 4.6% with the linear credibility approach, and equal to 4.4% with the expected value principle.

Let us use equation (11) to study the linear credibility predictor as a function of the frequency-premiums (or as a function of time). If the premiums per year have the average values, the credibility granted to claims not at fault (i.e. $(\widehat{\lambda_2}/\widehat{\lambda_1})\widehat{b_{12}}$) increases from 0 to 15.7% during 25 years, then decreases towards 0.

Suppose now: $\widehat{\lambda_1} = \widehat{\lambda_2} = 1$, which means about fifteen years of observation on average. We obtain: $\widehat{b_{11}} = 0.396$, $\widehat{b_{12}} = 0.136$. The bonus-malus coefficient derived from both types of claims with a linear credibility approach is then equal to $1 + (0.396 \times (n_1 - 1)) + (0.136 \times (n_2 - 1))$. Here, three claims not at fault are as significant as one claim at fault. But it must be kept in mind that this type of result depends on the level of premiums, and hence depends on time. We obtain for example

TABLE 1

BONUS-MALUS COEFFICIENTS FOR THE FREQUENCY OF CLAIMS AT FAULT (LINEAR CREDIBILITY APPROACH)

		n_1			
		0	1	2	3
claims at fault only		0.58	1	1.42	1.85
claims at fault or not at fault					
	$n_2 = 0$	0.47	0.86	1.26	1.66
	$n_2 = 1$	0.60	1	1.40	1.79
	$n_2 = 2$	0.74	1.14	1.53	1.93
	$n_2 = 3$	0.88	1.27	1.67	2.06

JEAN PINQUET

The bonus-malus coefficient derived from claims at fault only is equal to $1 + (0.425 \times (n_1 - 1))$. If both types of claims are taken into account, the claims at fault are less meaningful in the prediction, but the decrease is not important.

Let us use the average cost of claims of both types in the prediction of the third party liability risk. The linear credibility predictor for the frequency of claims not at fault is obtained from a linear system derived from equation (10). We obtain

$$\widehat{b_{21}} = 0.136; \quad \widehat{b_{22}} = 0.355.$$

A bonus-malus coefficient for third party liability can be obtained from an average of the coefficients related to both types of claims, with weights equal to their average cost, since here frequency-premiums are equal. The coefficient is equal to $1 + (0.367 \times (n_1 - 1)) + (0.161 \times (n_2 - 1))$. As compared to the preceding results, the claims not at fault become more significant, whereas an opposite effect is obtained for claims at fault.

With the same assumptions on the premiums, the coefficients for the frequency of claims at fault which are derived from the expected value principle are the following

TABLE 2

BONUS-MALUS COEFFICIENTS FOR THE FREQUENCY OF CLAIMS AT FAULT (EXPECTED VALUE PRINCIPLE)

		n_1			
		0	1	2	3
claims at fault only		0.65	0.94	1.30	1.74
claims at fault or not at fault					
	$n_2 = 0$	0.56	0.81	1.12	1.50
	$n_2 = 1$	0.67	0.94	1.28	1.68
	$n_2 = 2$	0.78	1.07	1.43	1.85
	$n_2 = 3$	0.89	1.20	1.58	2.03

For a given number of claims at fault, the bonus-malus coefficient increases at a steady pace with respect to the number of claims not at fault. If this number is fixed, the bonus-malus coefficient is a convex function of the number of claims at fault.

Let us compare the evolution throughout time of bonus-malus coefficients for the frequency of claims at fault, if this type of claims or both types are accounted for. Using both types of claims will increase the number of events to be used in the prediction, but on the other hand each claim at fault will be less meaningful. Credibility predictors are here compared on a portfolio. We consider a simulated portfolio, derived from

the working sample. The characteristics of each policyholder are those of the first period, and we suppose that they remain unchanged. Fixed effects are drawn at random from the bivariate Gaussian distribution estimated before. Then numbers of claims are simulated, and bonus-malus coefficients are computed with the linear credibility approach. Their dispersion is measured by the standard deviation, almost equal in the simulations to the coefficient of variation because of the fairness of the rating structure. The results are given after T years of observation. Because of the consistency of the bonus-malus coefficients, the limit of the standard deviation when T converges towards infinity is $\widehat{CV}(W_1) = \sqrt{\widehat{V_{11}}}^1$.

TABLE 3

EVOLUTION THROUGHOUT TIME OF BONUS-MALUS COEFFICIENTS

Standard deviation of the bonus-malus coefficients	$T=1$	$T=2$	$T=5$	$T=10$	$T=+\infty$
from claims at fault only	0.177	0.244	0.363	0.471	0.859
from claims at fault or not at fault	0.198	0.271	0.395	0.502	0.859

The addition of claims not at fault improves the prediction of the fixed effects. The result may seem disappointing, since the increase of the standard deviation is equal to 12% after one year. The improvement depends on the frequency of claims not at fault, and on the squared covariance between the two random effects (see Section 3.2).

3. APPENDIX

3.1. Do data allow distribution mixing from the a priori rating model? Geometrical conditions for the Lagrangian

On the data, the model with random effects will outperform the a priori rating model on a likelihood criterion if the Lagrangian does not belong to Θ_2^-, the negative dual of Θ_2, which is equal to

$$\Theta_2^- = \left\{ \mathcal{L} = \sum_{1 \leq j \leq k \leq q} \mathcal{L}_{jk} e_{jk} / \sum_{1 \leq j \leq k \leq q} \mathcal{L}_{jk} V_{jk} \leq 0 \forall \theta_2 \in \Theta_2, \ \theta_2 = \sum_{1 \leq j \leq k \leq q} V_{jk} e_{jk} \right\}.$$

If $q = 1$, this simply means that $\mathcal{L}_{11}$ is nonnegative, which implies an overdispersion of residuals. If $q = 2$, we have

$$\Theta_2 = \left\{ \sum_{1 \leq j \leq k \leq 2} V_{jk} e_{jk} / V_{11} + V_{22} \geq 0, \ V_{11} V_{22} - V_{12}^2 \geq 0 \right\};$$

$$\Theta_2^- = \left\{ \mathcal{L} = \sum_{1 \leq j \leq k \leq 2} \mathcal{L}_{jk} e_{jk} / \mathcal{L}_{11} + \mathcal{L}_{22} \leq 0, \ 4\mathcal{L}_{11}\mathcal{L}_{22} - \mathcal{L}_{12}^2 \geq 0 \right\}$$

(the proof for the expression of Θ_2^- is available from the author upon request). The model with random effects will outperform the a priori rating model on a likelihood criterion if one of the two last conditions is not fulfilled.

Degenerate distributions for the random effects are investigated by Larsen et al. (1991) and Partrat (1992, 1993) for bonus-malus systems with two guarantees. If the Lagrangian belongs neither to Θ_2 nor to Θ_2^-, the direction defined by the projection of the Lagrangian on Θ_2 will correspond to the steepest ascent for the log-likelihood. It is related to degenerate distributions for which $U_2 = aU_1$, or $U_1 \equiv 0$.

3.2. Rate of revelation of hidden features in the number distributions

Owing to the consistency of an optimal bonus-malus system, the limit of the bonus-malus coefficient related to type $j, j = 1, ..., q$ and to the policyholder i is equal to $w_j^i / E(W_j)$, with the notations of the paper. On a portfolio, the limit distribution is that of $W_j / E(W_j)$.

Let us consider a simulated portfolio, with time independent rating factors, and start with $q = 1$. We reason in continuous time, and write

$$N_1^t \sim P(\lambda_1 w_1 t) \tag{12}$$

for the distributions of claims reported by a policyholder between 0 and t years (we drop the individual index). This distribution depends on λ_1, a function of the rating factors, and on the fixed effect w_1. The bonus-malus coefficient derived from the negative binomial model (expected value principle) or from the linear credibility approach can be written as

$$BM_1^t = \frac{1 + \widehat{V_{11}}^{-1} N_1^t}{1 + \widehat{V_{11}}^{-1} \widehat{\lambda_1} t}$$

Hence $BM_1^{dt} = 1 + \widehat{V_{11}}^{-1}(N_1^{dt} - \widehat{\lambda_1} dt) + o(dt)$. If we replace $\widehat{\lambda_1}$, the premium for one year in the a priori rating model, by its limit $\lambda_1 E(W_1)$, we have

$$E(BM_1^{dt}) = 1 + \lambda_1 \widehat{V_{11}}^{-1}(w_1 - E(W_1))dt + o(dt); \ V(BM_1^{dt}) = \lambda_1 w_1 \left(\widehat{V_{11}}^{-1} \right)^2 dt + o(dt)$$

for the risk distribution given in (12). On a portfolio, $\lambda_1 w_1$ is the outcome of Λ_1, a random variable. Then we obtain

$$V(BM_1^{dt}) = E(\Lambda_1) \left(\widehat{V_{11}}^{-1} \right)^2 dt + o(dt)$$

for the unconditional variance. At the beginning, the rate of revelation increases with the frequency of claims, and with the squared variance of the random effect.

Let us consider several types of claims. The bonus-malus coefficient for one type of claim (say, type $j, j = 1, ..., q$) obtained from the linear credibility approach is equal to

$$BM_{j/1,...,q}^t = 1 + \sum_{k=1}^q \widehat{b_{jk}^t} \frac{\widehat{\lambda_k}}{\widehat{\lambda_j}} \left(\frac{N_k^t}{\widehat{\lambda_k}t} - 1 \right) .$$

The $\widehat{b_{jk}^t}$ were computed in Section 1.3.2, and it is easily seen from equation (10) that

$$\widehat{b_{jk}^{dt}} = \widehat{\lambda_j} \widehat{V_{jk}}^{\,-1} dt + o(dt),$$

since premiums are negligible at the beginning. Hence

$$BM_{j/1,...,q}^{dt} = 1 + \sum_{k=1}^q \widehat{V_{jk}}^{\,-1} \left(N_k^{dt} - \widehat{\lambda_k} dt \right) + o(dt) .$$

Computing first the conditional expectation and variance, then the unconditional variance, we obtain

$$V(BM_{j/1,...,q}^{dt}) = \left[\sum_{k=1}^q E(\Lambda_k) \left(\widehat{V_{jk}}^{\,-1} \right)^2 \right] dt + o(dt)$$

and the conclusion given in the paper.

REFERENCES

BOULANGER, F. (1994) Système de Bonus-Malus Multi-Garanties. *Proceedings of the XXV ASTIN Colloquium*, Cannes, 183-194.

BÜHLMANN, H. (1967) Experience Rating and Credibility. *ASTIN Bulletin* 4, 199-207.

DIONNE, G. and VANASSE, C. (1989) A Generalization of Automobile Insurance Rating Models: the Negative Binomial Distribution with a Regression Component. *ASTIN Bulletin* 19, 199-212.

DIONNE, G. and VANASSE, C. (1997) The Role of Memory and Saving in Long-Term Contracting with Moral Hazard: An Empirical Evidence in Automobile Insurance. *Mimeo, Risk Management Chair*, HEC-Montreal.

GERBER, H., JONES, D. (1975) Credibility Formulas of the Updating Type. *Transactions of the Society of Actuaries* Vol. XXVII, 31-52.

LARSEN, C.R., DENGSOE, C., EGEBO, J.H., HANSEN, J.E. (1991) *Risk analysis of customers – not products*. SCOR, International Prize in Actuarial Science, 71-120.

LEMAIRE, J. (1985) *Automobile Insurance: Actuarial Models*. Huebner International Series on Risk, Insurance and Economic Security.

LEMAIRE, J. (1995) *Bonus-Malus Systems in Automobile Insurance*. Huebner International Series on Risk, Insurance and Economic Security.

NEYMAN, J. (1959) Optimal Asymptotic Tests of Composite Statistical Hypotheses. *Probability and Statistics. The Harald Cramér Volume*, 213-214. Wiley, New York.

NEYMAN, J., SCOTT, E.L. (1966) On the Use of $C(\alpha)$ Optimal Tests of Composite Hypotheses. *Bulletin of the International Statistical Institute*, 41 I, 477-497.

PARTRAT, C. (1992) Risque et Statistique. *Bulletin de l'I.A.F.* 356-357, 11-24.

PARTRAT, C. (1993) Compound Poisson Models for Two Types of Claims. *Proceedings of the XXIV ASTIN Colloquium*, Cambridge.

PICARD, P. (1976) Généralisation de l'Etude sur la Survenance des Sinistres en Assurance Automobile. *Bulletin Trimestriel de l'Institut des Actuaires Français*, 204-267.

PINQUET, J. (1997a) Allowance for Costs of Claims in Bonus-Malus Systems. *ASTIN Bulletin* 27, 33-57.

PINQUET, J. (1997b) Testing for Heterogeneity Through Consistent Estimators. *Document de Travail THEMA 9714*.

PINQUET, J. (1998) *Experience Rating Through Heterogeneous Models*. Forthcoming in Insurance Handbook, Kluwer Academic Publishers, Huebner International Series on Risk, Insurance and Economic Security (Editor: Georges Dionne).

SUNDT, B. (1981) Credibility Estimators with Geometric Weights. *Insurance: Mathematics and Economics* 7, 113-122.

JEAN PINQUET
Université de Paris X, U.F.R. de Sciences Economiques
200, Avenue de la République 92001 NANTERRE CEDEX
FRANCE
Phone: 33 1 40 97 47 58; Fax 33 1 40 97 71 42;
E-mail: pinquet@u-paris10.fr