

1994

Predicting Automobile Insurance Multi-Regional Base Pure Premiums


Edward Nissan

University of Southern Mississippi, nissan@cba.usm.edu

Iskandar S. Hamwi

University of Southern Mississippi, hamwi@cba.usm.edu

Follow this and additional works at: <http://digitalcommons.unl.edu/joap>

 Part of the [Accounting Commons](#), [Business Administration, Management, and Operations Commons](#), [Corporate Finance Commons](#), [Finance and Financial Management Commons](#), [Insurance Commons](#), and the [Management Sciences and Quantitative Methods Commons](#)

Nissan, Edward and Hamwi, Iskandar S., "Predicting Automobile Insurance Multi-Regional Base Pure Premiums" (1994). *Journal of Actuarial Practice 1993-2006*. 157.

<http://digitalcommons.unl.edu/joap/157>

This Article is brought to you for free and open access by the Finance Department at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Journal of Actuarial Practice 1993-2006 by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Predicting Automobile Insurance Multi-Regional Base Pure Premiums

Edward Nissan* and Iskandar S. Hamwi†

Abstract‡

Multi-regional insurance base premiums are customarily computed by a top-down method where national or state projections are adjusted to reflect regional differences. This paper proposes a methodology for a bottom-up projection. A weighing scheme that minimizes the variance of the estimator is suggested as a criterion to establish an overall multi-regional rate.

Key words and phrases: *ratemaking, loss severity, minimum variance, casualty insurance*

1 Introduction

Sometimes it is necessary for an insurance company to determine premium rates for a particular line of business solely on the basis of its

*Edward Nissan, Ph.D., is Professor of Economics and International Business at the University of Southern Mississippi. He has published articles in the areas of regional and urban economics, concentration in U.S. businesses, quality-of-life indexes, and actuarial methodologies in insurance ratemaking.

Dr. Nissan's address is: Department of Economics and International Business, University of Southern Mississippi, PO Box 5072, Southern Station, Hattiesburg, MS 39406. Internet address: nissan@cba.usm.edu

†Iskander S. Hamwi, Ph.D., is Professor of Insurance at the University of Southern Mississippi. He is a past president of the Western Risk and Insurance Association and currently serves as an associate editor of the *Journal of Insurance Issues*.

Dr. Hamwi's address is: Department of Marketing and Finance, University of Southern Mississippi, PO Box 5072, Southern Station, Hattiesburg, MS 39406. Internet address: hamwi@cba.usm.edu

‡Professors Nissan and Hamwi are grateful to the anonymous referees for their many helpful comments and suggestions that have improved considerably the quality and scope of the paper. They also thank the editor for his patience and encouragement. The usual caveat still applies: all errors are, however, the responsibility of the authors.

own loss experience.¹ When the number of insureds within a class or region is too small to rate accurately (without reference to a larger body of data), the general approach is to use experience rating techniques,² of which credibility theory is a special case, to combine loss information belonging to several classes and/or regions. If the data are combined, this larger source of data can be used to determine the current year's experience rated multi-class or multi-regional base pure premium. This base pure premium then must be adjusted to reflect the previous year's pure premium and class or regional differences. Credibility theory can be used to decide on the relative weight to be placed on previous year's pure premium versus this year's base pure premium. The objective of this paper is to provide a way to calculate the current year's multi-class or multi-regional base pure premium and not to decide the way that this base pure premium has to be adjusted to produce a final pure premium for each class or region.

The seminal work on the estimation techniques for premium rates across class or territories was done by Bailey (1963). Bailey suggests the calculation of territorial or class differential rates by iterative approximation to arrive at a set of estimates that provides the best fit and to ensure fairness and equity. In addition, he recognizes and gives formulation to the additive relativities at higher levels of classification to modify basic rates.

A comprehensive study of pricing in the state of Illinois was conducted by Witt (1979) to discover whether the rate intended to cover loss costs, expenses, and underwriting profit margins is adequate and at the same time equitable to consumers. This concern for equity was addressed further by Chang and Fairly (1978, 1979), who discuss the traditional multiplicative and the closely associated log-linear methods. They find both are biased toward some drivers when applied to the State of Massachusetts. They suggest using an additive procedure that they claim would eliminate biases and improve overall accuracy.

Subsequently, Fairley, Tomberlin, and Weisberg (1981), in their study of pricing in New Jersey, address the issues of the merits and drawbacks of the multiplicative and additive methods. They point out that typically only a single state is used during a single period. They recommend the inclusion of more regions and time periods, in a scheme that they demonstrate for the years 1975, 1976, and 1977.

¹Anti-trust considerations in the future may require companies to use their own data in all but the most extreme situations.

²For an overview of experience rating and credibility theory, see, for example, Daykin, Pentikäinen, and Pesonen (1994, pp. 179-189) or Venter (1990, Chapter 7).

In a more recent article, Brown (1988) provides a clear summary description of the multiplicative and the additive approaches. He explains that a driver's rate by the multiplicative approach is obtained by $BR_m \times x_i \times y_j$, while for the same driver the rate would be $BR_a + (x_i + y_j)$, using the additive approach (where BR_m and BR_a are base rates for the multiplicative and the additive portion, respectively, and x_i and y_j are the adjustments, such as the class and driving record of the insured). The difference between the two methods may be simply stated as *percents versus cents* adjustments. Brown goes on to suggest the use of the generalized linear models approach for estimating the components of the multiplicative as well as the additive versions.

The traditional approach in multi-regional ratemaking is excellently summarized by Finger (1990) who provides a variety of examples. The traditional multi-regional approach relies on an iterative procedure employing regional and class relativities to adjust the pure premium for each region. A final iteration that uses base exposures instead of earned exposures produces a convergent rate for all regions and all classes that accurately may represent the historical experience. Note that both regional and class relativities are employed simultaneously to produce a convergent base rate; these, in turn, are adjusted further for higher levels of classification when appropriate.

McClenahan (1990, Chapter 2) believes that the traditional approach of finding a state-wide average rate that subsequently is distributed, using territorial relativities, among the various territories within the state and then, using classifications relativities, among the classes within each territory has worked fairly well in practice.

Excellent comparative assessments of alternative approaches available for predicting multi-regional premiums are provided by Sant (1980), Weisberg and Tomberlin (1982), Weisberg, Tomberlin, and Chatterjee (1984), and Jee (1989). Jee also makes an important contribution by classifying the methodologies according to the functional form of the model and estimation method. The methods of most relevance to this research are those based directly on observed pure premium data, in contrast to those that divide the observed data into frequency and severity components.³

³According to Finger (1990) the pure premium approach, because it requires more information and also can produce frequency, severity and pure premium relativities, is more accurate than the loss ratio method. Under the loss-ratio method, incurred losses are divided by earned premiums; under the pure premium method, incurred losses are divided by number of exposures. Interestingly, Brown (1993) shows that the loss ratio and pure premium methods are algebraically equivalent when used in calculating classification differentials and for changing the average portfolio rate.

2 Objectives

The purpose of this paper is to present a statistical method for estimating multi-regional base premiums for a particular line of insurance: automobile physical damage coverage. Here we have insureds who are in the same class, but who are located in more than one territory or region. The multi-regional base premium is the total dollar amount of claims spread over the whole number of insured persons. The focus is only on that portion of the pure premium that is directly attributable to claims. Neither profit nor expense margins are included in the rate. In its final form, however, the base rate can be adjusted to include a risk margin to ensure that the probability of the total claims exceeding the funds generated by the base premium is less than some specified quantity, such as 5 percent or 1 percent.⁴ This margin will compensate for the variability of the underwriting risk.

The contribution this paper makes is in the implicit use of a sampling prototype methodology akin to stratified sampling, employing observed data to estimate a basic rate that can be adjusted by either multiplicative or additive factors. According to Deming (1950, p. 213), in stratified sampling, random samples are drawn from a universe divided into separate strata or classes. The purpose of stratification is to find out what properties of the various classes govern the variance of the estimate of the mean of the entire universe. Furthermore, it is desired that the estimator be efficient (minimum variance). In this paper, the sampling universe is divided into separate strata by the geographic location (called regions) of drivers. The concern then is to find an overall linearly weighted mean that has the minimum variance.

What remains to be defended next is the assumption that the claims resulting from automobile accidents and by implication, pure premiums, constitute random samples. Support for this assumption comes from Darnell and Evans (1990, p. 13) who explain that the conditions of the world within which data are generated are outside the control of the investigator and therefore do not satisfy the foundation of the classical probability model that requires the assumption of repeated experiments. Economists and social scientists almost exclusively deal with data generated outside such an experimental context.

Darnell and Evans explain that observed economic variables may be treated as if resulting from a single drawing from a population.⁵

⁴When risk margins are determined in this manner, the resulting premium is called a *percentile premium* (Gerber 1979, Chapter 5).

⁵It is precisely this type of argument that Butler (1993) puts forth in an interesting article in which he proposes the use of car-mile exposure rather than car-year as a basis

An advantage of a sampling prototype is that the observations need not follow any particular probability distribution such as, for example, when regression models are employed.

In regression models, the error terms depend on many factors including omission of explanatory variables, model specification, aggregation of variables, and functional misspecifications. The combined effects of these factors may render the coefficients of the least squares regression not to be most efficient because they may lead to false conclusions in hypothesis testing. For example, Brown (1988) mentions the exponential family of probability distributions, each with specific estimators that must be determined from a sample. Similarly, the procedure suggested by Chang and Fairly (1978, 1979) uses regression models that assume the error term follows the customary requirement of normal distribution. Such distributional assumptions are not binding in sampling. With a sufficiently large number of observations, as common in practice, there is good reason to assume, according to Cochran (1953), that the estimators of population parameters such as the mean are approximately normally distributed.

The method proposed in this paper differs from other techniques advocated for multi-regional rate prediction because it uses information on the mean and variance of loss severity for many regions and for several prior years. It uses a minimum-variance criterion to assign yearly weights for regions.

3 The Model

Consider an insurer that sells automobile physical damage coverage to several classes of insureds. Each class of insureds consists of policyholders spread over several regions. It is assumed that the claims generated by a single class of insureds in the same region and the same year are mutually independent and have identical policies.

The following notation is used throughout this paper:

For $i = 1, \dots, L$ and $j = 1, \dots, K$, let us define:

- K = Number of regions, $K = 1, 2, \dots$;
- L = Number of years, $L = 1, 2, \dots$;
- n_{ij} = Number of claims in year i and region j ;
- N_{ij} = Number of insureds in year i and region j ;

of pricing and explains that in the Bailey and Simon (1959, 1960) model, automobile accidents can be envisioned as a random sampling of the class population on the road.

$$\begin{aligned}
 X_{ijk} &= \text{Observed claim severity from the } k\text{th claim} \\
 &\quad \text{in year } i \text{ and region } j, \text{ for } k = 1, 2, \dots, n_{ij}; \\
 \bar{X}_{ij} &= \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} X_{ijk} = \text{observed average claim severity;} \\
 E[\bar{X}_{ij}] &= \mu_{ij}; \\
 s_{ij}^2 &= \frac{1}{n_{ij} - 1} \sum_{k=1}^{n_{ij}} (X_{ijk} - \bar{X}_{ij})^2 \\
 &= \text{observed variance of claim severity;} \\
 E[s_{ij}^2] &= \sigma_{ij}^2, \\
 \bar{P}_{ij} &= \frac{1}{N_{ij}} \sum_{k=1}^{n_{ij}} X_{ijk} = \text{Observed pure premium;} \\
 E[\bar{P}_{ij}] &= \frac{n_{ij}}{N_{ij}} \mu_{ij}; \\
 a_{ij} &= \text{Weight for year } i \text{ and region } j, a_{ij} \geq 0; \text{ and} \\
 b_i &= \text{Weight for year } i, b_i \geq 0; \text{ and} \\
 w_j &= \text{Weight for region } j, w_j \geq 0.
 \end{aligned}$$

Note that for each year, the a_{ij} weights sum to one, i.e., $\sum_j a_{ij} = 1$ while $\sum_i b_i = 1$ and $\sum_j w_j = 1$.

Clearly, the variance of \bar{X}_{ij} for year i and region j is

$$\text{Var}[\bar{X}_{ij}] = \frac{\sigma_{ij}^2}{n_{ij}} \quad (1)$$

while the variance of the observed pure premium per insured is

$$\text{Var}[\bar{P}_{ij}] = \frac{n_{ij}}{N_{ij}} \sigma_{ij}^2. \quad (2)$$

The multi-regional base premium for year $L + 1$, $\bar{P}(L + 1)$, is determined by a linear combination of *all* of the observed pure premiums \bar{P}_{ij} (across all regions and for the preceding L years). Thus $\bar{P}(L + 1)$ can be written in its most general form as

$$\bar{P}(L + 1) = \sum_{i=1}^L \sum_{j=1}^K c_{ij} \bar{P}_{ij} \quad (3)$$

where the c_{ij} s are general nonnegative weights that sum to one. In Section 4 we will describe procedures for choosing the weights.

4 Determination of Base Premiums

4.1 Independent Regions

The assumption of statistical independence among regions can be defended on the grounds that regions are physically separate and, therefore, what occurs in one region will have no bearing on what occurs in another region. Except in certain infrequent occurrences like highway pile-ups and natural disasters such as floods and earthquakes, it is unlikely for automobile physical damages to be statistically dependent events. Furthermore, independence assumption routinely is made in many statistical studies for the sake of simplicity. Another assumption that is made is that the group of insureds is homogeneous. According to Tiller (1990, p. 91), "While homogeneity is the goal of manual ratemaking, it is not usually possible to achieve."

In the case where insureds come from the same class but are located in different territories, however, there is greater degree of homogeneity among them in terms of expectation of loss than among insureds who belong to different classes and different territories. When, in one year, a group of insureds can be considered homogeneous for the purpose of auto physical damage coverage, such homogeneity is likely to continue over time.

To assign proper weights a_{ij} to each region, let \bar{P}_i be the weighted observed pure premiums across regions for year i , i.e.,

$$\bar{P}_i = \sum_{j=1}^K a_{ij} \bar{P}_{ij}, \quad \text{for } i = 1, \dots, L, \quad (4)$$

where $\sum_j a_{ij} = 1, j = 1, \dots, K$. Under the independence assumption, its variance is given by

$$\text{Var}[\bar{P}_i] = \sum_{j=1}^K a_{ij}^2 \frac{n_{ij}}{N_{ij}^2} \sigma_{ij}^2, \quad \text{for } i = 1, \dots, L. \quad (5)$$

In a similar manner, multi-regional base premium for year $L + 1$, $\bar{P}(L+1)$, is defined as a weighted average of the \bar{P}_i s by assigning weights b_i such that $\sum_i b_i = 1$. It follows that

$$\bar{P}(L+1) = \sum_{i=1}^L b_i \bar{P}_i, \quad (6)$$

with mean and variance given by

$$E[\bar{P}(L + 1)] = \sum_{i=1}^L b_i \sum_{j=1}^K a_{ij} \frac{n_{ij}}{N_{ij}} \mu_{ij} \tag{7}$$

$$\text{Var}[\bar{P}(L + 1)] = \sum_{i=1}^L b_i^2 \left[\sum_{j=1}^K a_{ij}^2 \frac{n_{ij}}{N_{ij}^2} \sigma_{ij}^2 \right]. \tag{8}$$

It should be noted that both n_{ij} and N_{ij} are treated as given constants rather than random variables as is done in some models.⁶ Here, it is assumed that in each year the population N_{ij} is fixed, and a corresponding sample n_{ij} is drawn.

Consider the linear estimator given in equation (4). Its estimated variance is given by:

$$\text{est. Var}[\bar{P}_{i.}] = \sum_{j=1}^K a_{ij}^2 \frac{n_{ij}}{N_{ij}^2} s_{ij}^2.$$

By the constraint imposed on the weights whereby $\sum_j a_{ij} = 1$, we can eliminate a_{iK} to get

$$\text{est. Var}[\bar{P}_{i.}] = \sum_{j=1}^{K-1} a_{ij}^2 \frac{n_{ij}}{N_{ij}^2} s_{ij}^2 + (1 - \sum_{j=1}^{K-1} a_{ij})^2 \frac{n_{iK}}{N_{iK}^2} s_{iK}^2. \tag{9}$$

Differentiating equation (9) with respect to a_{ij} , for $j = 1, \dots, K - 1$ and setting each of the $K - 1$ equations equal to zero to satisfy the first order condition for minimization yields the following linear system of equations: for $j = 1, 2, \dots, K - 1$

$$\frac{n_{iK}}{N_{iK}^2} s_{iK}^2 \sum_{r=1}^{K-1} a_{ir} + a_{ij} \frac{n_{ij}}{N_{ij}^2} s_{ij}^2 = \frac{n_{iK}}{N_{iK}^2} s_{iK}^2. \tag{10}$$

In spite of its initial appearance, the set of equations (10) easily can be solved by simple row operations. Let

$$\alpha_{ij} = \frac{n_{ij}}{N_{ij}^2} s_{ij}^2.$$

It easily can be proved that the solution to the system of equations (10) is

$$a_{ij}^* = \frac{\alpha_{iK}}{\alpha_{ij}} \left(1 + \sum_{j=1}^{K-1} \frac{\alpha_{iK}}{\alpha_{ij}} \right)^{-1} \quad \text{for } j = 1, 2, \dots, K - 1 \tag{11}$$

⁶Mercer (1985), Stroinski and Currie (1989), and Langford and Capella (1994) use models where n_{ij} and N_{ij} are treated as random variables.

with $a_{iK}^* = 1 - \sum_{j=1}^{K-1} a_{ij}^*$.

Note that the matrix of coefficients of the a_{ij} s arising from the system of equations (10) is positive definite, and by implication it is strictly convex. Hence, the second order condition for global minimum is assured. That is, the a_{ij}^* s do minimize $\text{est. Var}[\bar{P}_i]$; see Hadley (1964, pp. 83-93).

The discussion thus far has been limited to obtaining the regional weights for the yearly averages of premiums. To establish a criterion by which the yearly weights b_i in equation (6) are chosen, information from several prior years is utilized. This is in line with Jee (1989) who suggests the incorporation of trends in the projected estimate. A scheme that fulfills this suggestion is one that uses the standardized ratio of consecutive observed yearly means as follows: let

$$\beta_1 = 1 \quad \text{and} \quad \beta_i = \frac{\bar{P}_i}{\bar{P}_{i-1}}, \quad \text{for } i = 2, \dots, L.$$

The yearly weights b_i are determined as

$$b_i = \frac{\beta_i}{\sum_{i=1}^L \beta_i} \quad \text{for } i = 1, 2, \dots, L. \quad (12)$$

The rate-maker may use other weighing schemes. For example, the rate-maker can recognize the full effect of inflation by adjusting prior years' observed premiums by an appropriate cost index. In the case of automobile physical damage, for example, the index for repair costs would be an appropriate choice. On the other hand, if the rate-maker wants to attach greater importance to the more recent experience, a weighting scheme that assigns larger weights to more recent years than to earlier years is appropriate.

An empirical example to demonstrate the computations is provided in the appendix. The example pertains to automobile physical damage coverage. The procedure can be applied to other insurance coverages, however, with similar aspects.

4.2 Dependent Regions

Suppose that after a series of statistical tests that measure the degree of statistical association for bivariate data,⁷ it is found that the regions are dependent. Then, following Cardoso (1993), we can compute

⁷See, for example, Rohatgi (1984, pp. 762-771) or Sachs (1984, Chapter 5) for examples of such tests.

a projected multi-regional base premium by using the past annually observed pure premiums for each of the K regions. This ensures that the spatial dependencies among regions are considered because all regions may be affected by the same economic factors as well as changes in accident frequency. A procedure that takes regional dependency into account not only requires the K regional pure premium means and variances, but also requires the computation of the covariances between the $\binom{K}{2}$ regional pairs.

The following are the suggested steps:

Step 1: Tabulate the \bar{P}_{ij} s and their mean and variance across the years. A schematic representation of the set of data is given in Table 1.

Table 1
Yearly Pure Premiums By Region

Year	Region 1	Region 2	...	Region K
1	P_{11}	P_{12}	...	P_{1K}
2	P_{21}	P_{22}	...	P_{2K}
\vdots	\vdots	\vdots		\vdots
L	P_{L1}	P_{L2}	...	P_{LK}
Mean	$\bar{P}_{.1}$	$\bar{P}_{.2}$...	$\bar{P}_{.K}$
Variance	ν_1^2	ν_2^2	...	ν_K^2

Step 2: Compute the means, variances, and correlation coefficients for $j, k = 1, \dots, K$:

$$\bar{P}_{.j} = \frac{1}{L} \sum_{i=1}^L P_{ij} \tag{13}$$

$$\nu_{jj} = \nu_j^2 = \frac{1}{L-1} \sum_{i=1}^L (P_{ij} - \bar{P}_{.j})^2 \tag{14}$$

$$\nu_{jk} = \frac{\sum_{i=1}^L (P_{ij} - \bar{P}_{.j})(P_{ik} - \bar{P}_{.k})}{(L-1)} \quad \text{for } j \neq k; \tag{15}$$

Step 3: The projected multi-regional base premium is obtained as a linear combination of the separate K regional means

$$\bar{P}(L+1) = \sum_{j=1}^K w_j \bar{P}_{.j} \tag{16}$$

with observed variance estimated by

$$\text{est. Var}[\bar{P}] = \sum_{j=1}^K w_j^2 v_j^2 + 2 \sum_{j=1}^{K-1} \sum_{m=j+1}^K v_{jm} w_j w_m. \quad (17)$$

Step 4: Derive the set of weights $\{w_j^*\}$ that minimize equation (17) subject to the condition $\sum_{j=1}^K w_j^* = 1$. Let \mathbf{V} be the variance-covariance matrix given by

$$\mathbf{V} = \{v_{ij}\} \text{ for } i, j = 1, 2, \dots, K,$$

and \mathbf{w} be the column vector of weights $\{w_j\}$. Our problem is as follows:

$$\min \mathbf{w}^T \mathbf{V} \mathbf{w} \text{ subject to } \sum_{j=1}^K w_j = 1.$$

Using the method of Lagrange multipliers, we have

$$\min \mathcal{L} = \mathbf{w}^T \mathbf{V} \mathbf{w} - \lambda (\mathbf{w}^T \mathbf{1} - 1)$$

where $\mathbf{1}$ is a column vector of 1s. Differentiating \mathcal{L} with respect to the w_j s and the λ and then setting these derivatives to zero yields:

$$\begin{pmatrix} v_1^2 & v_{12} & \cdots & v_{1K} & -1 \\ v_{21} & v_2^2 & \cdots & v_{2K} & -1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ v_{K1} & v_{K2} & \cdots & v_K^2 & -1 \\ 1 & 1 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_K \\ \lambda \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}. \quad (18)$$

Once the variances and correlations have been calculated, the system of equations (18) can be solved using standard numerical methods such as Cramer's Rule or Gaussian elimination; see, for example, Burden and Faires (1985, Chapter 6). See the appendix for an example.

If any of the w_j^* s found is negative, then we must solve the system of equations given in equation (17) subject to the following additional constraints: $w_j \geq 0$, for $j = 1, 2, \dots, K$. This is now a basic quadratic programming problem.⁸

⁸For more on quadratic programming see Rao (1978, Chapter 12.4). In addition, there is commercial software available through International Mathematical Subroutine Library (IMSL), Visual Numerics, Inc., Houston, Tex., to solve quadratic programming problems.

As an alternative to quadratic programming, we can derive approximate minimum variance weights based on those a_{ij}^* s that were derived in equation (11). Define the regional weights as

$$\bar{w}_j = \sum_{i=1}^L \frac{a_{ij}^*}{L}. \quad (19)$$

These weights then are used in equation (16) to determine the overall base premium.

An empirical example to demonstrate the procedures outlined above is provided in the appendix.

5 Advantages of Proposed Methodology

The approach discussed in Section 4 and the accompanying empirical examples of the appendix are beneficial in a variety of ways as outlined below:

- The calculated overall base rate can be used as an alternative indication in credibility considerations. As explained by McClenahan (1990), a credibility-weighted indication is desirable when a rate is less than fully credible. Thus, for instance, if a rate for a specific class is established by a traditional (manual) method, which, in the assessment of the actuary, is not fully credible, then a complementary rate may be advisable in computing a credibility-weighted indication.
- When trending pure premiums, the most commonly used models according to Cardoso (1993) are the linear model given by $P = a + bt$ and the exponential or log-linear model given by $P = ae^{bt}$, where a and b are constant and t is a time trend. In either case, however, the data used to calculate the pure premium may contain significant serial correlation, making tests of hypotheses for the significance of the regression coefficients invalid (Dougherty, 1992). In our method, this correlation is used in Section 4.2 to project a base rate for year $L + 1$ instead of using a linear trend equation.
- Our proposed method also can be used to calculate an overall average rate for several territories, regardless of the number of classes included in each territory.

- The suggested approach, for the case of dependence (see Section 4.2), takes into account the variation existing within and among regions through variances and covariances, while in the case of independence (see Section 4.1) it takes into account only within-regional variation through variances. In both cases, the variation is embodied in the final rate through the regional weights.

6 Summary

A statistical method has been presented for estimating a multi-regional base premium rate for a class of insureds who are located in different regions. When all of the regions are incorporated in the analysis, the method generates a country-wide rate for a particular class of insureds. Of course, in some insurance lines the grouping of risks into separate regional schedules is as important in the interest of equity as grouping them into different classifications, as pointed out recently by Harrington and Doerpinghaus (1993). Nonetheless, we have focused only on determining an average class rate for all regions combined.

The prevailing practice has been to develop a state-wide rate that is subsequently adjusted using relativities first among the various regions within the state and then among the classes within each region. In contrast, the proposed method finds a multi-regional class rate first that later can be adjusted to reflect any possible regional differences in loss experience. With respect to automobile physical damage insurance, members of the same class who are located in different geographic areas are likely to represent a more homogeneous group than those who belong to different classes within the same region. Homogeneity imparts statistically reliable experience that should allow for the determination of a fairly accurate rate. Interclass subsidies, characteristic of the current system (particularly between rural and city dwellers), would be minimized, if not eliminated, because the proposed method starts with a multi-regional class rate. This bypasses the step of having to calculate an overall state-wide rate for all classes.

References

- Bailey, R.A. and Simon, L.J. "An Actuarial Note on the Credibility of Experience of a Single Private Passenger Car." *Proceedings of the Casualty Actuarial Society* 46 (1959): 159-164.

- Bailey, R.A. and Simon, L.J. "Two Studies in Automobile Insurance Rate-making: Section A. Effectiveness of Merit Rating and Class Rating." *Proceedings of the Casualty Actuarial Society* 47 (1960): 1-7.
- Bailey, R.A. "Insurance Rates with Minimum Bias." *Proceedings of the Casualty Actuarial Society* 50 (1963): 4-11.
- Brown, R.L. "Minimum Bias with Generalized Linear Models." *Proceedings of the Casualty Actuarial Society* 75 (1988): 187-217.
- Brown, R.L. "On the Equivalence of the Loss Ratio and Pure Premium Methods of Determining Property and Casualty Rating Relatives." *Journal of Actuarial Practice* 1 (1993): 97-110.
- Burden, R.L. and Faires, J.D. *Numerical Analysis*. 3rd Edition. Boston, Mass.: Prindle, Weber & Schmidt, 1985.
- Butler, P. "Cost-Based Pricing of Individual Automobile Risk Transfer: Car-Mile-Exposure Unit Analysis." *Journal of Actuarial Practice* 1 (1993): 51-66.
- Cardoso, R.A. "Estimating the Effect of Statutory Changes on Insured Casualty Losses Using Generalized Indicator Variables." *Journal of Actuarial Practice* 1 (1993): 127-140.
- Chang, L. and Fairley, W.B. "An Estimation Model for Multivariate Risk Classification." In *Automobile Insurance Risk Classification: Equity and Accuracy*. Boston, Mass.: Massachusetts Division of Insurance, 1978: 25-55.
- Chang, L. and Fairley, W.B. "Pricing Automobile Insurance under Multivariate Classification of Risks: Additive versus Multiplicative." *Journal of Risk and Insurance* 46 (1979): 75-98.
- Cochran, W.G. *Sampling Techniques*. New York: Wiley, 1953.
- Darnell, A.C. and Evans, J.L. *The Limits of Econometrics*. Hants, UK: Edward Elgar, 1990.
- Daykin, C.D., Pentikäinen, T. and Pesonen, M. *Practical Risk Theory*. London: Chapman & Hall, 1994.
- Deming, W.E. *Some Theory of Sampling*. New York: Dover, 1950.
- Dougherty, C. *Introduction to Econometrics*. New York: Oxford University Press, 1992.
- Fairley, W.B., Tomberlin, T.J. and Weisberg, H.I. "Pricing Automobile Insurance Under a Cross-Classification of Risks: Evidence from New Jersey." *Journal of Risk and Insurance* 48 (1981): 505-514.
- Finger, R.J. "Risk Classification," Chapter 5 of *Foundations of Casualty Actuarial Science*. Arlington, Va.: Casualty Actuarial Society, 1990.

- Hadley, G. *Nonlinear and Dynamic Programming*. Reading, Mass.: Addison-Wesley, 1964.
- Harrington, S.E. and Doerpinghaus, H.I. "The Economics and Politics of Automobile Insurance Rate Classification." *Journal of Risk and Insurance* 60 (1993): 59-84.
- Jee, B. "A Comparative Analysis of Alternative Pure Premium Models in the Automobile Risk Classification System." *Journal of Risk and Insurance* 56 (1989): 434-459.
- Langford, B.E. and Capella, L.M. "A Model of Personal Automobile Insurance Shopping Intentions (PAPSI)." *Journal of Insurance Issues* 17 (1994): 1-21.
- McClenahan, C.L. "Ratemaking." Chapter 2 of *Foundations of Casualty Actuarial Science*. Arlington, Va.: Casualty Actuarial Society, 1990.
- Mercer, A. "A Decision Support System for Insurance Marketing." *European Journal of Operational Research* 20 (1985): 10-16.
- Rao, S.S. *Optimization: Theory and Methods*. New Delhi, India: Wiley Eastern Ltd., 1978.
- Rohatgi, V.K. *Statistical Inference*. New York: Wiley, 1984.
- Sachs, L. *Applied Statistics: A Handbook of Techniques*. 2nd Edition. New York: Springer-Verlag, 1984.
- Sant, D.T. "Estimating Expected Losses in Auto Insurance." *Journal of Risk and Insurance* 47 (1980): 133-151.
- Stroinski, K.J. and Currie, I.D. "Selection of Variables for Automobile Insurance Rating." *Insurance: Mathematics and Economics* 8 (1989): 35-46.
- Tiller, M.W. "Individual Risk Rating." Chapter 3 of *Foundation of Casualty Actuarial Science*. Arlington, Va.: Casualty Actuarial Society, 1990.
- Venter, G.G. "Credibility." Chapter 7 of *Foundation of Casualty Actuarial Science*. Arlington, Va.: Casualty Actuarial Society, 1990.
- Weisberg, H.I. and Tomberlin, T.J. "A Statistical Perspective on Actuarial Methods for Estimating Pure Premiums from Cross-Classified Data." *Journal of Risk and Insurance* 49 (1982): 539-563.
- Weisberg, H.I., Tomberlin, T.J. and Chatterjee, S. "Predicting Insurance Losses Under Cross-Classification: A Comparison of Alternative Approaches." *Journal of Business and Economic Statistics* 2 (1984): 170-178.

Witt, R.C. "An Economic Appraisal of Territorial Pricing Issues in Automobile Insurance." *Journal of Risk and Insurance* 46 (1979): 33-60.

Appendix: An Empirical Example

Tables A1, A2 and A3 show the data supplied by an insurance company for nine years ($L = 9$) and for regions denoted by 1, 2, and 3 respectively ($K = 3$). These data are used to illustrate the procedures outlined in the paper. For each region, the data given are the number of claims (n_{ij}), the number of exposures (N_{ij}), the average collision claim (C_{ij}), the standard deviation of claims (s_{ij}), and the pure premium (P_{ij}).

Table A1
Summary Data for Region 1

Region 1 (Indianapolis and Gary, Indiana)					
Year	n_1	N_1	C_1	s_{i1}	P_1
1	411	1,279	735	1,411	236.19
2	398	1,462	785	1,562	213.70
3	364	1,518	867	1,635	207.90
4	447	1,618	855	1,880	236.21
5	464	1,505	856	1,703	263.91
6	260	1,107	811	1,642	190.48
7	178	924	828	1,618	159.51
8	181	798	815	1,536	184.86
9	168	828	819	1,658	166.17
Mean					206.55

Source: Data from a major automobile insurance company for unmarried, male principle operators under 21.

Table A4 shows the derived weights a_{ij} (from equation (11)), b_i (from equation (12)), the yearly averages \bar{P}_i (from equation (4)), and the estimated variance $\text{Var}[\bar{P}_i]$ (from equation (5)). Substituting the information from Tables A1, A2, A3 and A4 into equations (6) and (8) yields, under the assumption of independence (See Section 4.1), $\bar{P}(10) = 196.24$ and estimated $\text{Var}[\bar{P}(10)] = 7.24$.

Let R_α be the $100\alpha\%$ upper confidence limit of multi-regional premium, then for say $\alpha = 0.05$ and the normal approximation, the esti-

Table A2
Summary Data for Region 2

Region 2 (Los Angeles and San Francisco, Calif.)					
Year	n_2	N_2	C_2	s_{i2}	P_2
1	1,240	5,690	752	1,601	163.88
2	2,007	9,512	794	1,663	167.53
3	1,969	9,797	861	1,833	173.04
4	2,319	10,215	941	1,906	213.63
5	2,302	9,671	948	1,920	225.65
6	1,859	9,297	1,049	1,941	209.76
7	1,779	8,312	980	1,917	209.75
8	1,274	6,566	909	2,000	176.37
9	1,259	6,769	956	2,036	177.81
Mean					190.73

Source: Data from a major automobile insurance company for unmarried, male principle operators under 21.

mated projected (year 10) multi-regional premium rate is

$$R_{0.05} = 196.24 + 1.645 \times \sqrt{7.24} = 200.67.$$

Under the assumption of dependence (See Section 4.2), the use of Lagrange's multipliers for equations (17) results in the following weights:

$$w_1^* = -0.053, \quad w_2^* = 0.433 \quad \text{and} \quad w_3^* = 0.620.$$

Because one of the weights is negative (w_1 in this case), quadratic programming is used to produce the solution

$$w_1^* = 0.0, \quad w_2^* = 0.448 \quad \text{and} \quad w_3^* = 0.552.$$

Substituting the quadratic programming solution into equations (16) and (17) yields the projected base premium $\bar{P}(10) = 201.19$, the estimated $Var[\bar{P}(10)] = 57.57$, and the estimated projected (year 10) multi-regional premium rate

$$R_{0.05} = 201.19 + 1.645 \times \sqrt{57.57} = 213.67.$$

The third alternative, using average regional weights, as suggested by equation (19), gives the following weights:

$$w_1^* = 0.117, \quad w_2^* = 0.734 \quad \text{and} \quad w_3^* = 0.149,$$

Table A3
Summary Data for Region 3

Region 3 (Rural Mississippi)					
Year	n_3	N_3	C_3	s_{i3}	P_3
1	379	1,592	875	1,772	208.31
2	435	1,842	846	1,861	199.79
3	448	1,907	891	1,868	209.32
4	671	2,906	977	1,945	225.59
5	582	2,464	1,061	1,963	250.61
6	405	1,940	964	2,020	201.25
7	313	1,641	1,018	2,203	194.17
8	313	1,398	1,004	1,893	224.79
9	323	1,844	986	2,135	172.71
Mean					209.61

Source: Data from a major automobile insurance company for unmarried, male principle operators under 21.

giving $\bar{P}(10) = 195.46$ and the estimated $Var[\bar{P}(10)] = 47.18$. Note that this alternative produces a variance that is smaller than the variance obtained by quadratic programming. These variances, however, should not be compared because as equation (19) does not use the covariances while the variance obtained by quadratic programming does. The estimated projected (year 10) multi-regional premium rate is

$$R_{0.05} = 195.46 + 1.645 \times \sqrt{47.18} = 206.76.$$

Table A4
Regional Weights, Yearly Weights, and Multi-Regional Base Premium

Year	a_{i1}^*	a_{i2}^*	a_{i3}^*	b_i	\bar{P}_i	$Var[\bar{P}_i]$	$b_i \times \bar{P}_i$
1	0.140	0.712	0.149	0.111	180.59	69.86	20.05
2	0.106	0.785	0.109	0.108	175.93	48.18	19.00
3	0.123	0.756	0.121	0.115	181.74	52.08	20.90
4	0.095	0.713	0.192	0.133	218.07	57.57	29.00
5	0.109	0.715	0.176	0.119	234.22	64.89	27.87
6	0.107	0.754	0.139	0.098	206.51	61.10	20.24
7	0.129	0.746	0.125	0.108	201.30	70.56	21.74
8	0.128	0.723	0.149	0.102	184.67	85.52	18.84
9	0.118	0.698	0.184	0.106	175.50	79.53	18.60
Mean	0.117	0.734	0.149				\$196.24

Source: Based on calculations from Tables A1 to A3.

Notes: a_{ij}^* is the regional weight for year i and region j and is found from equation (11); b_i is the weight for year i and is found from equation (12); \bar{P}_i is the calculated from equation (4); and $Var[\bar{P}_i]$ is calculated using equation (5).

