© *The Journal of Risk and Insurance*, 2003, Vol. 70, No. 4, 577-599

Bonus-Malus Scales in Segmented Tariffs With Stochastic Migration Between Segments

Natacha Brouhns
Montserrat Guillén
Michel Denuit
Jean Pinquet

Abstract

This article proposes a computer-intensive methodology to build bonus-malus scales in automobile insurance. The claim frequency model is taken from Pinquet, Guillén, and Bolancé (2001). It accounts for overdispersion, heteroskedasticity, and dependence among repeated observations. Explanatory variables are taken into account in the determination of the relativities, yielding an integrated automobile ratemaking scheme. In that respect, it complements the study of Taylor (1997).

Introduction and Motivation

A Priori Risk Classification

Nowadays, it has become extremely difficult for insurance companies to maintain cross-subsidies among different risk categories in a competitive setting. Therefore, actuaries have to design a tariff structure that will fairly distribute the burden of claims among policyholders. If the risks in the portfolio are not all identically distributed, it is fair to partition all policies into homogenous classes with all policyholders belonging to the same class paying the same premium.

Natacha Brouhns and Michel Denuit work at the Institut de Statistique, Université Catholique de Louvain, Belgium; Montserrat Guillén at the Department at Economics, University of Barcelona, Spain; and Jean Pinquet at U.M.R. de Sciences Economiques, Université de Paris X, France. The authors can be contacted via e-mail: brouhns@stat.ucl.ac.be, guillen@eco.ub.es, denuit@stat.ucl.ac.be, and pinquet@u-paris10.fr, respectively. Natacha Brouhns thanks the Université Catholique de Louvain for financial support through a "Fonds Spéciaux de Recherche" research grant. This work has been partly performed while Natacha Brouhns and Michel Denuit were visiting the Department of Econometrics of the University of Barcelona. The warm hospitality of the members of the Department and grant SEC2001-3672 is gratefully acknowledged. The authors benefited from stimulating discussions with Richard Derrig and Jean Lemaire. They are indebted to Jean Lemaire for providing them information about the Massachusetts safe driver insurance plan. Finally, the authors thank two anonymous referees for their careful reading of the original manuscript and for numerous remarks that greatly improved the readibility of the article.

The classification variables introduced to partition risks into cells are called *a priori* variables (as their values can be determined before the policyholder starts to drive). In automobile third-party liability insurance, they commonly include the age, gender, and occupation of the policyholders; the type and use of their car; the place where they reside; and sometimes even the number of cars in the household, marital status, smoking behavior, or the color of the vehicle. It is convenient to achieve *a priori* classification with the help of generalized linear models; see, e.g., Renshaw (1994) or Pinquet (2000) for applications in actuarial sciences, and Dobson (1990) for a general overview of the statistical theory.

A Posteriori Corrections

Many important factors cannot be taken into account *a priori*; think, for instance, of the swiftness of reflexes, aggressiveness behind the wheel, or knowledge of the highway code. Consequently, tariff cells are still quite heterogenous despite the use of many classification variables. Data involving claim counts exhibit variability exceeding that explained by Poisson models. This is due to residual heterogeneity and can be modeled by a random effect in a statistical model.

It is reasonable to believe that the hidden characteristics are partly revealed by the number of claims reported by the policyholders. Hence, the adjustment of the premium from the individual claims experience in order to restore fairness among policyholders. In that respect, the allowance of past claims in a rating model derives from an exogenous explanation of serial correlation for longitudinal data. In this case, correlation is only apparent and results from the revelation of hidden features in the risk characteristics.

It is worth mentioning that serial correlation for claim numbers can also receive an endogeneous explanation. In this framework, the history of individuals modifies the risk they represent; this mechanism is termed "true contagion," referring to epidemiology. For instance, a car accident may modify the perception of danger behind the wheel and lower the risk to report another claim in the future. Insurance rating schemes also provide incentives to careful driving and should induce negative contagion. Nevertheless, the main interpretation for automobile insurance is exogenous, since positive contagion (i.e., policyholders who reported claims in the past are more likely to produce claims in the future than those who did not) is always observed for numbers of claims, whereas true contagion should be negative.

Once estimated, the statistical model incorporating the portfolio heterogeneity can be used to perform prediction on longitudinal data and allows for experience rating in casualty insurance. In an empirical Bayesian setting, the prediction is derived from the expectation of a random effect with respect to a posterior distribution taking into account the history of the individual. According to Lemaire (1995), a bonus-malus system calculated using Bayesian analysis is said to be optimal. Such a system is fair: any insured has to pay at each renewal a premium proportional to the estimate of his claim frequency taking into account, through Bayes theorem, all the information gathered in the past. An introduction to these concepts can be found in Pinquet (2000).

Dynamic Heterogeneity

The vast majority of the articles appearing in the actuarial literature considered time-independent heterogenous models. Noticeable exceptions are the pioneering articles

by Gerber and Jones (1975), Sundt (1981, 1988), and Pinquet, Guillén, and Bolancé (2001). The allowance for an unknown underlying random parameter that develops over time is justified since unobservable factors (hidden and relevant exogenous variables) influencing the driving abilities are not constant. One might consider either shocks (induced by events like divorces or nervous breakdown, for instance) or continuous modifications (e.g., due to a learning effect).

Another reason to allow for random effects that vary with time relates to moral hazard. Indeed, individual efforts to prevent accidents are unobserved and feature temporal dependence. The policyholders may adjust their efforts for loss prevention according to their experience with past claims, the amount of premium, and awareness of future consequences of an accident (due to experience rating schemes). The effort variable determines the moral hazard and is modeled by a dynamic unobserved factor.

Of course, it is hopeless in practice to discriminate between residual heterogeneity due to unobservable characteristics of drivers that significantly affect the risk of accident and their individual efforts to prevent such accidents. Both effects get mixed in the latent process. Since the observed contagion between annual claim numbers is always positive, the effect of omitted explanatory variables seems to dominate moral hazard. Anyway, this issue has no practical implication since predictions depend on observed contagion, but not on its nature. See, e.g., Pinquet (2000, Section 14.5) for more details on the identifiability of the nature of contagion.

It is essential at this stage to recognize explicitly the complementary nature of *a priori* and *a posteriori* ratemakings. In seminal articles, Dionne and Vanasse (1989, 1992) proposed a credibility model that integrates *a priori* and *a posteriori* information on an individual basis. Pinquet, Guillén, and Bolancé (2001) have recently generalized the Dionne–Vanasse model by allowing the random effects to vary with time, in the spirit of Gerber and Jones (1975) and Sundt (1988). Specifically, random effects are now described by a stationary process.

European and Asian Bonus-Malus Scales

In many countries, the amount of premium paid by a policyholder is the product of a base premium by a bonus-malus coefficient. The bonus-malus scale is the set of premium levels. Future premiums depend on the number of claims at fault, and are usually derived from a transition table (see, e.g., Lemaire, 1995, for numerous examples).

Bonus-malus scales consist of a finite number of levels, each with its own relativity. New policyholders have access to a specified level. After each year, the policy moves up or down according to transition rules. As an example, let us consider the two experience rating systems defined in Taylor (1997). There are nine bonus-malus levels, of which level 6 is the starting one. A higher level number indicates a higher premium. The level for year $t+1$, L_{t+1}, is given in the function of the level L_t of year t. If no claims have been reported by the policyholder then

$$L_{t+1} = \max\{1, L_t - 1\}.$$

If a number of claims, $n_t > 0$, have been reported during year t then

$$L_{t+1} = \min\{9, L_t + 4n_t\} \text{ (for the severe bonus-malus system),}$$
$$L_{t+1} = \min\{9, L_t + 2n_t\} \text{ (for the mild bonus-malus system).}$$

We will use these systems to illustrate our work.

In case a bonus-malus system is in force, all policies in the same tariff class can be partitioned according to the level they occupy in the bonus-malus scale. In that respect, the bonus-malus mechanism can be considered as a refinement of *a priori* risk evaluation splitting each risk class in a number of subcategories according to individual past claims histories. With the noticeable exception of Taylor (1997), all the authors who integrated *a priori* and *a posteriori* ratemaking in fact worked within the frequential credibility framework; none of them considered the problem of designing bonus-malus scales.

The relativity associated to level ℓ is r_ℓ; the meaning is that an insured at level ℓ pays an amount of premium equal to r_ℓ percent of the base premium. Let us denote as P_ℓ the amount of premium corresponding to level ℓ, i.e., $P_\ell = P_6 \times r_\ell$ percent, where P_6 is the base premium, corresponding to the starting level 6 ($r_6 \equiv 100$). Since the policyholders have been classified according to their observable characteristics, the base premium P_6 varies with policyholders' profiles. The main task of the actuary is to determine the premium relativities associated with each level of the scale. For a thorough presentation of the techniques relating to bonus-malus systems, we refer the reader to Lemaire (1995).

North-American Bonus-Malus Scales

The experience rating systems in force in North America differ from those described in the preceding sections, as they not only incorporate accidents at fault but also elements of the policyholders' driving record. The classes and the premium levels are usually very crude (typically, reporting one accident entails a 50 percent penalty for 3 years). The Massachusetts system, described next, is more sophisticated than the others in force in North America.

Let us briefly present the Massachusetts safe driver insurance plan. This program is mandated by state law and encourages safe driving by rewarding drivers who do not cause an accident, or incur a traffic law violation. Specifically, each policyholder is assigned a level between 9 and 35, based on his driving record during the previous 6 years. A new driver begins at level 15 (relativity of 100 percent). Occupying any level below 15 entails a premium discount, whereas above level 15 the driver pays a surcharge. For each incident-free year of driving, the policyholder goes down one level. The driver will move up a certain number of levels based on the type of accident: two levels for a minor traffic violation, three levels for a minor at-fault accident, four levels for a major at-fault accident, and five levels for a major traffic violation. The Massachusetts system forgets all incidents after 6 years.

The actuarial modeling of such a system would require an analysis in the vein of Pinquet (1998). The approach developed in this article, focused on European and Asian bonus-malus mechanisms, can nevertheless be adapted to systems like the one in force in Massachusetts. Of course, the actuary needs to have at his disposal appropriate data bases. Technical aspects are deferred to a subsequent article.

Bonus-Malus Systems as Markov Chains

Bonus-malus systems can be modeled using Markov chains. This route has been followed for a long time by numerous researchers, such as Norberg (1976), Gilde

and Sundt (1989), Lemaire (1995), Denuit and Dhaene (2001), and Centeno and Silva (2001).

We assume that the scale possesses the following Markovian property: the knowledge of the present level and of the number of claims of the present year suffice to determine the level to which the policy is transferred. This ensures that the bonus-malus system may be represented by a Markov chain (at least conditionally on the observable characteristics and random effects). This can be stated as follows: the future (the level for year $t+1$) depends on the present (the level for year t and the number of accidents reported during year t) and not on the past (the claim history and the levels occupied during years $1, 2, \ldots, t-1$).

A fundamental difference with the traditional approaches is that we lose the homogeneity of the chain in a dynamic segmented environment. Indeed, if the observable characteristics of the policyholders are allowed to vary in time, the claim frequencies are no longer constant and the trajectory of the policyholder in the bonus-malus scale is no longer described by a homogenous Markov chain, but well by a nonhomogenous one. Consequently, the classical techniques based on stationary distributions cannot be applied to the problem of determining the relativities.

Technical Weaknesses of Bonus-Malus Scales

Compared to frequency credibility models, bonus-malus scales suffer from some drawbacks. First and foremost is the progressive clustering of policyholders at the lowest levels of the scale (because transition rules are often not severe enough). Moreover, since the relativities are the same whatever the risk class to which the policyholders belong, those scales overpenalize *a priori* bad risks. Let us explain this phenomenon, put in evidence, e.g., by Taylor (1997). Over time, policyholders will be distributed over the levels of the bonus-malus scale. Since their trajectory is a function of past claims history, policyholders with low *a priori* claim frequencies will tend to gravitate at the lowest levels of the scale. The opposite is true for individuals with high *a priori* claim frequencies. So, those policyholders who have been granted premium discounts at policy issuance (on the basis of their observable characteristics) will also be rewarded *a posteriori* (because they occupy the lowest levels of the bonus-malus scale). Conversely, the policyholders who have been penalized at policy issuance (because of their observable characteristics) will cluster at the highest bonus-malus levels and will consequently be penalized again. The only solution to this problem is to let the relativities depend on observable characteristics. However, this makes the scale difficult to enforce in practice; this route will not be followed here. Instead, we will try to determine the r_i's in accordance with the extent of *a priori* classification.

Agenda

Let us now describe the content of this article. In the following section, we present the rating model incorporating both *a priori* and *a posteriori* risk classification. We distinguish between static and dynamic random effects. We also separate the latent process from the observed covariate process. In both cases, we make use of a Markovian assumption. The section "Numerical Illustrations" presents numerical illustrations. The final section concludes. Several appendices gather technical details about different aspects of the work.

RATING MODEL

Panel Structure

During the observation period, n policies have been in the portfolio, each one observed during T_i periods. Let N_{it} be the number of claims reported by policyholder i during year t, $i = 1, 2, \ldots, n$, $t = 1, 2, \ldots, T_i$. Let d_{it} be the length of this period. Usually, $d_{it} = 1$, but there is a variety of situations where this is not the case. Indeed, a new period of observation starts as soon as some policy characteristics are modified (think, for instance, of a moving of the policyholder for a company using postcode as rating factor, policyholder's wedding for a company using marital status, or simply the policyholder buying a new car). Moreover, in the year the policy is issued and in the one it is possibly canceled the length of the observation period is generally less than unity.

We face a nested structure: each policyholder generates a sequence $\boldsymbol{N}_i = (N_{i1}, N_{i2}, \ldots, N_{iT_i})$ of claim numbers. It is reasonable to assume independence between the series $\boldsymbol{N}_1, \boldsymbol{N}_2, \ldots, \boldsymbol{N}_n$, but this assumption is very questionable inside the $\boldsymbol{N}_i$'s (in fact, if the components of the $\boldsymbol{N}_i$'s were independent, *a posteriori* ratemaking would be senseless from the purely credibility point of view, even if these systems remain important because they counteract moral hazard and favor long-term commitment).

Regarding *a priori* ratemaking, the dependence between the components of each $\boldsymbol{N}_i$ is a nuisance. This point has been extensively treated by Denuit, Pitrebois, and Walhin (2003). Of course, when *a posteriori* corrections are of interest, the actuary has to focus on the serial dependence exhibited by the components of the $\boldsymbol{N}_i$'s.

Automobile insurance data have a panel structure: typically, n is large whereas the T_i's are small. Therefore, statistical methods appropriate for such data are particularly interesting, as the GEE approach developed by Liang and Zeger (1986), Zeger, Liang, and Albert (1988), and Zeger (1988). For an overview, see the book by Diggle, Liang, and Zeger (1994).

A Priori Risk Classification

The idea now is to incorporate in the N_{it}'s exogenous information (like age, gender, power of the car, and so on) summarized in the vectors $\boldsymbol{x}_{it}$; to this end, we resort to a regression model. The nonnegative integer nature of the annual number of claims implies that regression models based on continuous functional forms are inappropriate. While the general linear model assumes that the dependent variable is continuously distributed along the real line, claim frequencies are both censored at 0 and discrete. A statistical procedure that fails to account for this count data nature of the dependent variable can generate impossible predicted values, resulting in biased or inconsistent parameter estimates and invalid inferences.

The distributional assumption for the random component of the regression model has to account for the nonnegativity of the data, as well as their integer valuedness. We begin with Poisson regression and assume that the N_{it}'s conform to the Poisson distribution. The typical assumption in these circumstances is that the conditional mean of the N_{it}'s can be written as an exponential function of a linear combination $\boldsymbol{\beta}^t \boldsymbol{x}_{it}$ of the explanatory variables $\boldsymbol{x}_{it}$, with unknown regression coefficients $\boldsymbol{\beta}$ to be estimated from the data. This specification implies that the conditional mean of the dependent variable is log-linear and ensures that the parameter is nonnegative. This

yields a multiplicative tariff structure (which is particularly desirable for practical implementation).

Despite its prevalence as a starting point in the analysis of count data, the Poisson specification is often inappropriate because of unobserved heterogeneity and failure of the independence assumption if the data consist of repeated observations of the same policyholders. Unobserved heterogeneity and serial dependence will often cause the variance to exceed the mean (a phenomenon termed overdispersion). Failing to account for overdispersion yields underestimated standard errors, thereby conveying erroneously high levels of significance. Most often, this is due to the fact that important explanatory variables may not have been measured and are consequently incorrectly excluded from the regression relationship. A convenient way to take this phenomenon into account is to introduce a random effect in this model. This is the classical credibility construction.

Static Credibility Model

Traditionally, actuaries considered that heterogeneity did not vary with time. Every policy is thus affected by a risk parameter that can be interpreted as an error term in the regression model. For policyholder i, the risk parameter Θ_i represents the influence of unknown risk characteristics having a significant impact on the occurrence of claims. The ith policy of the portfolio, $i = 1, 2, \ldots, n$, is then represented by a sequence $(\Theta_i, \boldsymbol{N}_i)$ where Θ_i is a positive random variable with unit mean representing the unexplained heterogeneity. Specifically, the classical model introduced by Dionne and Vanasse (1989) is based on the following assumptions:

A1 given $\Theta_i = \theta$, the random variables N_{it}, $t = 1, 2, \ldots, T_i$, are independent and conform to the Poisson distribution with mean $\lambda_{it}\theta$, i.e.,

$$\Pr[N_{it} = k \mid \Theta_i = \theta] = \exp(-\lambda_{it}\theta)\frac{(\lambda_{it}\theta)^k}{k!}, \quad k \in \mathbb{N},$$

with $\lambda_{it} = d_{it}\exp(\boldsymbol{\beta}^t \boldsymbol{x}_{it})$;

A2 at the portfolio level, the Θ_i's are assumed to be independent and to admit the same distribution function $U(\cdot)$ called the structure function;

A3 the sequences $(\Theta_i, \boldsymbol{N}_i)$, $i = 1, 2, \ldots, n$, are assumed to be independent.

It is essential to understand the philosophy of the classical actuarial construction A1–A3. Here dependence between annual claim numbers $N_{i1}, N_{i2}, \ldots$ is a consequence of the heterogeneity of the portfolio: the dependence is only apparent. If we had a complete knowledge of policy characteristics then Θ_i would become deterministic and there would no longer be dependence between the N_{it}'s for fixed i. The unexplained heterogeneity is thus revealed by the claims and premiums' histories in a Bayesian way. These histories modify the distribution of Θ_i, and hence modify the premium.

Dynamic Credibility Model

In the classical credibility construction A1–A3 recalled above, the risk parameter Θ_i relating to policyholder i has been assumed constant over time. This is, of course,

rather unrealistic since driving ability may vary during the driving career (because of the learning effect, or a modification in the risk characteristics). In this article, we will assume that the unknown characteristics relating to policyholder i during year t are represented by a positive random variable Θ_{it}. The annual numbers of claims $N_{i1}, N_{i2}, \ldots, N_{iT_i}$ are assumed to be independent given the sequence $\boldsymbol{\Theta}_i = (\Theta_{i1}, \Theta_{i2}, \ldots, \Theta_{iT_i})$ of random effects. The latent unobservable vector $\boldsymbol{\Theta}_i$ characterizes the correlation structure of the N_{it}'s.

Now, the ith policy of the portfolio, $i = 1, 2, \ldots, n$, is represented by a double sequence $(\boldsymbol{\Theta}_i, \boldsymbol{N}_i)$ where $\boldsymbol{\Theta}_i$ is a positive random vector with unit mean representing the unexplained heterogeneity. Specifically, the model is based on the following assumptions:

B1 given $\boldsymbol{\Theta}_i = \boldsymbol{\theta}_i$, the random variables N_{it}, $t = 1, 2, \ldots, T_i$, are independent and conform to the Poisson distribution with mean $\lambda_{it}\theta_{it}$, i.e.,

$$\Pr[N_{it} = k \mid \Theta_{it} = \theta_{it}] = \exp(-\lambda_{it}\theta_{it})\frac{(\lambda_{it}\theta_{it})^k}{k!}, \quad k \in \mathbb{N},$$

with $\lambda_{it} = d_{it}\exp(\boldsymbol{\beta}^t \boldsymbol{x}_{it})$;

B2 at the portfolio level, the $\boldsymbol{\Theta}_i$'s are assumed to be independent. Moreover, $(\Theta_{i1}, \ldots, \Theta_{iT_i})$ is distributed as $(\Theta_1, \ldots, \Theta_{T_i})$ where $\boldsymbol{\Theta} = (\Theta_{i1}, \ldots, \Theta_{iT_{\max}})$ is a stationary random vector (with $T_{\max} = \max T_i$);

B3 the sequences $(\boldsymbol{\Theta}_i, \boldsymbol{N}_i)$, $i = 1, 2, \ldots, n$, are assumed to be independent.

It is further assumed that each $\boldsymbol{\Theta}_i$ is generated by a nonnegative stationary time series with $\mathbb{E}[\Theta_{it}] = 1$ for all i, t. The unit mean condition is imposed for identification (otherwise, the mean could be absorbed into the intercept term of the Poisson regression). This condition means that the *a priori* ratemaking is correct on average.

A fundamental difference between static A1–A3 and dynamic B1–B3 credibility models is that the latter incorporates the age of the claims in the risk prediction, whereas the former neglects this information. Since we intuitively feel that the predictive ability of a claim should decrease with its age, dynamic specification seems more acceptable. As pointed out by Pinquet, Guillén, and Bolancé (2001), dynamic credibility models agree with the economic analysis of multiperiod optimal insurance under moral hazard. In this view, the stationarity of the $\boldsymbol{\Theta}_i$'s implies that the predictive ability of claims depends mainly on the lag between the date of prediction and the date of occurrence (because of time translation invariance of the marginals of the $\boldsymbol{\Theta}_i$'s).

AR(1) Poisson-LogNormal Distribution

The multivariate LogNormal mixture of independent Poisson distributions is a parametric class of multivariate count distributions supporting a rich correlation structure. In our context, the multivariate Poisson-LogNormal law is a natural candidate to describe the joint distribution of the $\boldsymbol{N}_i$'s. We refer the reader to Appendix A for a complete description of the model.

A simple and powerful time series model for count data is obtained by specifying a Gaussian autoregressive process of order one (AR(1), in short) for the Θ_{it}'s on the

log-scale, that is,

$$\ln \Theta_{it} = \varrho \ln \Theta_{i,t-1} + \epsilon_{it}, \quad t = 2, 3, \ldots,$$

where $|\varrho| < 1$ and the ϵ_{it}'s are independent Gaussian errors with mean $(\varrho - 1)\frac{\sigma^2}{2}$ and variance $\sigma^2(1 - \varrho^2)$ to ensure that the $\ln \Theta_{it}$'s remain $\mathcal{N}(-\frac{\sigma^2}{2}, \sigma^2)$ distributed. Moreover, $\ln \Theta_{i1}$ conforms to the stationary $\mathcal{N}(-\frac{\sigma^2}{2}, \sigma^2)$ distribution.

This approach has been followed, for example, by Chan and Ledolter (1995) in analyzing the well-known polyo incidence data first studied by Zeger (1988). For a reference in actuarial science, see Pinquet, Guillén, and Bolancé (2001).

In the AR(1) Poisson-LogNormal case,

$$\sigma(h) = \mathbb{C}\text{ov}[\ln \Theta_{it}, \ln \Theta_{it+h}] = \sigma^2 \varrho^h,$$

so that the autocorrelation of the Θ_{it}'s is given by

$$\rho_\Theta(h) = \mathbb{C}\text{orr}[\Theta_{it}, \Theta_{it+h}] = \frac{\exp(\sigma^2 \varrho^h) - 1}{\exp(\sigma^2) - 1}.$$

Since ρ_Θ features the temporal dependence, the AR(1) Poisson-LogNormal model postulates that the predictive power of the claims decreases exponentially with their age.

The decreasing nature of ρ_Θ is in line with the exogenous interpretation of residual heterogeneity (generated by unobservable risk characteristics). The correlation between rating factors relating to different periods should decrease with the lag, for observable variables as well as for hidden features. Similarly, endogeneous components of latent processes (e.g., effort variables modeling moral hazard) yield the same conclusion: a driver will tend to be more prudent after an accident but this effect certainly decreases with the age of the claim.

Observed Covariate Process

We assume that *a priori* ratemaking has led to a partition of the portfolio into K risk classes $C_1, C_2, \ldots, C_K$. To each of these classes is attached a frequency premium λ_i (which is the expected annual number of claims for a policyholder in class C_i at the beginning of the year).

Since the partition of policyholders into $C_1, C_2, \ldots, C_K$ cannot account for the total heterogeneity present in the data, each policyholder is described by observable characteristics and a positive vector of random effects $\mathbf{\Theta}_i$ (representing hidden features influencing the occurrence of the covered peril) in accordance with assumptions B1–B3.

Now, observable characteristics of policyholders may evolve over time. Therefore, we denote as $\mathcal{I}_i = \{I_i(t),\ t = 1, 2, \ldots\}$ the random process describing the trajectory of the policyholder across risk classes $C_1, C_2, \ldots, C_K$; $I_i(t)$ represents the index of the risk class occupied by policyholder i at the beginning of year t. Thus, policyholder i

will be in class $C_{I_i(t)}$ during time interval $[t-1, t)$, $t = 1, 2, \ldots$. At the portfolio level, the $\mathcal{I}_i$'s are assumed to be independent and identically distributed. Moreover, each $\mathcal{I}_i$ is modeled as a homogenous Markov chain, independent of $\boldsymbol{\Theta}_i$, with transition probabilities $q_{j_1 j_2} = \Pr[I_i(t+1) = j_2 \mid I_i(t) = j_1]$.

It is worth mentioning that the assumption of independence between $\mathcal{I}_i$ and $\boldsymbol{\Theta}_i$ might be questionable. This apparent contradiction is solved if $\boldsymbol{\Theta}_i$ is seen as allowing for a residual heterogeneity (and is therefore independent of observable explanatory variables); see Pinquet (2000, p. 465) for more details on this issue.

Premium Amounts

The amount of premium charged to policyholder i in year t should be

$$\lambda_{I_i(t)}\Theta_{it}. \tag{1}$$

Unfortunately, Θ_{it} is unknown to the actuary. Hence, the use of past claims to reveal the Θ_{it}'s in a Bayesian way, that is, Equation (1), is replaced with

$$\underbrace{\lambda_{I_i(t)}}_{\text{Base premium}} \times \underbrace{\mathbb{E}[\Theta_{it} \mid \mathcal{H}_{i,t-1}]}_{\text{theoretical bonus-malus coefficient}},$$

where $\mathcal{H}_{i,t-1}$ gathers the claim history of the policyholder up to (and including) time $t-1$ (i.e., $N_{i1}, \ldots, N_{i,t-1}$) together with the observable characteristics for periods 1 to t (i.e., $I_i(1), \ldots, I_i(t-1)$). At this stage, it is interesting to compare our model to Taylor's:

1. we explicitly break the premium relating to year t into its *a priori* $\lambda_{I_i(t)}$ component and the heterogeneity component Θ_{it} to be corrected by the bonus-malus scale;
2. the individual's true underlying risk premium is not bound to remain constant but may vary with time because of modification in both observable and hidden characteristics.

Computing the Relativities

Now, let $\mathcal{I}$ describe the evolution of observable characteristics for a single policyholder chosen at random from the portfolio. His accident proneness is represented by the multivariate risk parameter $\boldsymbol{\theta}$ and is regarded as the outcome of a positive random vector $\boldsymbol{\Theta}$. Let us denote L_t as the level of the bonus-malus scale occupied by this policyholder in year t. The relativity $r(t, L_t)$ is the bonus-malus scale analog of the theoretical bonus-malus coefficient $\mathbb{E}[\Theta_t \mid \mathcal{H}_{t-1}]$. Unlike its theoretical counterpart, the relativity does not depend on observable characteristics. At this stage, the relativities still depend on the number t of years the policy is in force. This will be integrated further to produce time-independent relativities.

The predictive accuracy criterion is classically used to determine the premium relativities; it is a measure of the efficiency of a bonus-malus system. The idea behind this notion is as follows: a bonus-malus system is good at discriminating among the good and the bad risks if the premium they pay is close to their "true" premium. In

accordance with the predictive accuracy criterion, the relative premium associated with level L_t and year t is determined to minimize the expected squared difference between the relativity $r(t, L_t)$ and the true bonus-malus coefficient, that is, $\mathbb{E}[(r(t, L_t) - \Theta_t)^2]$. The solution of this minimization problem is the conditional expectation of Θ_t given $r(t, L_t)$, which corresponds to

$$r(t, j) = \mathbb{E}[\Theta_t \mid L_t = j]. \tag{2}$$

Of course, it is not convenient for commercial purposes to have a relativity depending on time since policy issuance. Therefore, the $r(t, j)$'s are averaged through time to produce the relativities r_j's, which is precisely our goal in this article.

The analytical determination of the r_j's clearly requires a huge amount of computation and complicated high-dimensional numerical integrations. For portfolios rated on the basis of numerous *a priori* variables, only a simulation-based approach is tractable.

NUMERICAL ILLUSTRATIONS

Data Description

The data relate to the automobile portfolio of a major insurance company operating in Spain. We only considered private use cars in this sample. The panel data contain information from 1991 to 1998. Our sample contains 80,994 policyholders who stayed in the portfolio for seven complete yearly periods. It is true that this sample does not represent the real portfolio because some form of survivor bias may have been introduced. We have not included here policyholders who canceled their policies because the procedures would need to account for duration. The illustrations given below can be accomodated to this circumstance, but we have not considered this at this moment.

We have seven exogenous variables that are kept in the panel plus the yearly number of accidents: gender of the policyholder, driving zone (urban–rural), risk associated with the driving zone (high–not high), coverage of the policy, age of the policy, age of the policyholder, and power of the car. This information is coded with the help of eight binary variables, described in Table 1. For every policy we have the initial

TABLE 1
Binary Coding of Explanatory Variables

Variable	Definition
Woman	Equals 1 for women and 0 for men
Urban	Equals 1 when driving in urban area, 0 otherwise
North	Equals 1 when zone is high risk (Northern Spain)
Coverage	Equals 1 if coverage includes comprehensive except fire
Client 1	Equals 1 if the client has been in the company between 3 and 5 years
Client 2	Equals 1 if the client is in the company for more than 5 years
Young	Equals 1 if the insured is 30 years old or younger
Power	Equals 1 if vehicle power is larger than or equal to 5,500 cc

information at the beginning of the period and the total number of claims at fault that took place within this yearly period.

A Priori Ratemaking With GEEs

In order to estimate the model parameters, a direct likelihood maximization is extremely difficult since the likelihood has no explicit form. It is worth mentioning that methods have been developed for Poisson-LogNormal count processes (as the Monte Carlo approach of Chan and Ledolter (1995)), but seem to be out of reach for the very large data sets the actuaries have to treat. Therefore, we prefer a GEE approach, which is in line with the Poisson-LogNormal model, except that it is only based upon first- and second-order moments. Since all estimators are consistent, we do not expect much difference in practice between the GEE and ML estimations. The advantage of the GEE approach over the classical Poisson regression (assuming serial independence, where estimators are also consistent) is that the standard errors of the regression parameters are corrected, as pointed out in Denuit, Pitrebois, and Walhin (2003). We refer the reader to Appendix B for a brief presentation of the GEE approach in the Poisson-LogNormal model.

We fitted the AR(1) Poisson-LogNormal distribution to our data. The estimated parameters are displayed in Table 2. The Poisson-LogNormal estimates with serial independence are displayed in the first place. Then the estimates that take the AR(1) covariance structure into account are also presented. We stopped the iterative procedure after four steps, once a 6-decimal places stability had been reached.

Estimation of the Transition Probabilities of the $\mathcal{I}_i$'s

For each policyholder, we have at our disposal the sequence $C_{I_i(1)}, C_{I_i(2)}, \ldots, C_{I_i(T_i)}$ of risk classes he occupied during the observation period. If we denote $q_{k_1k_2}$ as the probability of moving from C_{k_1} to C_{k_2}, the likelihood associated to policyholder i is

$$q_{I_i(1)I_i(2)} q_{I_i(2)I_i(3)} \cdots q_{I_i(T_i-1)I_i(T_i)}.$$

Among the explanatory variables we have at our diposal, the gender variable is obviously fixed, variables related to age and number of years since the client entered the company have a deterministic behavior. The other variables are stochastic. Since their behavior is almost not influenced by the others, we have opted for independent two-state stationary Markov chains describing the evolution of the following variables: Urban, North, Coverage, and Power. The transition probabilites estimated from the data are presented in Table 3.

As the probability of change is extremely small for variable North, we will consider it as fixed in the simulations.

TABLE 2
Parameter Estimates for the Dynamic Credibility Models

Model	Constant	Woman	Urban	North	Coverage	Client 1	Client 2	Young	Power	$\hat{\sigma}^2_\Theta$	$\hat{\varrho}$
Indep.	−2.6542	0.0692	−0.0503	0.1968	0.1345	−0.1490	−0.2266	0.2080	0.1179	1.3945	0.5250
AR(1)	−2.6576	0.0742	−0.0511	0.1974	0.1340	−0.1470	−0.2276	0.2091	0.1211	1.3951	0.5251

TABLE 3
Estimated Transition Probabilities Between States for Variables Changing Over Time

Variable	q_{00}	q_{01}	q_{10}	q_{11}
Urban	0.9938	0.0062	0.0153	0.9847
North	>0.999	<0.001	<0.001	>0.999
Coverage	0.9797	0.0203	0.1128	0.8872
Power	0.9014	0.0986	0.0066	0.9934

Determination of the Relativities

In this section, we work with the mild and severe scales of Taylor (1997). As explained above, a simulation approach is the only tractable solution to the problem. This computer-intensive method has the great advantage of allowing for tailor-made solutions. The actuary may opt for the most relevant hypotheses and end up with a result in accordance with the situation he faces. In our case, we assume that the bonus-malus scale will be proposed to new policyholders and that these persons will provide the company at policy issuance with a summary of their past driving behavior. As in many European countries, the information is supposed to consist of the claims at fault reported during the last 10 years. The company places the individuals in their levels according to past claim history. Moreover, the system is required to work for an initial period of 10 years.

We build the simulation as follows: to each level ℓ we associate the counters $\lambda_\bullet(\ell)$, $\theta_\bullet(\ell)$ and $\delta_\bullet(\ell)$:

1. $\lambda_\bullet(\ell)$ sums the values of the λ_{it}'s relating to policyholders occupying level ℓ at some time t;
2. $\theta_\bullet(\ell)$ sums the values of the θ_{it}'s relating to policyholders occupying level ℓ at some time t;
3. $\delta_\bullet(\ell)$ measures the occupation of level ℓ (in policyholders-years), namely the exposure.

The trajectory of policyholders is simulated as follows. The policyholder i is placed in level 6 and an initial risk class $C_{I_i(1)}$ is selected according to the profiles of new policyholders entering the portfolio of the company. The risk class ($C_{I_i(1)}$) fully determines the characteristics of the policyholder, except his precise age. Indeed, variable Young only indicates if the insured is 30 years old or younger. His age in years is then simulated from the empirical age distribution of new policyholders, taking into account the value of variable Young. The value of $\ln\theta_{i1}$ is drawn from the $\mathcal{N}(-\frac{1}{2}\sigma^2, \sigma^2)$ distribution. Then, the 10-year past claims record of the policyholder is generated from the $\mathcal{P}(\lambda_{I_i(1)}\theta_{i1})$ distribution, without updating risk class and heterogeneity component. The policyholder is placed at the level he would have attained were he in the system for 10 years (starting from level 6). Then, we start following the trajectory of the policyholder. The number n_{i1} of claims is taken from $\mathcal{P}(\lambda_{I_i(1)}\theta_{i1})$ and the policyholder is moved to another level (ℓ, say) according to the mild bonus-malus scale's rules given in the section "European and Asian Bonus-Malus Scales." Then,

the risk class $C_{I_i(2)}$ for the second coverage period is drawn from the distribution $\{q_{I_i(1)k},\ k = 1, \ldots, K\}$. The value of $\ln \theta_{i2}$ is taken as $\varrho \ln \theta_{i1} + \epsilon_{i1}$ where ϵ_{i1} conforms to $\mathcal{N}(-\frac{1}{2}(1-\varrho)\sigma^2, (1-\varrho^2)\sigma^2)$. All counters are updated. The number n_{i2} of claims is taken from $\mathcal{P}(\lambda_{I_i(2)}\theta_{i2})$ and the policyholder is moved to another level, according to the rules of the mild bonus-malus system.

The trajectory is so simulated during 10 years and for 100,000 policyholders. Finally, we compute the averages

$$\bar{\lambda}(\ell) = \frac{\lambda_\bullet(\ell)}{\delta_\bullet(\ell)} \quad \text{and} \quad \bar{\theta}(\ell) = \frac{\theta_\bullet(\ell)}{\delta_\bullet(\ell)},$$

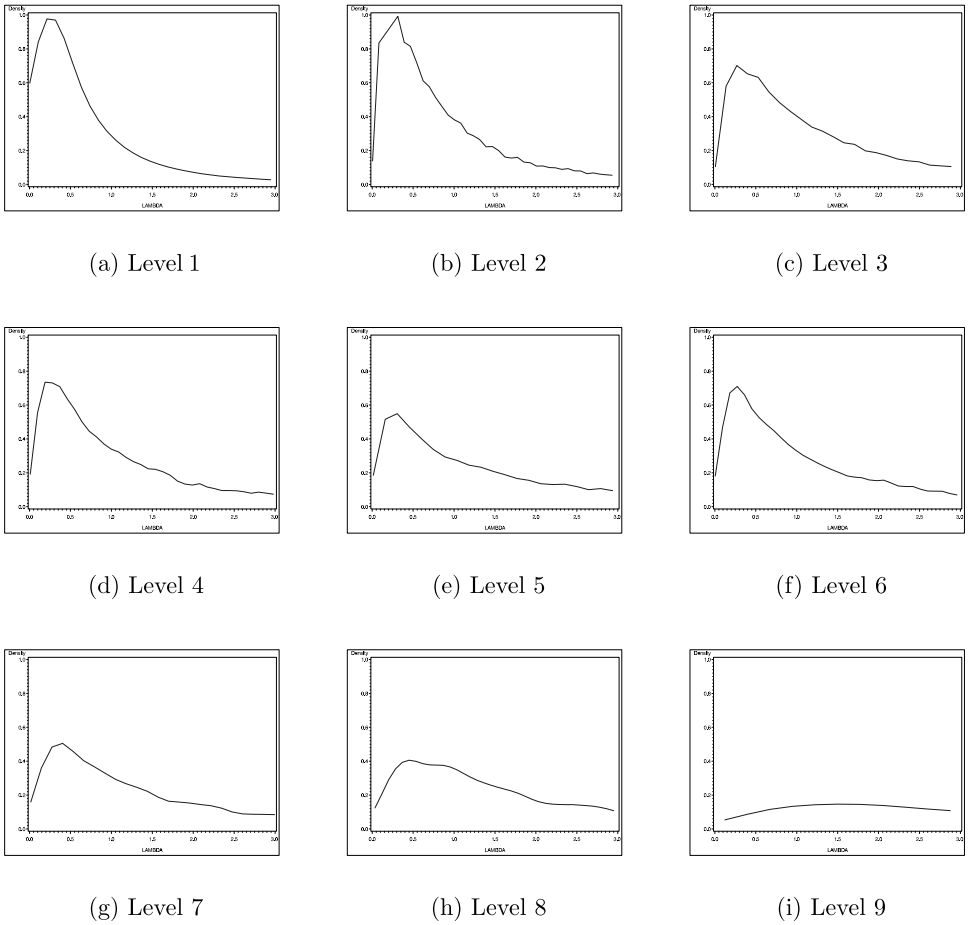
which correspond to the mean frequency premium and the mean heterogeneity, respectively. These results and the relativities, r_l, are displayed in Table 4. We clearly see in Table 4 that the average *a priori* premium amount paid by policyholders at each level of the scale (mean frequency premium) increases with bonus-malus level. This indicates that in case the premium system effectively uses the different rating variables, then bonus-malus levels with low average claim frequencies will tend to have low base premiums too. This means that the two rating mechanisms interact and thus cannot be considered in isolation. In other words, the relativities attached to each bonus-malus level must take *a priori* ratemaking into account. The aim is that the bonus-malus relativities discriminate between good and bad drivers only to the extent that this differentiation is not already recognized in the base premium. This is precisely the reason to compute relativities according to Equation (2).

Let us now examine the conditional distribution of Θ given $L_t = j$. Figure 1 proposes a kernel estimation of the structure function for each level of the scale. The higher the level, the more dispersed the structure function. Considering the mean heterogeneity of each level, we observe a jagged behavior with increases in the uneven levels 3–5–7–9 of the scale. This is due to the transition rules of the mild bonus-malus system, which induces a two-level penalty for each claim. It is worth mentioning at this stage that the

TABLE 4
Average Measures and Relativities for the Mild Bonus-Malus System in the Dynamic Credibility Model, With *a Priori* Risk Classification

Level ℓ	Mean Heterogeneity $\bar{\theta}(\ell)$	Mean Frequency Premium $\bar{\lambda}(\ell)$	Exposure $\delta_\bullet(\ell)$	Adjusted Heterogeneity $\bar{\theta}^*(\ell)$	Relativity r_ℓ
1	0.7755	0.08206	799,735	0.7730	0.2657
2	1.1457	0.08320	58,576	1.2003	0.4126
3	1.8900	0.08361	67,254	1.6275	0.5594
4	1.5449	0.08459	23,444	2.0548	0.7063
5	2.6465	0.08481	19,850	2.4820	0.8531
6	1.8163	0.08583	11,188	2.9093	1
7	2.8813	0.08651	9,549	3.3366	1.1469
8	2.5411	0.08748	5,937	3.7638	1.2937
9	7.7932	0.08822	4,467	4.1911	1.4406

FIGURE 1
Estimation of the Structure Function for Each Level of the Mild Bonus-Malus Scale



(a) Level 1 (b) Level 2 (c) Level 3

(d) Level 4 (e) Level 5 (f) Level 6

(g) Level 7 (h) Level 8 (i) Level 9

relativities in the bonus-malus level are not guaranteed to increase even for the Bayes scales built in Norberg's (1976) framework. To overcome this problem, we follow the idea of Gilde and Sundt (1989) and we smooth the obtained $\bar{\theta}(\ell)$'s by applying a linear regression (weighted by the $\delta_\bullet(\ell)$'s) on the data $(\ell, \bar{\theta}(\ell))$. We then substitute the fitted $\bar{\theta}^*(\ell)$ for the original $\bar{\theta}(\ell)$'s. Finally, since the $\bar{\theta}^*(\ell)$'s do not match the constraint $r_6 = 1$, we normalize them to get the values of the last column of Table 4.

Comparison With Other Approaches for the Mild Bonus-Malus System

In this section, we compare our results to those obtained with other approaches. First of all, we consider relativities in the static credibility model A1–A3 in the spirit of Taylor (1997). This model can be seen as a particular case of the dynamic one with $\varrho = 1$. The estimated coefficients $\hat{\beta}$ and the residual variance $\hat{\sigma}^2_\Theta$ are displayed in Table 5.

As in the dynamic case, convergence occurred after four steps. The resulting relativities are displayed in Table 6. Comparing Tables 4 and 6 yields the following remarks.

TABLE 5
Parameter Estimates for the Static Credibility Models

Model	Constant	Woman	Urban	North	Coverage	Client 1	Client 2	Young	Power	$\hat{\sigma}^2_{\Theta}$
Indep.	−2.6542	0.0692	−0.0503	0.1968	0.1345	−0.1490	−0.2266	0.2080	0.1179	1.3945
GEE	−2.6557	0.0715	−0.0509	0.1971	0.1344	−0.1490	−0.2273	0.2095	0.1200	1.3939

TABLE 6
Averages and Relativities for the Mild Bonus-Malus System in the Static Credibility Model, With *a Priori* Risk Classification

Level ℓ	Mean Heterogeneity $\bar{\theta}(\ell)$	Mean Frequency Premium $\bar{\lambda}(\ell)$	Exposure $\delta_{\bullet}(\ell)$	Adjusted Heterogeneity $\bar{\theta}^*(\ell)$	Relativity r_ℓ
1	0.6122	0.08226	821,862	0.5681	0.1321
2	1.3647	0.08336	45,151	1.3148	0.3057
3	1.4848	0.08369	51,404	2.0616	0.4792
4	2.3688	0.08453	18,245	2.8083	0.6528
5	2.7566	0.08442	15,068	3.5550	0.8264
6	3.6032	0.08460	11,059	4.3017	1
7	4.3365	0.08482	10,968	5.0484	1.1736
8	5.7266	0.08505	11,096	5.7951	1.3472
9	8.3573	0.08581	15,147	6.5418	1.5208

In the static case, the $\bar{\theta}(\ell)$'s are more dispersed and now increase monotonically with the level ℓ. It is not clear whether this increase is a general consequence of the time invariance of the random effects. Comparing the occupancy of each level, we see that the highest levels are visited more often in the static case than in the dynamic one. Quite surprisingly, the impact of letting random effects evolve over time is rather important. Even if it was not necessary, we also linearized the $\bar{\theta}(\ell)$'s and computed associated relativities, for the sake of comparison with the dynamic credibility framework. Resulting relativities are displayed in the last column of Table 6; they show that the static credibility model grants more discounts and induces more severe penalties than its dynamic B1–B3 counterpart.

We also determine the relativities resulting from disregarding *a priori* ratemaking. In this case, N_{it} follows the $\mathcal{P}(\lambda\Theta_{it})$ distribution. On the basis of our data set, the parameters were estimated at

$$\hat{\lambda} = 0.06949, \quad \hat{\sigma}^2_{\Theta} = 1.4632, \quad \text{and} \quad \hat{\varrho} = 0.5499.$$

The results are presented in Table 7. The $\bar{\lambda}(\ell)$'s are now all equal, since no *a priori* risk classification is achieved. Again, irregular behavior of the $\bar{\theta}(\ell)$'s is apparent: levels 3–5–7–9 have higher $\bar{\theta}(\ell)$'s, due to the two-step penalty of the mild bonus-malus scale. The difference $\bar{\theta}(9) - \bar{\theta}(1)$ becomes larger when the *a priori* ratemaking is not taken into account, as expected. The same remark applies to the linearized $\bar{\theta}^*(\ell)$'s as well as to the resulting relativities r_ℓ's.

TABLE 7
Averages and Relativities for the Mild Bonus-Malus System in the Dynamic Credibility Model, Without *a Priori* Risk Classification

Level ℓ	Mean Heterogeneity $\bar{\theta}(\ell)$	Mean Frequency Premium $\bar{\lambda}(\ell)$	Exposure $\delta_\bullet(\ell)$	Adjusted Heterogeneity $\bar{\theta}^*(\ell)$	Relativity r_ℓ
1	0.7897	0.06949	835,967	0.7833	0.2264
2	1.2442	0.06949	51,095	1.3184	0.3812
3	2.0848	0.06949	58,234	1.8536	0.5359
4	1.7238	0.06949	17,987	2.3887	0.6906
5	3.0167	0.06949	15,108	2.9239	0.8453
6	2.1400	0.06949	8,061	3.4590	1
7	3.5614	0.06949	6,697	3.9942	1.1547
8	3.1568	0.06949	3,919	4.5293	1.3094
9	9.9815	0.06949	2,932	5.0645	1.4641

TABLE 8
Averages and Relativities for the Mild Bonus-Malus System in the Static Credibility Model, Without *a Priori* Risk Classification

Level ℓ	Mean Heterogeneity $\bar{\theta}(\ell)$	Mean Frequency Premium $\bar{\lambda}(\ell)$	Exposure $\delta_\bullet(\ell)$	Adjusted Heterogeneity $\bar{\theta}^*(\ell)$	Relativity r_ℓ
1	0.6400	0.06949	851,558	0.5952	0.1178
2	1.5248	0.06949	40,707	1.4865	0.2943
3	1.6635	0.06949	46,196	2.3778	0.4707
4	2.7551	0.06949	14,811	3.2691	0.6471
5	3.1882	0.06949	11,996	4.1603	0.8236
6	4.2327	0.06949	8,343	5.0516	1
7	5.1375	0.06949	7,904	5.9429	1.1764
8	6.8669	0.06949	7,860	6.8342	1.3529
9	10.1311	0.06949	10,625	7.7255	1.5293

For the sake of completeness, we finally compute the relativities obtained without a priori ratemaking and in the static credibility model, that is when N_{it} is assumed to obey to $\mathcal{P}(\lambda\Theta_i)$. We obtained

$$\hat{\lambda} = 0.06949 \quad \text{and} \quad \hat{\sigma}^2_\Theta = 1.4632.$$

The results are presented in Table 8. Comparing Tables 6 and 8, we see that disregarding *a priori* risk classification yields more dispersed $\bar{\theta}(\ell)$'s. Considering Tables 7 and 8, we again see that the impact of introducing dynamic heterogeneity is nonnegligible. Finally, it is worth mentioning that the $\bar{\theta}(\ell)$'s are monotonic in the level ℓ.

Comparison of the Mild and the Severe Bonus-Malus Systems

This section aims to examine the consequences of a modification in the transition rules. Specifically, we now consider a four-step penalty for each claim. The comparative results are presented in Tables 9–11 .

As expected, if the penalties applied to policyholders reporting claims are increased, the discounts granted to policyholders filing no claims become more important. This is apparent from the $\bar{\theta}(\ell)$'s associated to low levels ℓ. On the contrary for high levels ℓ, the $\bar{\theta}(\ell)$'s are higher for the mild system than for the severe one. The same remark applies to the linearized $\bar{\theta}^*(\ell)$'s and to the resulting relativities r_ℓ's.

TABLE 9
Comparison of $\bar{\theta}(\ell)$'s for the Mild and Severe Bonus-Malus Systems

	Mild Bonus-Malus System				Severe Bonus-Malus System			
	With *a Priori*		Without *a Priori*		With *a Priori*		Without *a Priori*	
Level	Dynamic	Static	Dynamic	Static	Dynamic	Static	Dynamic	Static
1	0.7755	0.6122	0.7897	0.6400	0.7560	0.5311	0.7677	0.5569
2	1.1457	1.3647	1.2442	1.5248	0.8506	1.1443	0.8942	1.2689
3	1.8900	1.4848	2.0848	1.6635	0.9184	1.2343	0.9755	1.3695
4	1.5449	2.3688	1.7238	2.7551	1.1018	1.3441	1.1942	1.4847
5	2.6465	2.7566	3.0167	3.1882	1.7335	1.4753	1.9266	1.6238
6	1.8163	3.6032	2.1400	4.2327	1.1666	2.4142	1.2808	2.7625
7	2.8813	4.3365	3.5614	5.1375	1.3575	2.8647	1.4765	3.2627
8	2.5411	5.7266	3.1568	6.8669	1.8851	3.6008	2.1875	4.1238
9	7.7932	8.3573	9.9815	10.1311	4.9248	5.3051	5.6969	6.0814

TABLE 10
Comparison of $\bar{\theta}^*(\ell)$'s for the Mild and Severe Bonus-Malus Scales

	Mild Bonus-Malus System				Severe Bonus-Malus System			
	With *a Priori*		Without *a Priori*		With *a Priori*		Without *a Priori*	
Level	Dynamic	Static	Dynamic	Static	Dynamic	Static	Dynamic	Static
1	0.7730	0.5681	0.7833	0.5952	0.6970	0.4809	0.7144	0.5099
2	1.2003	1.3148	1.3184	1.4865	0.9467	0.9331	1.0069	1.0237
3	1.6275	2.0616	1.8536	2.3778	1.1964	1.3852	1.2995	1.5375
4	2.0548	2.8083	2.3887	3.2691	1.4460	1.8374	1.5920	2.0512
5	2.4820	3.5550	2.9239	4.1603	1.6957	2.2895	1.8845	2.5650
6	2.9093	4.3017	3.4590	5.0516	1.9454	2.7417	2.1770	3.0788
7	3.3366	5.0484	3.9942	5.9429	2.1951	3.1939	2.4695	3.5926
8	3.7638	5.7951	4.5293	6.8342	2.4448	3.6460	2.7620	4.1063
9	4.1911	6.5418	5.0645	7.7255	2.6944	4.0982	3.0545	4.6201

TABLE 11
Comparison of r_ℓ's for the Mild and Severe Bonus-Malus Systems

	Mild Bonus-Malus System				Severe Bonus-Malus System			
	With *a Priori*		Without *a Priori*		With *a Priori*		Without *a Priori*	
Level	Dynamic	Static	Dynamic	Static	Dynamic	Static	Dynamic	Static
1	0.2657	0.1321	0.2264	0.1178	0.3583	0.1754	0.3282	0.1656
2	0.4126	0.3057	0.3812	0.2943	0.4866	0.3403	0.4625	0.3325
3	0.5594	0.4792	0.5359	0.4707	0.6150	0.5052	0.5969	0.4994
4	0.7063	0.6528	0.6906	0.6471	0.7433	0.6702	0.7313	0.6662
5	0.8531	0.8264	0.8453	0.8236	0.8717	0.8351	0.8656	0.8331
6	1	1	1	1	1	1	1	1
7	1.1469	1.1736	1.1547	1.1764	1.1283	1.1649	1.1344	1.1669
8	1.2937	1.3472	1.3094	1.3529	1.2567	1.3298	1.2687	1.3338
9	1.4406	1.5208	1.4641	1.5293	1.3850	1.4948	1.4031	1.5006

CONCLUSIONS

This article proposes a computer-intensive method to calibrate bonus-malus scales. It complements the pioneering study of Taylor (1997) and clearly enlightens the strong complementarity of *a priori* and *a posteriori* ratemakings. The main originality of this article is to compare four different credibility models on the basis of real data: static versus dynamic heterogeneity, with and without recognizing *a priori* risk classification. The impact of the different assumptions becomes clear in the numerical illustrations.

Of course, the present article is only a first step toward real bonus-malus computations. In practice, commercial tariffs are likely to involve a much finer segmentation. Hence, the number K of risk classes becomes large and the matrix of transition probabilities $q_{j_1 j_2}$ becomes correspondingly large. Estimation of the $q_{j_1 j_2}$'s can then be problematic and it may become necessary to impose some parametric structures to these probabilities.

APPENDIX A

Multivariate Poisson-LogNormal Distribution

Definition. (This multivariate counting distribution has been considered by Aitchinson and Ho (1989); see also Joe (1997). According to this model, we have

$$\Pr[\boldsymbol{N}_i = \boldsymbol{n}_i] = \int_{\theta_{i1}\in\mathbb{R}^+} \cdots \int_{\theta_{iT_i}\in\mathbb{R}^+} \left\{ \prod_{t=1}^{T_i} \exp(-\theta_{it}\lambda_{it}) \frac{(\theta_{it}\lambda_{it})^{n_{it}}}{n_{it}!} \right\} f_{\boldsymbol{\Theta}_i}(\boldsymbol{\theta}_i \mid \boldsymbol{\Sigma})\, d\theta_{i1} \ldots d\theta_{iT_i},$$

where the density $f_{\boldsymbol{\Theta}_i}$ corresponds to the multivariate $\mathcal{LN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution, i.e.,

$$f_{\boldsymbol{\Theta}_i}(\boldsymbol{\theta}_i|\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{T_i/2}\theta_{i1}\ldots\theta_{iT_i}|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\ln\boldsymbol{\theta}_i - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1}(\ln\boldsymbol{\theta}_i - \boldsymbol{\mu})\right\},$$

with $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ the mean vector and covariance matrix of the Gaussian random vector $\ln \boldsymbol{\Theta}_i$. In our context of serial dependence, $\boldsymbol{\Sigma} = \{\sigma_{s,t}\}_{1 \le s,t \le T_i}$, $\sigma_{tt} = \sigma^2$, $\sigma_{s,t} = \sigma(|t-s|)$ for $s \neq t$, and $\boldsymbol{\mu} = -\frac{1}{2}\sigma^2 \mathbf{1}$. We conventionally put $\sigma(0) = \sigma^2$ for convenience. All the $\ln \Theta_{it}$'s conform to the same $\mathcal{N}(-\frac{1}{2}\sigma^2, \sigma^2)$ distribution.

First- and Second-Order Moments. Although the multiple integral giving $\Pr[\boldsymbol{N}_i = \boldsymbol{n}_i]$ cannot be simplified, the moments of $\boldsymbol{N}_i$ are easily obtained as a function of those of $\boldsymbol{\Theta}_i$. Let us recall that the moment generating function $m(t; \mu, \sigma^2)$ of the $\mathcal{N}(\mu, \sigma^2)$ distribution is $\exp(\mu t + \frac{t^2\sigma^2}{2})$. Therefore, the moments of $\boldsymbol{\Theta}_i$ are given by

$$\begin{aligned}
\mathbb{E}[\Theta_{it}] &= m(1; -\sigma^2/2, \sigma^2) = 1 \\
\mathbb{V}\mathrm{ar}[\Theta_{it}] &= \mathbb{E}\big[\Theta_{it}^2\big] - \{\mathbb{E}[\Theta_{it}]\}^2 = m(2; -\sigma^2/2, \sigma^2) - 1 = \exp(\sigma^2) - 1 = \sigma_\Theta^2 \\
\mathbb{C}\mathrm{ov}[\Theta_{it}, \Theta_{is}] &= \mathbb{E}[\Theta_{it}\Theta_{is}] - \mathbb{E}[\Theta_{it}]\mathbb{E}[\Theta_{is}] \\
&= m(1; -\sigma^2, 2\sigma^2 + 2\sigma(|t-s|)) - 1 = \exp(\sigma(|t-s|)) - 1.
\end{aligned}$$

Let us now compute the means and variances of the N_{it}'s:

$$\mathbb{E}[N_{it}] = \lambda_{it}\mathbb{E}[\Theta_{it}] = \lambda_{it} \tag{A1}$$

$$\begin{aligned}
\mathbb{V}\mathrm{ar}[N_{it}] &= \mathbb{E}[\mathbb{V}\mathrm{ar}[N_{it} \mid \Theta_{it}]] + \mathbb{V}\mathrm{ar}[\mathbb{E}[N_{it} \mid \Theta_{it}]] \\
&= \lambda_{it} + \lambda_{it}^2 \mathbb{V}\mathrm{ar}[\Theta_{it}] = \lambda_{it}\{1 + \lambda_{it}\{\exp(\sigma^2) - 1\}\}.
\end{aligned} \tag{A2}$$

The covariance between N_{it} and N_{is}, $s \neq t$, is given by

$$\begin{aligned}
\mathbb{C}\mathrm{ov}[N_{it}, N_{is}] &= \mathbb{E}[\mathbb{C}\mathrm{ov}[N_{it}, N_{is} \mid \boldsymbol{\Theta}_i]] + \mathbb{C}\mathrm{ov}[\mathbb{E}[N_{it} \mid \boldsymbol{\Theta}_i], \mathbb{E}[N_{is} \mid \boldsymbol{\Theta}_i]] \\
&= \lambda_{it}\lambda_{is}\mathbb{C}\mathrm{ov}[\Theta_{it}, \Theta_{is}] = \lambda_{it}\lambda_{is}\{\exp(\sigma(|t-s|)) - 1\}.
\end{aligned} \tag{A3}$$

From Equations (A1)–(A2), we see that the marginals of $\boldsymbol{N}_i$ have overdispersion relative to the Poisson distribution, as expected from any Poisson mixture. From Equation (3) we have a correspondence between positive and negative values of correlations for the random effects Θ_{it} and the counts N_{it}.

Autocorrelation Function of the Latent Process. Let us now introduce the autocorrelation function ρ_Θ, defined as

$$\rho_\Theta(h) = \mathbb{C}\mathrm{orr}[\Theta_{it}, \Theta_{i,t+h}] = \frac{\exp(\sigma(h)) - 1}{\exp(\sigma^2) - 1}, \quad h = 1, 2, \ldots.$$

This function features the temporal dependence between the random effects. If the predictive power of the claims decreases with their age, $\rho_\Theta(h)$ should decrease with h.

Computing

$$\begin{aligned}\mathbb{C}\text{orr}[N_{it}, N_{is}] &= \frac{\mathbb{C}\text{ov}[N_{it}, N_{is}]}{\sqrt{\mathbb{V}\text{ar}[N_{it}]\mathbb{V}\text{ar}[N_{is}]}} \\ &= \frac{\sqrt{\lambda_{it}\lambda_{is}}\{\exp(\sigma(|t-s|))-1\}}{\sqrt{1+\lambda_{it}\{\exp(\sigma^2)-1\}}\sqrt{1+\lambda_{is}\{\exp(\sigma^2)-1\}}} \quad \text{for } s \neq t\end{aligned}$$

shows that

$$\mathbb{C}\text{orr}[N_{it}, N_{it+h}] = \frac{\rho_\Theta(h)}{\sqrt{1+\frac{1}{\lambda_{it}\{\exp(\sigma^2)-1\}}}\sqrt{1+\frac{1}{\lambda_{it+h}\{\exp(\sigma^2)-1\}}}}, \quad h \geq 1.$$

APPENDIX B

Generalized Estimating Equations (GEEs)

For convenience, let us consider a panel data set (T observation periods for each policyholder) as in the numerical illustration we have in mind. Let us define the $T \times T$ diagonal matrix $\boldsymbol{A}_i$ with λ_{it} on the diagonal. We also introduce $\boldsymbol{S}_i = \boldsymbol{N}_i - \boldsymbol{\lambda}_i$ and $\boldsymbol{X}_i$ as part of the design matrix relating to policyholder i (i.e., $\boldsymbol{X}_i$ is of dimension $p \times T$, where p is the number of observed covariates, and has columns $\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{it}$). We denote as $\boldsymbol{\alpha}$ the vector of parameters relating to the dependence structure; we chose

$$\alpha_h = \mathbb{C}\text{ov}[\Theta_t, \Theta_{t+h}], \quad h = 0, 1, \ldots,$$

but other parameterizations are of course possible. Now, it is clear from Equation (A1) that the expectation $\boldsymbol{\lambda}_i$ of $\boldsymbol{N}_i$ only depends on $\boldsymbol{\beta}$ (and not on $\boldsymbol{\alpha}$). Let us denote as $\boldsymbol{V}_i$ the covariance matrix of $\boldsymbol{N}_i$, which depends on $\boldsymbol{\alpha}$.

It is well known that the maximum likelihood equation for $\boldsymbol{\beta}$ in the Poisson model with serial independence is

$$\sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{S}_i = \boldsymbol{0}. \tag{B1}$$

The solution $\hat{\boldsymbol{\beta}}^{\perp}$ of Equation (B1) remains consistent if the elements of $\boldsymbol{N}_i$ are correlated (provided the λ_{it}'s are correctly specified).

Let us now present estimating equations taking the correlation structure of the $\boldsymbol{N}_i$'s into account to increase efficiency. Since the resulting estimator of $\boldsymbol{\beta}$ remains consistent, we do not expect much difference for point estimation of the regression coefficients. Accounting for serial correlation inflates the standard errors of the β_j's, which is important, e.g., in the context of variable selection as well as to correctly assess the accuracy of the frequency premium (λ_{it}) estimates.

Let us define $\boldsymbol{D}_i = \boldsymbol{A}_i \boldsymbol{X}_i^t$ of dimension $T \times p$. We can view Equation (B1) as the sum of $\boldsymbol{D}_i^t \boldsymbol{A}_i^{-1} \boldsymbol{S}_i$, where A_i is the covariance matrix of $\boldsymbol{N}_i$ in the model with serial

independence. The idea is now to substitute the covariance matrix $\boldsymbol{V}_i$ of $\boldsymbol{N}_i$ in the Poisson mixture model to $\boldsymbol{A}_i$, and to estimate $\boldsymbol{\beta}$ by the solution of

$$\sum_{i=1}^{n} \boldsymbol{D}_i^t \boldsymbol{V}_i^{-1} \boldsymbol{S}_i = \boldsymbol{0}. \tag{B2}$$

Liang and Zeger (1986) showed that this estimator possesses the desirable property of consistency. Under suitable specification conditions, it is also efficient.

As suggested by these authors, to compute the solution of Equation (B2), we iterate between a modified Fisher scoring for $\boldsymbol{\beta}$ and moment estimation of $\boldsymbol{\alpha}$. Specifically, the iterative procedure for $\boldsymbol{\beta}_i$ is

$$\hat{\boldsymbol{\beta}}^{(j+1)} = \hat{\boldsymbol{\beta}}^{(j)} + \left\{ \sum_{i=1}^{n} \boldsymbol{D}_i^t \big(\hat{\boldsymbol{\beta}}^{(j)}\big) \boldsymbol{V}_i \big(\hat{\boldsymbol{\beta}}^{(j)}, \boldsymbol{\alpha}\big(\hat{\boldsymbol{\beta}}^{(j)}\big)\big) \boldsymbol{D}_i \big(\hat{\boldsymbol{\beta}}^{(j)}\big) \right\}^{-1} \times \left\{ \sum_{i=1}^{n} \boldsymbol{D}_i^t \big(\hat{\boldsymbol{\beta}}^{(j)}\big) \boldsymbol{V}_i \big(\hat{\boldsymbol{\beta}}^{(j)}, \boldsymbol{\alpha}\big(\hat{\boldsymbol{\beta}}^{(j)}\big)\big) \boldsymbol{S}_i \big(\hat{\boldsymbol{\beta}}^{(j)}\big) \right\}.$$

The starting value $\hat{\boldsymbol{\beta}}^{(0)}$ is taken to be the solution of Equation (B1), that is $\hat{\boldsymbol{\beta}}^{\perp}$. At each step, we update the estimation of $\boldsymbol{\alpha}$ by

$$\hat{\alpha}_0\big(\hat{\boldsymbol{\beta}}^{(j)}\big) = \frac{\sum_{i=1}^{n} \sum_{t=1}^{T} \left\{ \big(n_{it} - \hat{\lambda}_{it}^{(j)}\big)^2 - n_{it} \right\}}{\sum_{i=1}^{n} \sum_{t=1}^{T} \big\{\hat{\lambda}_{it}^{(j)}\big\}^2},$$

where $\hat{\lambda}_{it}^{(j)} = d_{it} \exp(\boldsymbol{x}_{it} \hat{\boldsymbol{\beta}}^{(j)})$, and

$$\hat{\alpha}_h\big(\hat{\boldsymbol{\beta}}^{(j)}\big) = \hat{\rho}_{\Theta}(h) \hat{\sigma}_{\Theta}^2 = \frac{\sum_{i=1}^{n} \sum_{t=h+1}^{T} \big(n_{it} - \hat{\lambda}_{it}^{(j)}\big)\big(n_{it-h} - \hat{\lambda}_{it-h}^{(j)}\big)}{\sum_{i=1}^{n} \sum_{t=h+1}^{T} \hat{\lambda}_{it}^{(j)} \hat{\lambda}_{it-h}^{(j)}} \quad \text{for } h = 1, \ldots, T-1.$$

It is worth mentioning that the repeated statement of the SAS procedure GENMOD cannot be applied (since the number of levels of the class variable indicating clusters corresponding to individual policyholders has too many levels in our case). Therefore, we had to program the GEE formulas in the IML environment of SAS. The code is available from the authors upon request.

REFERENCES

Aitchinson, J., and C. H. Ho, 1989, The Multivariate Poisson-Log Normal Distribution, *Biometrika*, 75: 621-629.

Chan, K. S., and J. Ledolter, 1995, Monte Carlo EM Estimation for Time Series Models Involving Counts, *Journal of the American Statistical Association*, 90: 242-252.

Centeno, M., and J. M. A. Silva, 2001, Bonus Systems in an Open Portfolio, *Insurance: Mathematics & Economics*, 28: 341-350.

Denuit, M., and J. Dhaene, 2001, Bonus-Malus Scales Using Exponential Loss Functions, *German Actuarial Bulletin*, 25: 13-27.

Denuit, M., S. Pitrebois, and J.-F. Walhin, 2003, Tarification Automobile Sur Données de panel, *Bulletin of the Swiss Association of Actuaries*, (in press).

Diggle, P. J., K. Y. Liang, and S. L. Zeger, 1994, *Analysis of Longitudinal Data* (New York: Oxford University Press).

Dionne, G., and C. Vanasse, 1989, A Generalization of Actuarial Automobile Insurance Rating Models: The Negative Binomial Distribution with a Regression Component, *ASTIN Bulletin*, 19: 199-212.

Dionne, G., and C. Vanasse, 1992, Automobile Insurance Ratemaking in the Presence of Asymmetrical Information, *Journal of Applied Econometrics*, 7: 149-165.

Dobson, A. J., 1990, *An Introduction to Generalized Linear Models* (London: Chapman and Hall/CRC).

Gerber, H. U., and D. Jones, 1975, Credibility Formulas of the Updating Type, *Transactions of the Society of Actuaries*, 27: 31-52.

Gilde, V., and B. Sundt, 1989, On Bonus Systems with Credibility Scales, *Scandinavian Actuarial Journal*, 13-22.

Joe, H., 1997, *Multivariate Models and Dependence Concepts* (London: Chapman and Hall).

Lemaire, J., 1995, *Bonus-Malus Systems in Automobile Insurance* (Boston: Kluwer Academic Publisher).

Liang, K. Y., and S. L. Zeger, 1986, Longitudinal Data Analysis Using Generalized Linear Models, *Biometrika*, 73: 13-22.

Norberg, R., 1976, A Credibility Theory for Automobile Bonus System, *Scandinavian Actuarial Journal*, 92-107.

Pinquet, J., 1998, Designing Optimal Bonus-Malus Systems from Different Types of Claims, *ASTIN Bulletin*, 205-220.

Pinquet, J., 2000, Experience Rating Through Heterogeneous Models, in: G. Dionne, ed., *Handbook of Insurance* (Amsterdam: Kluwer Academic Publishers).

Pinquet, J., M. Guillén, and C. Bolancé, 2001, Allowance for the Age of Claims in Bonus-Malus Systems, *ASTIN Bulletin*, 31(2): 337-348.

Renshaw, A. E., 1994, Modelling the Claim Process in the Presence of Covariates, *ASTIN Bulletin*, 24: 265-285.

Sundt, B., 1981, Recursive Credibility Estimation, *Scandinavian Actuarial Journal*, 3-21.

Sundt, B., 1988, Credibility Estimators With Geometric Weights, *Insurance: Mathematics & Economics*, 7: 113-122.

Taylor, G., 1997, Setting a Bonus-Malus Scale in the Presence of Other Rating Factors, *ASTIN Bulletin*, 27: 319-327.

Zeger, S. L., 1988, A Regression Model for Time Series of Counts, *Biometrika*, 75: 621-629.

Zeger, S. L., K. Y. Liang, and P. S. Albert, 1988, Models for Longitudinal Data: A Generalized Estimating Equation Approach, *Biometrics*, 44: 1049-1060.