North American Actuarial Journal
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/uaaj20

Bonus-Malus Systems

Jean Lemaire A.S.A., Ph.D. [a]

[a] Insurance and Risk Management Department at The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104

Published online: 04 Jan 2013.

Bonus-Malus Systems: The European and Asian Approach to Merit-Rating

Jean Lemaire*

Abstract

Bonus-malus is a merit-rating technique used in most of Europe and Asia, and some Latin American and African countries. Policyholders from a given risk cell are subdivided into bonus-malus classes. Their claims histories then modify the class upon each renewal. Markov chain theory provides the tools for the design, evaluation, and comparison of these systems. In this article, definitions and examples of bonus-malus systems are provided (Section 2). The main actuarial tools for the study and design of bonus-malus systems are reviewed (Section 3). In the discussions that follow, Krupa Subramanian outlines a model for analyzing market shares in a competitive environment, a crucial research topic given current deregulation trends, and Pierre Lemaire compares actuarial with regulatory approaches to bonus-malus.

1. Introduction

In most developed countries, insurers use some form of merit-rating, in addition to other classification variables, in automobile third-party liability insurance. In the U.S. and Canada, insurance companies tend to use many *a priori* variables, such as age, sex, marital status and driving experience of the policyholder, car model, use of car, county of residence, and so on. Compared to other countries, the U.S. uses *a posteriori* or merit-rating to a limited extent: in many states and provinces, at-fault accidents and moving traffic violations are translated into penalty "points" and lead to a premium surcharge for three years. In other countries, insurance carriers usually rely on far fewer rating variables. For instance, in Switzerland, until recently insurers were authorized to use only one variable, the power of the engine, with more than 70% of the policies clustered in one single cell. In these countries, insurers rely on a much more sophisticated form of merit-rating, called *bonus malus*.

Bonus-malus systems (BMSs) were introduced in Europe in the early 1960s, following the seminal works of Delaporte (1965), Bichsel (1964), and Bühlmann (1964). The very first ASTIN Colloquium, held in La Baule, France, in 1959, was devoted exclusively to bonus-malus. According to ASTIN legend, when Général De Gaulle became President of France in 1958, he ordered French companies to introduce BMSs in automobile insurance. French actuaries then convened the first ASTIN meeting. There exists a vast literature on BMSs in actuarial journals, mainly in the *ASTIN Bulletin* and the *Swiss Actuarial Journal*. A recent book (Lemaire 1995) summarizes this literature and provides more than 140 references and the complete description of 31 systems. The goal of this paper is to introduce the main ideas underlying the design of a BMS to North American actuaries and to present some recent research developments.

BMSs are not commonly used in North America. This may change in the future because:

- *A posteriori* rating is a very efficient way of classifying policyholders into cells according to their risk. Several studies have shown that, if insurers are allowed to use only one rating variable, it should be some form of merit-rating (for instance, Lemaire 1985, ch. 7). The best predictor of the number of claims of a driver in the future is not age, car, or the township of residence, but past claims behavior.
- Several variables commonly used in North America, such as age, sex, and territory, are subject to intense scrutiny. Regulatory authorities may prohibit insurers from using some of these in the future, thereby forcing the insurance industry to rely more on merit-rating. In Massachusetts, where insurers do not use age, sex, and marital status, a Safe Driver

*Jean Lemaire, A.S.A., Ph.D., is Chairperson of the Insurance and Risk Management Department at The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104.

Insurance Plan that is essentially a BMS was introduced in 1990.

2. DEFINITION AND EXAMPLES

By definition, an insurance company uses a BMS when

- The insureds of a given tariff group can be partitioned into a finite number of classes, denoted C_i or simply i ($i=1, \ldots, s$), so that the annual premium depends only on the class (s denotes the number of classes) and on the tariff group
- Policyholders begin their driving career in a specified starting class C_{i_0}
- An insured's class for a given period of insurance (usually a year) is determined uniquely by the class for the preceding period and the number of claims reported during the period.

Such a system is determined by three elements:

- The premium scale $\overline{b}=(b_1, \ldots, b_s)$
- The initial class C_{i_0}
- The transition rules—the rules that determine the transfer from one class to another when the number of claims is known.

These rules can be introduced as transformations T_k, such that $T_k(i)=j$ if the policy is transferred from class C_i into class C_j when k claims have been reported. The term T_k can be written in the form of a matrix

$$T_k = (t_{ij}^{(k)}),$$

where $t_{ij}^{(k)}=1$ if $T_k(i)=j$ and 0 otherwise. The probability $p_{ij}(\lambda)$ of a policy moving from C_i into C_j in one period, for a policyholder characterized by some parameter λ (for instance, claim frequency), is equal to

$$p_{ij}(\lambda) = \sum_{k=0}^{\infty} p_k(\lambda)\ t_{ij}^{(k)},$$

where $p_k(\lambda)$ is the probability that a driver with claim frequency λ has k claims in a year. Obviously $p_{ij}(\lambda)\geq 0$ and

$$\sum_{j=1}^{s} p_{ij}(\lambda) = 1.$$

The matrix

$$M(\lambda) = [p_{ij}(\lambda)] = \sum_{k=0}^{\infty} p_k(\lambda)\ T_k$$

is the transition matrix of this Markov chain.

An insured enters the system, in the initial class, when he or she obtains a driving license. Then, throughout the entire driving lifetime, the transition rules are applied upon each renewal to determine the new class as a function of claims history. If a policyholder decides to switch to a new carrier, he or she has to obtain a certificate from the former company specifying the current BMS class and recent claims that could influence the new class.

Usually only the number of claims at-fault in third-party liability is used to penalize policyholders. Traffic violations are not taken into consideration. Korea is the only country where transition rules depend on claim severity, with property-damage losses subdivided into two categories and bodily injury claims into 14.[1]

Regulatory environments in European and Asian countries are extremely diversified, from total freedom to government-imposed systems, with many intermediate situations. The approach to BMS design depends on regulation. If a tariff is imposed by the government and every insurer has to use it, there is no commercial pressure to match the premiums to the risks by making use of all available relevant information. Supervising authorities may choose, for sociopolitical reasons, to exclude from the tariff structure certain risk factors, even though they may be significantly correlated to losses. The government may then seek to correct for the inadequacies of the *a priori* system by using a "tough" BMS that penalizes claims more heavily. In a free market, carriers need to use a rating structure that matches the premiums to the risks as closely as possible, or at least as closely as the rating structure used by competitors. This entails using virtually every available classification variable correlated to the risks, because failing to do so would mean sacrificing the chance to select against competitors and incurring the risk of suffering adverse selection by them. The use of more classification variables is expected in free market countries, which decreases the need for a sophisticated BMS.

Some countries have adopted very simple BMSs. In Brazil, for instance, policyholders are subdivided into just seven classes, with premium levels 100, 90, 85, 80, 75, 70, and 65. New policyholders have to start in class 7, at level 100. Each claim-free year results in a one-class discount. Each at-fault claim is penalized by one class. The transition rules are presented in Table 1.

[1]Japan had a rule that each bodily injury loss was penalized as two property-damage claims, but recently abandoned it.

TABLE 1
BRAZILIAN SYSTEM (STARTING CLASS: 7)

Class	Premium	Class After 0 Claims	1 Claims	2 Claims	3 Claims	4 Claims	5 Claims	≥6 Claims
7	100	6	7	7	7	7	7	7
6	90	5	7	7	7	7	7	7
5	85	4	6	7	7	7	7	7
4	80	3	5	6	7	7	7	7
3	75	2	4	5	6	7	7	7
2	70	1	3	4	5	6	7	7
1	65	1	2	3	4	5	6	7

Belgium recently adopted a more sophisticated system, with 23 classes. Pleasure-users and commuters enter the system in class 11, at level 85. Business-users enter in class 14, at level 100. Each claim-free year leads to a one-class discount. The first claim in any policy year is penalized by four classes. Any subsequent claim in the same year results in a five-class penalty. Table 2 presents the transition rules of this system.

The preceding definition assumes that the BMS forms a Markov chain process. A (first-order) Markov chain is a stochastic process in which the future development depends only on the present state but not on the history of the process or the manner in which the present state was reached. It is a process without memory, such that the states of the chain are the different BMS classes. The knowledge of the present class and the number of claims for the year suffice to determine next year's class. It is not necessary to know how the policy reached the current class.

In fact, the Belgian system does not form a Markov process. In addition to the rules mentioned above,

TABLE 2
BELGIAN SYSTEM

Class	Premium	Class After 0 Claims	1 Claims	2 Claims	3 Claims	4 Claims	≥5 Claims
22	200	21	22	22	22	22	22
21	160	20	22	22	22	22	22
20	140	19	22	22	22	22	22
19	130	18	22	22	22	22	22
18	123	17	22	22	22	22	22
17	117	16	21	22	22	22	22
16	111	15	20	22	22	22	22
15	105	14	19	22	22	22	22
14	100	13	18	22	22	22	22
13	95	12	17	22	22	22	22
12	90	11	16	21	22	22	22
11	85	10	15	20	22	22	22
10	81	9	14	19	22	22	22
9	77	8	13	18	22	22	22
8	73	7	12	17	22	22	22
7	69	6	11	16	21	22	22
6	66	5	10	15	20	22	22
5	63	4	9	14	19	22	22
4	60	3	8	13	18	22	22
3	57	2	7	12	17	22	22
2	54	1	6	11	16	21	22
1	54	0	5	10	15	20	22
0	54	0	4	9	14	19	22

Starting class: 11 or 14

Belgian regulatory authorities have added a special transition rule that no policy can be in a class above 14 after four consecutive claim-free years. This last restriction is a concession to youthful operators who have many accidents in their early years and who suddenly improve. Very few policyholders are ever able to take advantage of this rule. Yet it makes the BMS non-Markovian. It forces insurance companies to memorize the claims history of some policyholders for four years, instead of simply the present class, had this restriction not been allowed. Indeed, after a claim-free year, a Belgian customer in class 17 will be sent to class 14 or 16, depending on the number of consecutive claim-free years earned before.

Fortunately, it is possible to modify the presentation of the system into a Markovian way, at the price of an increase of the number of classes. Classes are subdivided by adding an index that counts the number of consecutive claim-free years. In Markov chain terminology, the state variable is augmented with sufficient information so that a Markovian analysis is possible. Table 3 provides this extended presentation. The modified BMS requires 35 classes, up from 23 initially.

TABLE 3
BELGIAN BMS: MARKOVIAN PRESENTATION

		Class After				
Class	Premium	0 Claims	1 Claims	2 Claims	3 Claims	4 Claims
22	200	21.1	22	22	22	22
21.0	160	20.1	22	22	22	22
21.1	160	20.2	22	22	22	22
20.0	140	19.1	22	22	22	22
20.1	140	19.2	22	22	22	22
20.2	140	19.3	22	22	22	22
19.0	130	18.1	22	22	22	22
19.1	130	18.2	22	22	22	22
19.2	130	18.3	22	22	22	22
19.3	130	14	22	22	22	22
18.0	123	17	22	22	22	22
18.1	123	17.2	22	22	22	22
18.2	123	17.3	22	22	22	22
18.3	123	14	22	22	22	22
17	117	16	21.0	22	22	22
17.2	117	16.3	21.0	22	22	22
17.3	117	14	21.0	22	22	22
16	111	15	20.0	22	22	22
16.3	111	14	20.0	22	22	22
15	105	14	19.0	22	22	22
14	100	13	18.0	22	22	22
13	95	12	17	22	22	22
12	90	11	16	21.0	22	22
11	85	10	15	20.0	22	22
10	81	9	14	19.0	22	22
9	77	8	13	18.0	22	22
8	73	7	12	17	22	22
7	69	6	11	16	21.0	22
6	66	5	10	15	20.0	22
5	63	4	9	14	19.0	22
4	60	3	8	13	18.0	22
3	57	2	7	12	17	22
2	54	1	6	11	16	21.0
1	54	0	5	10	15	20.0
0	54	0	4	9	14	19.0

3. Tools for the Design and Evaluation of Bonus-Malus Systems

BMSs can be analyzed from the perspective of the policyholder or the insurance carrier. The tools are the same, but assumptions about probability distributions for the number of claims vary. For instance, if the Poisson distribution is acceptable to model the number of losses of an individual policyholder, differences among drivers make it inadequate to represent loss counts for an insurer. Distributions such as the negative binomial and the Poisson-Inverse Gaussian systematically outperform the Poisson to fit observed portfolio loss counts. The policyholder is emphasized in this paper. It is assumed that the distribution $\{p_k; k=0, 1, 2, \ldots\}$ of the number of claims of a specific driver conforms to a Poisson with parameter λ, where λ is called *the claim frequency* of the policyholder and is assumed to be constant over time.

$$p_k = \frac{e^{-\lambda}\,\lambda^k}{k!}$$

This section reviews the major tools that actuaries use to design, evaluate, and compare BMSs.

3.1. The Relative Stationary Average Level

The *relative stationary average level* (RSAL) measures the position of the average driver, on a scale from zero to one, once the BMS has reached its steady-state condition. It evaluates the degree of concentration of policies in the lower classes of the BMS.

An apparently inescapable consequence of the implementation of a BMS is a progressive decrease of the observed average premium level because of a clustering of the policies in the high-discount classes. For instance, the Belgian system penalizes accidents by four or five classes and awards a one-class discount per claim-free year. The average observed claim frequency in the country is, however, close to 10%. Consequently, in any given year, the total number of bonus classes awarded to claim-free policyholders is much larger than the number of malus classes given to drivers with claims, and the mean premium level decreases. After a few years, the majority of policies is concentrated in the lowest classes of the BMS. The insurance carrier is then forced to compensate for this decrease by increasing the dollar cost of the premium at level 100, which somewhat defeats the purpose of the system. To maintain the BMS in financial equilibrium without raising the basic premium, it would be necessary to penalize each claim by eight or nine classes. While statistically perfectly justified, such penalties seem commercially impossible to enforce.

A forecast of the future distribution of policies among the classes, say n years from now, can be obtained easily through simulation or by computing the n-th power of the transition matrix $M(\lambda)$. For many purposes, an asymptotic study is sufficient to compare BMSs. Only steady-state results are outlined in this paper.

By using the Kemeny and Snell (1960) terminology, a BMS forms a regular Markov chain: all its states are ergodic (it is possible to go from every state to every other state), and the chain is not cyclic. In that case the value 1 is a simple eigenvalue of the transition matrix $M(\lambda)$. The corresponding left eigenvector, row vector

$$\bar{a}(\lambda) = [a_1(\lambda), a_2(\lambda), \ldots, a_s(\lambda)]$$

defined by the equation

$$\bar{a}(\lambda) = \bar{a}(\lambda)\, M(\lambda),$$

and

$$\sum_{i=1}^{s} a_i(\lambda) = 1$$

is the *stationary probability distribution*. The term $a_i(\lambda)$ is the limit value for the probability that the policy is in class C_i, when the number of periods tends to infinity. It is also the fraction of the time a policyholder with claim frequency λ spends in class C_i, once stationarity has been reached. As an illustration, let us compute the stationary distribution for the Brazilian system, when $\lambda=0.1$. For this BMS the transition matrix is (with class 7 at the top):

$$M(\lambda) = \begin{matrix} 1-p_0 & p_0 & 0 & 0 & 0 & 0 & 0 \\ 1-p_0 & 0 & p_0 & 0 & 0 & 0 & 0 \\ 1-\Sigma p_i & p_1 & 0 & p_0 & 0 & 0 & 0 \\ 1-\Sigma p_i & p_2 & p_1 & 0 & p_0 & 0 & 0 \\ 1-\Sigma p_i & p_3 & p_2 & p_1 & 0 & p_0 & 0 \\ 1-\Sigma p_i & p_4 & p_3 & p_2 & p_1 & 0 & p_0 \\ 1-\Sigma p_i & p_5 & p_4 & p_3 & p_2 & p_1 & p_0 \end{matrix}$$

where the lower bound of Σp_i in each row is such that all row probabilities add up to 1. For $\lambda=0.10$, the left eigenvector of this matrix has the components:

$a_1 = 0.88948$; $a_2 = 0.09355$; $a_3 = 0.01444$; $a_4 = 0.00215$; $a_5 = 0.00032$; $a_6 = 0.00005$; $a_7 = 0.00001$.

It represents the asymptotic distribution of policies among the seven classes of the system. It indicates, for instance, that 89% of the policyholders with $\lambda=0.10$ will eventually belong to class 1. Multiplying each probability by the corresponding premium level, the mean asymptotic premium level is found to be 65.65, an extremely low value easily explained by the soft transition rules of that BMS.

Figure 1 presents the evolution of the mean premium level for five systems. For a simple system like the Taiwanese, the premium decreases abruptly in the first few years, the time it takes for the best policyholders to reach the largest discount. The system then stabilizes rapidly. For the more sophisticated systems, the premium decreases in a much smoother way, and the steady state is not reached until more than 30 years has elapsed. This is a very long time to stabilize, given the short driving lifetime of the average policyholder: because most insureds do not drive for more than 60 years, a period of 30 years to stabilize their premium around their "true" level seems excessive.

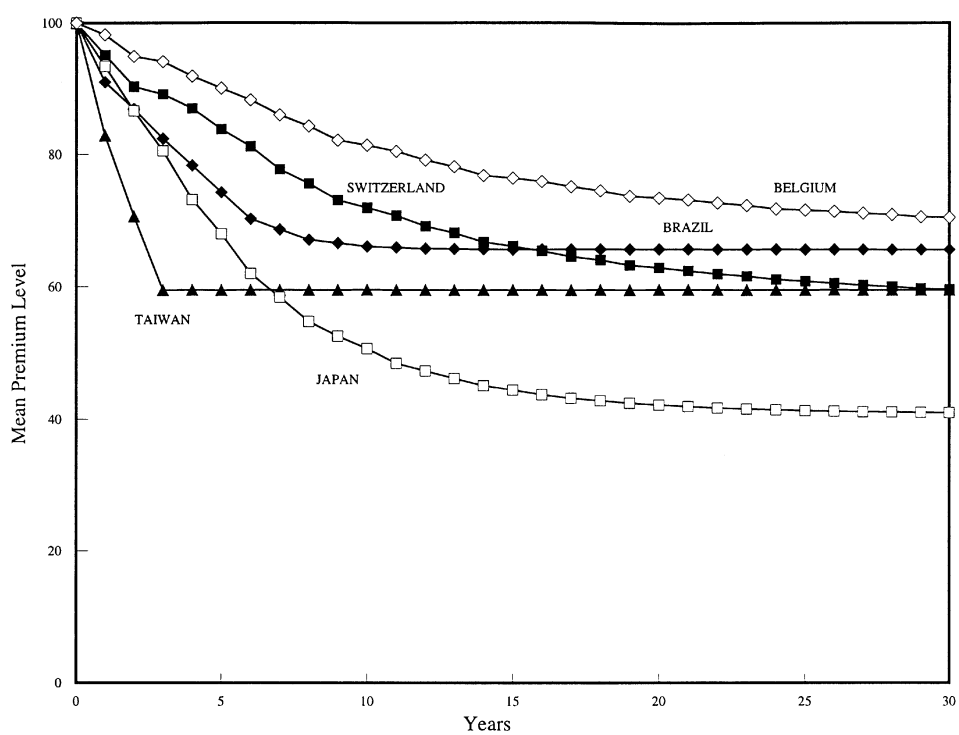
Given the wide variety of systems in force, stationary average levels are difficult to compare. Therefore, the relative stationary average level is defined as

$$\text{RSAL} = \frac{\text{stationary average level} - \text{minimum level}}{\text{maximum level} - \text{minimum level}}.$$

Expressed as a percentage, this index determines the relative position of the average policyholder, when the lowest premium is set equal to zero and the highest to 100. A low value of RSAL indicates a high clustering of policies in the high-discount BMS classes. A high RSAL suggests a better spread of policies among classes. The RSAL is only a crude measure of the severity of a BMS, because it is influenced by the premium level of the sparsely populated highest class. Most systems lead to a very low RSAL value. Only Kenya, Spain, and Malaysia-Singapore use BMSs with a RSAL higher than 20%. For the five systems of Figure 1, the RSAL is less than 10%. Note that the RSAL is measured here assuming no entries or exits. In an open portfolio, the RSAL would be somewhat higher, because exits are at lower levels than entries.

A consequence of a low RSAL is that all BMSs carry an implicit penalty for new drivers, since the premium level of the access class is substantially higher than the average stationary premium level. The first-year surcharge is easily computed as (entry premium $-$ stationary premium)/stationary premium. BMS-induced surcharges range from a low of 26.95% to a high of 212.97% in Germany. In addition to these implicit increases, many countries have introduced explicit penalties for inexperienced drivers, for example, a surcharge in France and a deductible after a claim in Belgium and Switzerland.

FIGURE 1
EVOLUTION OF MEAN PREMIUM LEVEL



3.2. The Coefficient of Variation of the Insured's Premium

Insurance consists of a transfer of risk from the policyholder to the carrier. Without experience rating, the transfer is total (perfect solidarity): the variability of insureds' payments is zero. With experience rating, personalized premiums vary from year to year, according to claims history; cooperation between drivers is weakened. Solidarity between insureds can be evaluated by a measure of the variability of annual premiums. The coefficient of variation (standard deviation divided by mean) was selected, because it is a dimensionless parameter. There is thus no need for currency conversions.

The Actuarial Institute of the Republic of China kindly provided marketwide observed loss severity distributions for property damage and bodily injury for accident years 1987 to 1989. These distributions are very well represented by a lognormal model (Lemaire 1993). Assuming that the aggregate claims process is compound Poisson with lognormal severities (Bowers et al. 1986, ch. 11), the coefficient of variation of aggregate losses is found to average 6.40. Although loss distributions in other countries of course differ from the Taiwanese experience, the coefficient of variation is not likely to be affected much. So, without insurance, the policyholder is exposed to a loss process with a coefficient of variation of 6.40. With full insurance and no *a posteriori* rating, the variability of premiums is zero. With a BMS, the coefficient of variation will be somewhere between 0 and 6.40. The tougher the BMS, the larger this coefficient.

Figure 2 shows the evolution of the coefficient of variation with time, for a policyholder with claim frequency $\lambda=0.10$, for the five selected systems. Typically, the coefficient of variation starts at zero for the first policy year, increases until the best policyholders reach the maximum discount, and then decreases until stationarity is reached, in some cases after more than 30 years.

A common criticism against BMSs is that premiums paid by policyholders are too variable. This argument seems unwarranted. The highest variability for $\lambda=0.10$ is achieved by the new Swiss system, with an asymptotic coefficient of variation of 0.4595. This represents only 7.18% of the variation of the loss process. Even with a severe system such as the Swiss, policyholders are requested to assume only a small part of the risk.

Figure 3 shows the coefficient of variation, when stationarity has been reached, as a function of the

FIGURE 2
EVOLUTION OF COEFFICIENT OF VARIATION

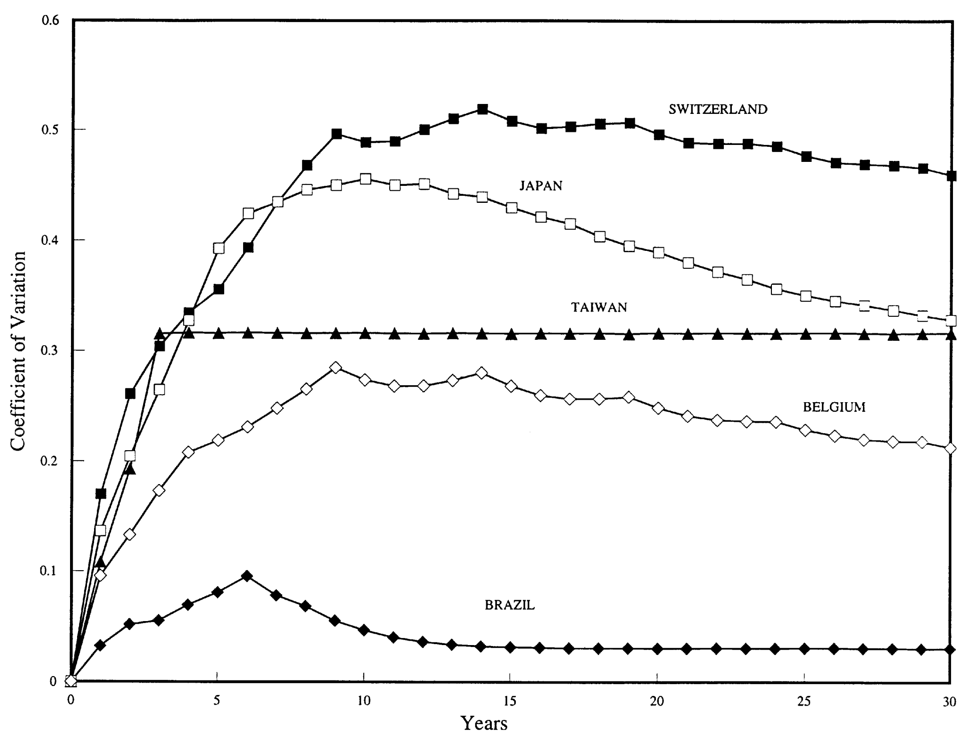
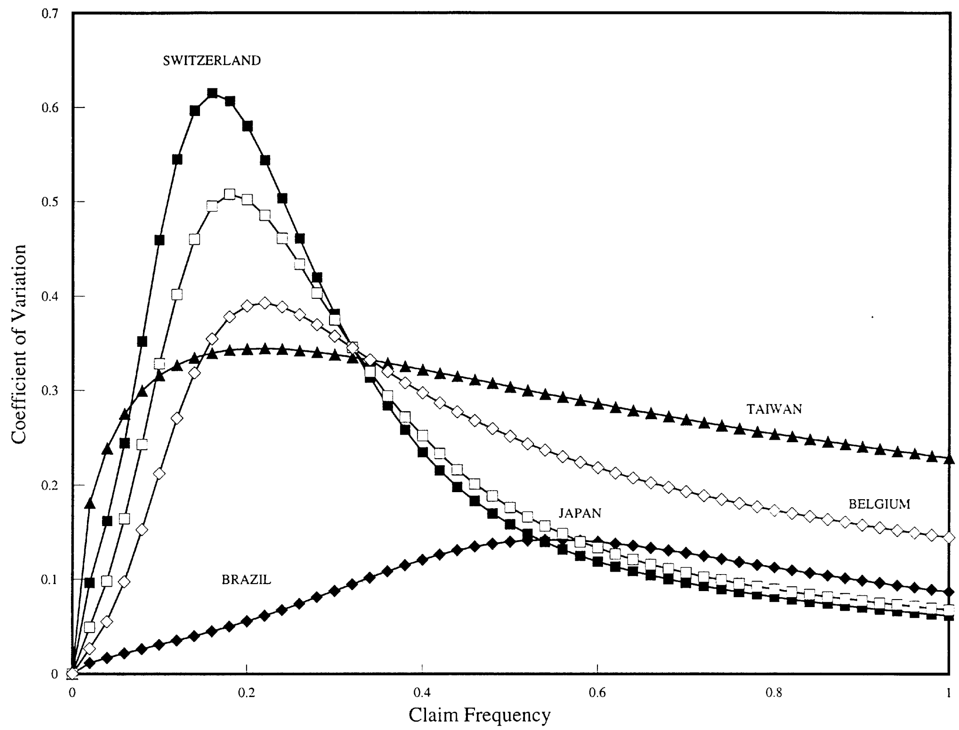


FIGURE 3
COEFFICIENT OF VARIATION AS A FUNCTION OF CLAIM FREQUENCY



claim frequency. The variability is zero for the policyholders who never cause a claim, because they will permanently stay in the best class. Similarly, the coefficient of variation tends to zero as $\lambda \rightarrow \infty$. With the exception of the Taiwanese system, the maximum variation occurs for claim frequencies that are significantly higher than the benchmark 10%.

3.3 The Elasticity of the Mean Stationary Premium with Respect to the Claim Frequency

The main goal of a BMS is to reduce the premium for good drivers and to increase it for bad risks. The random variables "number of claims" and "claim amount" are often assumed to be independent. This assumption essentially states that the cost of an accident is for the most part beyond the control of a policyholder. The degree of care exercised by a driver mostly influences the number of accidents, but in a much lesser way the cost of these accidents. The validity of the independence assumption has been questioned [for instance, by Lemaire (1985, ch. 5)]. Nevertheless, the fact that nearly all BMSs in force around the world penalize the number of claims, independently of their amount, is an indication that insurers and regulators accept it, at least as an approximation. It has the important implication that the risk of each driver can be measured by individual claim frequency λ.

The *elasticity* of a BMS measures the response of the system to a change in the claim frequency. Obviously, for any BMS, the lifetime premiums paid by policyholders have to be an increasing function of λ. Ideally, this dependence should be linear. A relative increase of the claim frequency should produce the same relative increase of the premium. To provide an intuitive description of the concept, consider two policyholders, one with a claim frequency of 0.10, the other with a λ of 0.11. Over a long time, the second driver should pay 10% more premiums than the first. A BMS with this property is called perfectly elastic. In practice, however, the mean premium increase in most cases is much lower than 10%. If the increase is, say, 2% instead of 10%, the system's elasticity is said to be 20%.

More rigorously, denote $P(\lambda)$ the mean stationary premium associated with a claim frequency λ. Ideally, an increment $d\lambda/\lambda$ of the claim frequency should lead to an equal change, $dP(\lambda)/P(\lambda)$, of the premium. A BMS is called *perfectly elastic* if

$$\frac{d\lambda/\lambda}{dP(\lambda)/P(\lambda)} = 1.$$

As a general rule, however, the change in premium is less than the change in λ. The elasticity $\eta(\lambda)$ of the BMS is defined as

$$\eta(\lambda) = \frac{dP(\lambda)/P(\lambda)}{d\lambda/\lambda} = \frac{d \ln P(\lambda)}{d \ln \lambda}.$$

It is the *elasticity of the mean stationary premium with respect to the claim frequency*. This concept was introduced in actuarial science, under the name of efficiency, by Loimaranta (1972).

Ideally, the elasticity should be close to 1 for the most common values of λ. Figure 4 shows the elasticity of the selected systems as a function of λ. For the most common values of λ ($\lambda \leq 0.15$), $\eta(\lambda)$ is very low. The elasticity reaches the neighborhood of 1 only for unusual values of λ. The Swiss system presents a rare case of overelasticity: $\eta(\lambda)>1$ for $\lambda \in [0.15-0.28]$. Only the new Swiss and Finnish systems exhibit an elasticity of more than 0.40 for $\lambda=0.10$.

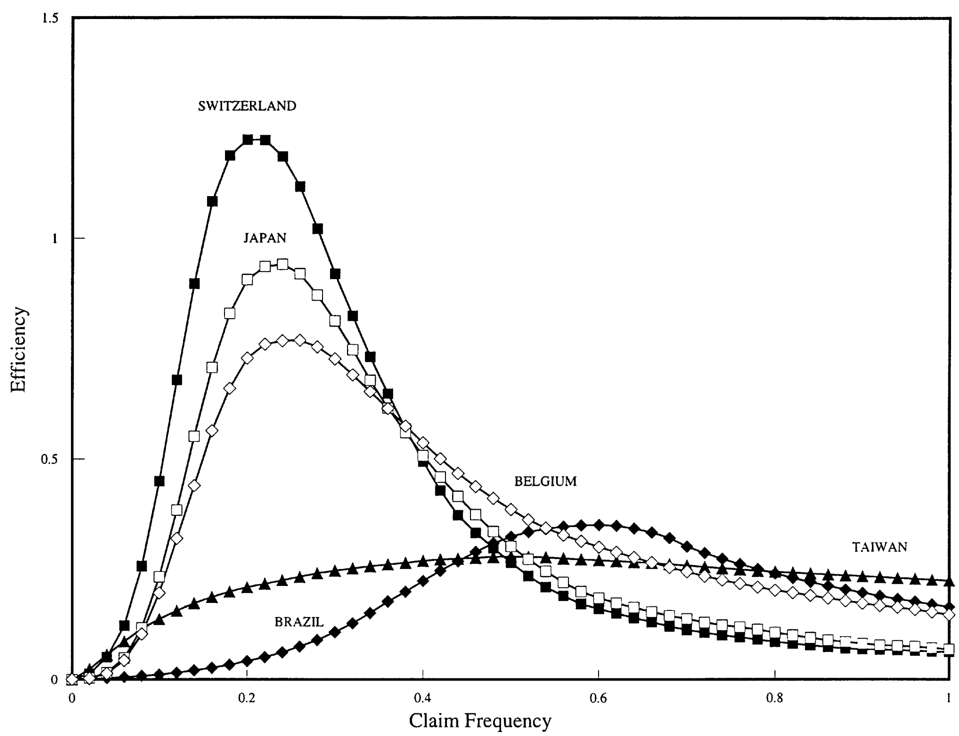
The low observed elasticities, along with the low coefficients of variation, indicate that the implementation of a BMS decreases, but does not eliminate, cross-subsidization between the risk classes. For all systems in force, the good risks still subsidize the poor drivers.

3.4 The Average Optimal Retention

A side effect of BMSs is a tendency of policyholders to pay small claims themselves and not report them to their carrier to avoid future premium increases. In some countries the existence of this phenomenon, called the *hunger for bonus*, has been explicitly recognized by regulators. In Germany, for instance, the policy wording specifies that, if the insured reimburses the carrier for the cost of the claim, the penalty will not be applied. Moreover, if the claim amount does not exceed DM 1,000, the company is required to inform the policyholder of the right to reimburse the loss.

If the claim amounts borne by policyholders are reasonable, the hunger for the bonus effect amounts to a small deductible and reduces administrative expenses. However, if a BMS induces drivers to pay claims in excess of, say, \$5,000, it definitely penalizes claims excessively and may create hit-and-run behaviors. The main objective of a BMS is to achieve a

FIGURE 4
ELASTICITY



better separation of the good and the bad risks; it is not to transfer most claims from the insurer to the insureds.

The determination of the optimal strategy of the policyholder is an interesting decision problem that has close links with infinite-horizon dynamic programming under uncertainty. For each class of the system, the following algorithm computes the optimal retention level—the level under which it is in the policyholder's interest not to report a claim. To simplify the presentation, retentions are computed here under an infinite-horizon assumption. Policyholders are supposed to drive forever. While this is obviously not the case in practice, calculations show that the effect of this assumption for long driving careers is minuscule. The decision problem can be formulated as follows.

Step 1. Evaluation of Given Strategies

Define a *strategy* for the policyholder as a vector $\bar{x}=(x_1, \ldots, x_i)$, where x_i is the retention limit for class C_i. The cost of any accident of amount less than or equal to x_i is borne by the policyholder. Claims of higher amounts are reported.

Consider a policyholder who has just caused an accident of amount x, at time t, with time units such that $0 \le t < 1$. Denote $f(x)$ the density function of the random variable X representing the cost of a claim. Given a specific strategy, the probability p_i of a claim not being reported by a policyholder in class C_i is

$$p_i = P(X \le x_i) = \int_0^{x_i} f(x)\, dx.$$

The probability $\bar{p}^i_k(\lambda)$ of reporting k claims during one period equals

$$\bar{p}^i_k(\lambda) = \sum_{h=k}^{\infty} p_h(\lambda) \binom{h}{k} (1 - p_i)^k\, p_i^{h-k}.$$

The expectation of the number of reported claims is

$$\bar{\lambda}_i = \sum_{k=0}^{\infty} k\, \bar{p}^i_k(\lambda).$$

The expected cost of a nonreported claim is equal to

$$E^i(X) = \frac{1}{p_i} \int_0^{x_i} x f(x)\, dx.$$

Assuming independence between the number and the amount of claims, the policyholder pays, on average, for the compensation of nonreported claims

$$E^i(X)\, (\lambda - \bar{\lambda}^i).$$

The expectation of the total cost for this period is

$$\mathrm{E}(x_i) = b_i + \beta^{1/2}\,\mathrm{E}^i(X)\, (\lambda - \bar{\lambda}^i),$$

introducing a discount factor β and assuming all claims take place in the middle of the period.

The vector

$$\bar{v}(\lambda) = [v_1(\lambda), \ldots, v_s(\lambda)]$$

of the discounted expectation of all the payments by the policyholder satisfies the set of s equations with s unknowns $v_i(\lambda)$ $(i=1, \ldots, s)$:

$$v_i(\lambda) = \mathrm{E}(x_i) + \beta \sum_{k=0}^{\infty} \bar{p}^i_k(\lambda)\, v_{T_k(i)}(\lambda).$$

This system always has a unique solution. The vector $\bar{v}(\lambda)$ of solutions to the system provides a numerical evaluation of the cost of every single possible strategy.

Step 2. Determination of the Optimal Strategy

The policyholder who causes a claim of amount x at time t has two possible courses of action: (1) if he or she does not report the accident, the expectation of total cost, discounted at the time of the claim, is

$$\beta^{-t}\,\mathrm{E}(x_i) + x + \beta^{1-t} \sum_{k=0}^{\infty} \bar{p}^i_k[\lambda(1 - t)]\, v_{T_{k+m}(i)}(\lambda),$$

where m is the number of claims already reported during the period; or (2) if the accident is reported to the company, the expectation is equal to

$$\beta^{-t}\,\mathrm{E}(x_i) + \beta^{1-t} \sum_{k=0}^{\infty} \bar{p}^i_k[\lambda(1 - t)]\, v_{T_{k+m+1}(i)}(\lambda).$$

The retention limit x_i is the claim cost x for which the two actions are equivalent:

$$x_i = \beta^{1-t} \sum_{k=0}^{\infty} \bar{p}^i_k[\lambda(1 - t)]\, [v_{T_{k+m+1}(i)}(\lambda) - v_{T_{k+m}(i)}(\lambda)].$$

These equations constitute a set of s equations with s unknowns x_i, as the x_i appear in an implicit way in the $\bar{p}^i_k[\lambda(1-t)]$. The system provides a new strategy vector $\bar{x}$, corresponding to a cost vector $\bar{v}$ (λ).

Summarizing, the first system provides $\bar{v}(\lambda)=[v_1(\lambda), \ldots, \bar{v}_s(\lambda)]$, for given $\bar{x}=(x_1, \ldots, x_s)$. The second system provides $\bar{x}$, for given $\bar{v}(\lambda)$. Taken together, the two systems consist in $2s$ equations with $2s$ unknowns, with, as solution, the optimal strategy $\bar{x}^*=(x_1^*, \ldots, x_s^*)$ and its associated cost $\bar{v}^*(\lambda)$. The system is best solved by successive approximations.

Application

The optimal strategy depends on the class C_i, the discount factor β, the claim frequency λ, the time of the accident t, and the number of claims already reported m.

The optimal strategy is an increasing function of t: the retention increases, as policy renewal (and the resulting likely discount) nears. However, the dependence of x on t is quite small, compared to the influences of C_i, β, and λ. Setting $t=0$ (and consequently $m=0$) simplifies all calculations and hardly reduces optimal retentions. The algorithm described above was applied to the Belgian BMS, first assuming a claim frequency $\lambda=0.10$ and a discount coefficient $\beta=0.90$. The 1989 observed property damage loss distribution in Taiwan was used, with some allowance for inflation and for the fact that Taiwanese losses may be slightly below world averages.[2]

Table 4 presents the optimal retention, for all classes, as a percentage of the average premium. Remember that some classes had to be subdivided to obtain a Markovian presentation of the system (Table 3). Given the toughness of the BMS, optimal retentions are very high. With the exception of a handful of classes (the most populated, however), it is always in a policyholder's interest to bear the cost of an accident that is more expensive than the average premium.

TABLE 4
OPTIMAL RETENTIONS: BELGIAN BMS

Class	Optimal Retention	Class	Optimal Retention
0	38.41%	17.2	296.85%
1	56.50	17.3	360.03
2	76.59	18.0	288.16
3	98.26	18.1	326.98
4	117.80	18.2	382.01
5	137.34	18.3	457.52
6	156.05	19.0	257.56
7	174.03	19.1	304.64
8	190.40	19.2	369.69
9	208.83	19.3	457.52
10	224.98	20.0	228.29
11	239.38	20.1	283.74
12	254.56	20.2	359.28
13	273.65	21.0	196.03
14	285.46	21.1	260.11
15	269.02	22	147.31
16	254.05		
16.3	305.99		
17	252.17		

The system has a special rule, that no policyholder can be in the malus zone after four consecutive claim-free years. The impact of the special transition rule on optimal retentions is evidenced in Table 4; a driver in class 18.0 (who had an accident last year) has an optimal retention of 288.16% of the average premium. This retention increases to 457.52% for an insured in class 18 with three claim-free years.

The retentions in Table 4 depend on the values taken by two crucial parameters: λ, the claim frequency, and β, the discount factor. Because the values taken by these parameters are difficult to estimate accurately by the policyholder, it is of crucial importance to study the variability of x^* in terms of λ or β. Figure 5 shows, for five different classes of the Belgian BMS, the optimal x_i as a function of λ, with given discount factor $\beta=0.9$. For values of the claim frequency in excess of 1, x_i^* rapidly drops to zero, because it does not pay to indemnify claims for someone who expects several claims per year. Note the low slope of these curves: the optimal strategy is not influenced much by a change in λ. A slight error in the estimation of λ has only minor consequences.

Figure 6 shows the variation of the optimal retention with the discount factor, for constant $\lambda=0.10$, for five selected classes. Obviously, all curves are increasing. A policyholder who discounts future payments a great deal does not have much interest in paying claims. For large discount factors, the slopes of the curves are rather large. A 1% error in the determination of the discount factor has a much larger impact than a 1% error in the claim frequency.

3.5 The Rate of Convergence of Bonus-Malus Systems

The first part of this section demonstrates that simple BMSs like the Taiwanese reach stationarity much faster than sophisticated systems like the Swiss. An evaluation of the rate of convergence of BMSs to their steady-state condition is of great importance, because many of the tools defined here assume that stationarity has been reached.

Let $p_{ij}(\lambda)$ be the transition probabilities of the Markov chain associated with each BMS, and $p_{ij}^n(\lambda)$ the n-step transition probabilities. The term $p_{ij}^n(\lambda)$ is the probability of moving from class C_i to class C_j in exactly n transitions. These probabilities are obtained by computing the n-th power of the transition matrix $M(\lambda)$. Let $a_j(\lambda)$, $j=1, \ldots, n$, be the stationary distribution. Consider a specific BMS class C_i (usually the

[2]Other applications of the algorithm have shown optimal strategies to be quite insensitive to the loss distribution.

FIGURE 5
OPTIMAL RETENTION AS A FUNCTION OF CLAIM FREQUENCY

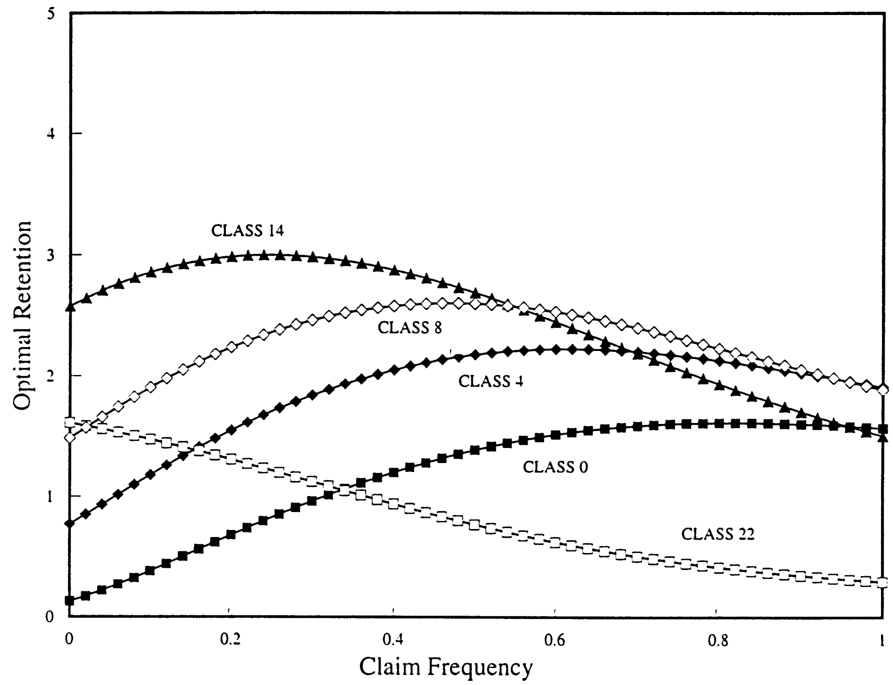
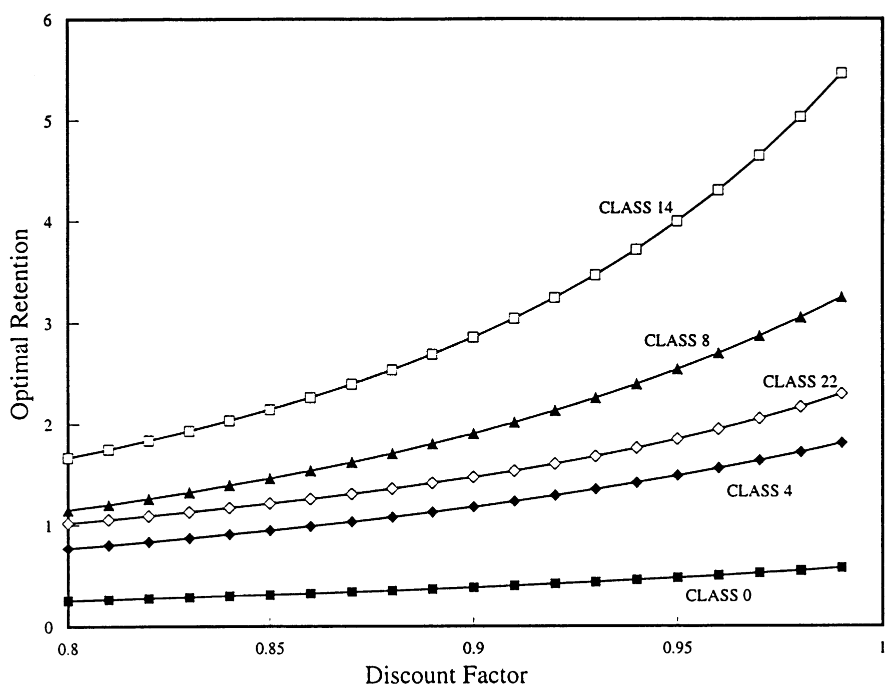


FIGURE 6
OPTIMAL RETENTION AS A FUNCTION OF THE DISCOUNT FACTOR



starting class of the system), a given λ, and a given number of steps n. Following Bonsdorff (1992), define the *total variation*

$$(TV)_n = \sum_{j=1}^{s} | p_{ij}^{n}(\lambda) - a_j(\lambda) |$$

as a measure of the degree of convergence of the system after n transitions. For any two probability distributions, the total variation is always between 0 and 2. The term $(TV)_n$ can prove to be a very useful tool for selecting factors that affect the pace to stationarity, such as the number of classes, the starting class, or the transition rules.

Table 5 shows the total variation for four systems, when $\lambda=0.10$ and C_i is the starting class of the BMS, as a function of time. It shows that sophisticated BMSs converge extremely slowly. While the Taiwanese system reaches full stationarity after only three years, there still exists significant variability after 30 years for other systems. The total variation for the Belgian BMS is still more than 20% of its original value. Even after 60 years, it still has not fully stabilized! This is a drawback of bonus-malus rating. The main objective of bonus-malus rating is to correct the inadequacies of *a priori* rating by separating the good from the bad drivers. This separation process should proceed as fast as possible. A 30-year separation phase—although justified by the low overall claim frequency and the resulting inherent variability of events—can be considered as excessive, because it exceeds half the driving lifetime of the majority of policyholders. It also exceeds the "life expectancy" of all BMSs. In fact, no single BMS has in the past been allowed to survive 30 years and to reach stationarity. All the countries that adopted bonus-malus rating in the late 1950s or early 1960s have switched to a second-generation BMS.

TABLE 5
TOTAL VARIATION FOR FOUR SYSTEMS

Years	Belgium	Japan	Taiwan	Switzerland
0	1.9913	1.9950	2.000	1.9742
10	1.7769	1.1551	0	1.0124
20	0.9120	0.3217	0	0.3541
30	0.4209	0.0529	0	0.1348
60	0.0382	0.0007	0	0.0061

REFERENCES

BICHSEL, F. 1964. "Erfahrungs-Tarifierung in der Motorfahrzeughalfplichtversicherung," *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker*, 119–129.

BONSDORFF, H. 1992. "On the Convergence Rate of Bonus-Malus Systems," *ASTIN Bulletin* 22:217–223.

BOWERS, N., HICKMAN, J., GERBER, H., JONES, D., AND NESBITT, C. 1986. *Actuarial Mathematics*. Itasca, Ill.: Society of Actuaries.

BÜHLMANN, H. 1964. "Optimale Prämienstufensysteme," *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker*, 193–213.

DELAPORTE, P. 1965. "Tarification du risque individuel d'accidents par la prime modelée sur le risque," *ASTIN Bulletin* 3:251–271.

KEMENY, J.G., AND SNELL, J.L. 1960. *Finite Markov Chains*. New York: Van Nostrand Reinhold.

LEMAIRE, J. 1985. *Automobile Insurance: Actuarial Models*. Boston: Kluwer.

LEMAIRE, J. 1993. "Selecting a Fitting Distribution for Taiwanese Automobile Losses," unpublished manuscript.

LEMAIRE, J. 1995. *Bonus-Malus Systems in Automobile Insurance*. Boston: Kluwer.

LOIMARANTA, K. 1972. "Some Asymptotic Properties of Bonus Systems," *ASTIN Bulletin* 6:233–245.

DISCUSSIONS

KRUPA SUBRAMANIAN*

BONUS-MALUS SYSTEMS IN A COMPETITIVE ENVIRONMENT

In this discussion, a simple model is developed to analyze the evolution of market shares and the financial stability of two insurers when one of them adopts an aggressive competitive behavior by modifying its bonus-malus system. The two scenarios that are developed show that rating freedom encourages insurers to adopt tougher systems.

Until a few years ago, in all but a handful of countries, all companies had to use the same BMS.[1] This situation is changing very fast. Deregulation ideas are gaining ground in Asia. European Economic Community directives have introduced complete rating

*Krupa Subramanian, A.S.A., is a doctoral student at The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104.

[1]A notable exception is the United Kingdom, where complete rating freedom has existed for many years. Drivers have always been able to choose among many different (but similar) systems.