J. King Saud Univ., Vol. 3, *Admin. Sci.* (2), pp. 83–94, Riyadh (1411 A.H./1991)

A Simple Actuarial Approach for Automobile Rating

Ahmed Kamhawey Abaza
Assistant Professor, Dept of Quantitative Methods, College of Administrative Sciences, King Saud University, Riyadh, Saudi Arabia
(Received 28/10/1410; Accepted for Publication 22/5/1411)

Abstract. In automobile insurance, it is traditional to cross-classify risks on the basis of two factors: **First,** drivers are divided according to a driver class variable formed from a complex combination of individual characteristics. **Second,** automobiles are also divided into discrete territories according to where they are principally engaraged. The purpose of this paper has been to adopt and modify the traditional actuarial methodology for calculating pure premium on the basis of cross-classified data as a simple actuarial approach for automobile rating.

Introduction

In the absence of other means of transportation, the automobile is a virtual necessity for many people. Given the strong dependency on cars and the substantial risks associated with owning and operating one, the purchase of automobile insurance is an important consumer decision, the significance of which is amplified by the high cost of insurance. In many countries it is characterized by the existence of a merit-demerit (bonus-malus) rating system. According to such a system the premium that the insured pays decreases if he or she does not make a claim, and increase if he or she does. While the number of claims and amount of total loss are often predictable with a high degree of precision, we cannot know in advance which individuals will have the claims.

For any individual, or class of individuals, the expected loss during a specified period (usually one year) is called the loss pure premium, or simply the pure premium. Data available to estimate the pure premium for such a given future period generally consists of the claim experience for a population (or large sample from the population) over a given past period. On the basis of this past data, an estimate of the pure premium is obtained by calculating the observed average claim cost, *i.e.*

$$\text{Estimated Pure Premium} = \frac{\text{Total Losses}}{\text{Total Exposures}}$$

An exposure represents an insured risk over the time period (usually one year) of an insurance contract.

Loosely speaking, if the number of exposures in our data set is very large and the claim generation process is stable over time, this estimate should be very precise. Suppose, however, that we sub-divide the total data set into groups on the basis of one or more variables.

For example, suppose we divide drivers of automobiles by age (under 25 or over 25) and by sex. Then each age-sex combination, or cell, will have a smaller number of exposures. Some cells may have very few exposures, leading to rather imprecise estimates in this cells.

The problem of estimation based on limited data is termed by actuaries the creditability problem. The sample size being used may not be sufficient to allow a credible estimate of the indicated pure premium for a particular class. In the case of automobile insurance, it is traditional to cross-classify risks on the basis of two factors. **First,** drivers are divided according to a driver class variable formed from a complex combination of individual characteristics. **Second,** automobiles are also divided into discrete territories according to where they are principally garaged which is reflecting variations in traffic conditions and other risk factors related to location, such as theft rate.

To address the problem of small cells, actuaries assume a certain structure to the effects of driver class and territory. They assume that the effect of being in a particular class or in a particular territory is to modify the expected pure premium by a constant multiplicative factors. These factors are called the relativities corresponding to the driver class or the territory.

Because the estimated relatives are based on all the exposures in a territory or class, these estimates will generally have much higher credibility than those based on the individual cells [1]. So, it is thought that an estimate based on the overall loss pure premium for the population and the relativities will be more accurate than the raw individual cell estimates. That is, for an individual in territory i and class j, the estimate would be,

$$\begin{array}{l} \text{Pure Premium for} \\ \text{particular territory} \\ \text{and driver class} \end{array} = \begin{array}{c} \text{Overall average} \\ \text{premium} \end{array} * \begin{array}{c} \text{Driver class} \\ \text{relativity} \end{array} * \begin{array}{c} \text{Territorial} \\ \text{relativity} \end{array}$$

In automobile rate setting the problem of pricing insurance to take account of driver characteristics is thus reduced to the problem of estimating the appropriate relativities based on recent data [1]. Actuaries have developed pragmatic techniques for determining these relativities that have been used with only minor variations for several decades.

The objectives of this paper are therefore: **first,** to analyze the traditional actuarial methodology from a statistical viewpoint. By a statistical viewpoint, we are referring to the general idea of specifying a simple stochastic model for the process of claim generation, and regarding the basic problem as of estimating the model parameters; **second,** to modify this method to suit a free compound stochastic model.

Methodology

The most straight forward approach is to estimate the territorial relativity x_i and the driver class relativity y_j by the ratios of observed class mean to overall pure premium, *i.e.*

$$P_{i.} = \sum_j \frac{n_{ij}\ p_{ij}}{n_{i.}} \tag{1}$$

$$P_{.j} = \sum_i \frac{n_{ij}\ p_{ij}}{n_{.j}} \tag{2}$$

and the overall mean,

$$P_{..} = \sum_i \sum_j \frac{n_{ij}\ p_{ij}}{n_{..}} \tag{3}$$

The "naive" estimates of the territorial relatives x_i and driver class relatives y_j are

$$x_i = \frac{P_{i.}}{P_{..}} \tag{4}$$

$$y_j = \frac{P_{.j}}{P_{..}}$$

This method is simple and intuitively appealing, but has long been recognized to have serious problems.

Intuitively the problem in using x is related to the fact that the value of $p_{i.}$ is dependent upon the distribution of driver types within the territory. Assuming that territory *perse* has a constant multiplicative effect regardless of driver class, and the value of $p_{i.}$ relative to $p_{..}$ will be larger if the territory contains relatively more "bad drivers" and smaller if it contains a disproportionate number of good drivers. In statistical language, if there is an interaction of driver class and territory in the distribution of exposures the estimated relativities for a territory will be indirectly influenced by the true driver class relativities, and vice versa. As a result, the average estimated pure premium for a given territory will not equal observed average pure premium. In actuarial language this condition is termed lack of "balance".

In mathematical terms, if we calculate the average estimated pure premium across a given territory i, we will obtain

$$\hat{p}_{i.} = \sum_j x_i \, y_j \, p_{..} \; \frac{n_{ij}}{n_{..}}$$

$$= \sum_j \frac{p_{i.} \, p_{.j} \, n_{ij}}{p_{..} \, n_{..}}$$

$$= p_{i.} \left[\sum_j \frac{p_{.j}}{p_{..}} \; \frac{n_{ij}}{n_{i.}} \right]$$

Suppose first that the distribution of exposures to classes and territories is independent, so that

$$\frac{n_{ij}}{n_{i.}} = \frac{n_{.j}}{n_{..}} \qquad (6)$$

Then, it is easy to show that the factor parentheses reduces to 1.0 [1]. In generaly, however, this factor will be greater than 1.0 if territory i contains a relatively bad mix of drivers, and smaller if it contains a relatively good mix of drivers. So this factor is a kind of "rating factor" reflecting the relative quality of drivers within the territory. Unless this value factor has the value 1.0, the relatives will not produce estimates that are balanced with respect to territories, Similarly, we have

$$\hat{p}_{.j} = p_{.j} \left[\sum_i \frac{p_{.j}}{p_{..}} \; \frac{n_{ij}}{n_{.j}} \right] \qquad (7)$$

where the factor in parentheses represents a rating factor reflecting the mix of territories within each driver class. Thus, the native method results in estimates that are not balanced by either territory or driver class. Moreover, it is easily shown that they are not even balanced overall, *i.e.*

$$\frac{\sum_i \sum_j x_i \; y_j \; p_{..}}{n_{..}} = p_{..}$$

From an actuarial perspective, this lack of balance means that some identifiable subgroups of the total exposure population are apparently being systematically over charged or under charged. While any system of rates must necessarily mischarge some insureds, the system should be unbiased in the sense that such over-charges and under-charges are essentially random and will tend to even out over time.

The problem of imbalance results from the different distributions of driver class within territory for different territories, and vice-versa. Thus it has occured to actuaries that some "mix-of-business" adjustment is necessary to bring the estimate into balance.

Adjusting the Naive Estimates

In 1979, Chang and Fairly [2], provided a clear explanation of the traditional multiplicative method. The traditional driver-class relativity is the ratio of the observed driver-class average claim amount per exposure to the observed statewide average claim amount per exposure. The traditional territorial relativity, adjusted for mix of driver class exposures in the territory, is the ratio of the adjusted territorial average claim amount per exposure to the adjusted statewide average claim amount per exposure. An adjusted territorial average claim amount per exposure is the observed territorial average divided by the "average rating factor" in the territory. The "average rating factor" in the territory is the weighted average of the diver-class relatives weighted by the driver-class exposure distribution in that territory. Moreover, they present a formula that can be written in our notation as

$$x_i = \frac{p_{i.}}{P_{..}} \Big/ \sum \frac{p_{.j}}{p_{..}} \; \frac{n_{ij}}{n_{i.}}$$

$$y_j = p_{.j} / p_{..} \tag{8}$$

Referring to (5) and (7) we see that this adjusted estimate simply divides the naive estimate of the territorial relativity of the average rating factor, and it is clear that this adjusted relativity results in cell estimates that are balanced for each territory i.

What about the driver classes? Note that if we average across class j we obtain

$$\hat{p}_{.j} = \sum_i x_i \; y_i \frac{n_{ij}}{n_{.j}} \; p_{..}$$

$$= p_{.j} \left[\sum_i x_i \frac{n_{ij}}{n_{.j}} \right] \tag{9}$$

The imbalance for driver class i now determined by the rating factor for class j, using the adjusted relativities defined by (8). So the estimates will be balanced for territories but not for driver classes.

While these estimates are easy to compute and rate balanced in one dimension, it is possible that they may be substantially imbalanced in the other. Certainly, it would seem preferable to adjust simultaneously for both territory and class rating factors, to produce a set of relativities resulting in estimates balanced for territories and classes [8].

The first hint of this possibility in the actuarial literature occurs in Baily and Simon's important 1960 paper [3]. Then, in his much quoted 1963 article [4], Baily explicitly proposed a method he called the "minimum bias" approach, which involve the solution of the system of equations.

$$\sum_j x_i \; y_j \frac{n_{ij}}{n_{i.}} \; p_{..} = p_{i.}$$

$$\sum_i x_i \; y_j \frac{n_{ij}}{n_{.j}} \; p_{..} = p_{.j} \tag{10}$$

or equivalently

$$x_i = \frac{p_{i.}}{p_{..}} \Big/ \sum_j y_j \frac{n_{ij}}{n_{i.}}$$

$$y_j = \frac{p_{.j}}{p_{..}} / \sum_i x_i \frac{n_{ij}}{n_{.j}} \tag{11}$$

The system of equations can be solved iteratively, starting with the naive estimates for the y, computing the x using (11) to obtain a new values for the y and continuing until the estimates converge.

A Simple Stochastic Model of Claim Generation

First, we shall adopt a very simple model of the claim generation process [1]. The frequency of accidents for each driver during the exposure will be assumed to follow a poisson distribution with mean depending only on the territory i and the driver class j. Further we assume that the claim cost, or severity, associated with each accident has a very small variance, so that it has effectively a constant values S. Finally, we define to be the expected value of the pure premium for drivers in cell (ij) and f to be the observed accident frequency in cell (i,j).

Then

$$\prod_{ij} = E(p_{ij}) = \frac{S.E(f_{ij})}{n_{ij}} = S\lambda_{ij} \tag{12}$$

Further, in keeping with the idea of multiplicative effects of territory and driver class, we assume

$$\lambda_{ij} = \alpha_i \ \beta_j \tag{13}$$

With this model it is now possible in principle to apply any standard statistical methods, such as maximum likelihood, minimum chi-square, and least-squares, to derive parameter estimates and therefore estimates of the "true" pure premium.

In 1981, Wiesberg and Tomberlin [5], developed an easy and effective approach to deal with the problem of estimating α and β. Assuming that S is effectively fixed and known, it will be easy by using the maximum likelihood method to get the likelihood function for the above model, taking the log for this function and differentiating it with respect to the α's and β's, we can get the following system of equations:

$$\frac{\partial \log L}{\partial \alpha_i} = \sum_j \frac{f_{ij}}{\alpha_i} - \sum_j \beta_j \ n_{ij} \tag{14}$$

$$\frac{\partial \log L}{\partial \beta_j} = \sum_i \frac{f_{ij}}{\beta_j} - \sum_i \alpha_i \; n_{ij}$$

Setting these equal to zero yields the following system of equations.

$$\alpha_i = \frac{\sum_j f_{ij}}{\sum_j \beta_j n_{ij}} \tag{15}$$

$$\beta_j = \frac{\sum_i f_{ij}}{\sum_i \alpha_i n_{ij}} \tag{16}$$

These equations can be solved iteratively. We can start with, say, the naive estimates for β and solve for the α using equation (15). These new values can then be substituted into (16) to solve for the β, and so on.

Computing the Moments of S when the Severity is Random

Suppose that the random variable S, representing *e.g.* the total amount of claims in cell (ij), which may be written as

$$S = \sum_{i=1}^{N} x_i \tag{17}$$

where $x_1, x_2, \ldots$ are i.i.d. random variables (claim-size), independent of the random variable N [f = the actual number of claims in cell (ij)]. We know two algorithms to compute the moments of S when N is poisson distributed. When λ is the poisson parameter and $p_k = E\, x^k$, according to Shiu [6] we have

$$E(S - \lambda p_i)^k = k! \left\{ \frac{\lambda p_k}{k!} + \frac{\lambda^2}{2!} \sum_{\substack{k_1 + k_2 = k \\ k_1, k_2 \geq 2}} \frac{p_{k_1} p_{k_2}}{k_1 ! k_2 !} + \frac{\lambda^3}{3!} \sum_{\substack{k_1 + k_2 + k_3 = k \\ k_1, k_2, k_j \geq 2}} \frac{p_{k_1} p_{k_2} p_{k_3}}{k_1 ! k_2 ! k_3 !} + \cdots \right\} \tag{18}$$

and Goovaerts [7, p.12], gives a useful recursion formula

$$E(S - \lambda p_i)^{k+1} = \lambda \sum_{t=0}^{k-1} \binom{k}{t} E(S - \lambda p_i)^t \; p_{k+1-t} \tag{19}$$

When the distribution of N is arbitrary, the number of arithmatic operations required for computing E(S) increase with k^3, and the storage needed is proportional to k^2 [8], to do that we can compute the conditional expectations of S given N = n. Observe that by symmetry, Newton's Binomial Theorem and independence, for all n = 0,1 ...

$$E\left(\sum_{i=1}^{n} x_i\right)^k = \sum_{i=1}^{n} ex_i \left(\sum_{j=1}^{n} x_j\right)^{k-1}$$

$$= nEx_n \left(\sum_{j=1}^{n} x_j\right)^{k-1}$$

$$= nEx_n \sum_{t=0}^{k-1} \binom{k-1}{t} X_n^t \left(\sum_{j=1}^{n-1} x_j\right)^{k-1-t}$$

$$= n\sum_{t=0}^{k-1} \binom{k-1}{t} p_{t+1} E\left(\sum_{j=1}^{n-1} x_j\right)^{k-1-t} \qquad (20)$$

Letting $n^{!k} = n(n-1) \ldots. (n-k-1)$, it is easy to show that coefficients a,j = 1,2, ... , k ; k = 1,2, ... exist, for all n = 1,2, ...

$$E\left(\sum_{i=1}^{n} x_i\right)^k = \sum_{j=1}^{k} a_{jk}\, n^{!j} \qquad (21)$$

If we take $a_{1k} = p_k$ and for j = 2,3, ..., k

$$a_{jk} = \sum_{t=0}^{k-j} \binom{k-1}{t} p_{t+1}\, a_{j-1,k-1-t} \qquad (22)$$

Using (21) we directly obtain

$$E(S^k) = \sum_{n=0}^{\infty} P(N = n)\, E(S^k \mid N = n)$$

$$= \sum_{n=0}^{\infty} P(N = n) \sum_{j=1}^{k} a_{jk}\, n^{[j]} \tag{23}$$

$$= \sum_{n=1}^{k} a_{jk}\, E(N^{[j]})$$

The coefficients a_{jk} in (23) are computed using (22); the factorial moments of N can be computed from the ordinary moments, but in fact often are more easily calculated themselves. In Janardan [9] one finds expressions for factorial and ordinary moments of many counting distributions, including those used in actuarial work.

Conclusion

The purpose of this paper has been to adopt and modify the traditional actuarial methodology for calculating pure premium on the basis of cross-classified data as a simple actuarial approach for automobile rating. This approach entails the use of row and column relativities to deal with the problem of cells that contain small number of exposures.

To illuminate the properties of this methodology, we have considered Weisberg and Tomas's simple stochastic model for the claim generation [1], we modify and extend the use of this model to suit any arbitrary compound distributed severity S, to do this we depend on the first few moments to compute the higher order moments for the cells claims and also the total claims. One of the main reasons one might be interested in computing moments of S, can be found when we have incomplete information concerning the claim-size distribution.

References

[1] Weisberg, Herbert I. and Tomberlin, Thomas J., "*A Statistical Interpretation of Actuarial Relatives.* Boston, USA: Analysis and Inference, Inc., 1982.

[2] Chans, Lena and Fairley, William B. "Pricing Automobile Insurance Under Multivariate Classification." *The Journal of Risk and Insurance,* XLVI, No. 1 (1979), 75-93.

[3] Bailey, Robert A. and Simon, Leroy J. "Two Studies in Automobile Insurance Rate-making." *Proceedings of the Casualty Actuarial Society,* Vol. XLVII, Part 1 (1960).

[4] Bailey, Robert A. "Insurance Rates with Minimum Bias." *Proceedings of the Casualty Acturial Society,* Vol. L, (1963).

[5] Weisberg, Herbert I. and Tomberlin, Thomas J. *Estimating Pure Premiums from Cross-Classified Data.* Boston, USA: Analysis and Inference, Inc., 1981.

[6] Shiu, E.S.W. "Moments of Two Distributions of Collective Risk Theory." *Scand. Actuarial Journal* (1977), 185-187.

[7] Goovaerts, M.J.; de Vylder, F. and Haezendonk, J. *Insurance Premiums.* North Holland, 1984.

[8] Kass, R. and Goovaerts, M.J. "Computing Moments of Compund Distributions." *Scand. Actuarial Journal* (1985), 185-187.

[9] Janardan, K.G. "Moments of Certain Series Distributions and Their Applications." *SIAM Jour. Appl. Math.* (1984), 854-968.

مدخـل اكتواري مبسط لتسعير تأمين السيارات

أحمـد قمحـاوي أباظـه

أستاذ مساعد، قسم الأساليب الكمية، كلية العلوم الإدارية،

جامعة الملك سعود، الرياض، المملكة العربية السعودية

ملخص البحث. نظرًا للأهمية المتزايدة للسيارات كوسيلة نقل سريعة ومستقلة، وبالتالي الأهمية المواكبة لها والمصاحبة لايجاد غطاء تأميني يقلل ـ بقدر المستطاع ـ من الآثار الناتجة من حوادث هذه السيارات. . فإن البحث يركز على عرض الطريقة الاكتوارية بمعالجة إحصائية وذلك بافتراض نموذج عشوائي بسيط لعملية التعويضات المستقبلية، وكيفية تقدير معلمات هذا النموذج مع مراعاة أخطاء التقدير ـ الناتجة من إعادة تقسيم مجموعة البيانات الأصلية لمجموعات جزئية ـ ثم محاولة تعميم هذه الطريقة بجعلها صالحة للتطبيق على أي نموذج عشوائي مركب.