THE MINIMUM BIAS PROCEDURE: A PRACTITIONER'S GUIDE

SHOLOM FELDBLUM AND J. ERIC BROSIUS

Abstract

The minimum bias classification ratemaking procedure, introduced by Robert Bailey and LeRoy Simon in 1960, determines rate relativities simultaneously for two or more classification dimensions. This paper summarizes the minimum bias procedure for the practicing actuary and provides the intuition for several bias functions: balance principle, least squares, χ-squared, and maximum likelihood. The exposition is structured around a series of illustrations using a two-dimensional private passenger automobile classification system: male/female and urban/rural.

ACKNOWLEDGMENTS

The authors are indebted to Robert A. Bailey, LeRoy J. Simon, Thomas A. Ryan, Dr. Ernesto Schirmacher, Dr. Doris Schirmacher, Daniel A. Lowen, and Ginda Kaplan Fisher for review of and corrections to this document. E. Schirmacher, D. Lowen, and G. Fisher, in particular, made extensive corrections to both the substance and the style of this paper, and D. Schirmacher provided the software application to run the minimum bias computations. Any remaining errors are the authors' own and should not be attributed to Messrs. Bailey, Simon, Ryan, Schirmacher, Lowen, Ms. Schirmacher, or Ms. Fisher.

1. THE MINIMUM BIAS PROCEDURE

Introduction

This paper is geared to the practicing actuary or actuarial student seeking to optimize classification relativities. It provides the intuition underlying the minimum bias procedure along with simple illustrations to show the computations required for each method.

Background

The minimum bias procedure was first introduced in a 1960 *Proceedings*[1] paper by Robert Bailey and LeRoy Simon, "Two Studies in Automobile Insurance." Bailey and Simon examined models with two types of arithmetic functions (multiplicative and additive), two types of bias functions (balance principle and χ-squared), and two data types (loss costs and loss ratios).

Bailey and Simon used their procedure (i) to judge the merits of an additive versus a multiplicative classification model for Canadian private passenger automobile business and (ii) to choose optimal rate relativities.[2] They discuss the rationale for the minimum bias procedure, the characteristics of a suitable rating model, and the rating scenarios that fit the various types of models. The authors concluded that: (i) the additive model fits the Canada private passenger automobile data better than the multiplicative model, and (ii) the χ-squared function is the optimal bias function. The first conclusion was based on a goodness-of-fit test; the second conclusion was based on the credibility assigned by the χ-squared function.

In a 1963 *Proceedings* paper, "Insurance Rates with Minimum Bias," Robert Bailey summarized the minimum bias theory, outlining the considerations that support the use of the balance principle as the bias function and explaining when loss ratios serve better than loss costs. This paper was on the CAS examination syllabus for many years.

In a 1988 *Proceedings* paper, "Minimum Bias with Generalized Linear Models," Robert Brown expanded the minimum bias

[1]References to the *Proceedings* are to the *Proceedings of the Casualty Actuarial Society*.
[2]The minimum bias procedure deals with loss cost relativities, which we refer to here as pure premium relativities. In practice, actuaries determine rate relativities. The two types of relativities may differ if expenses are not a fixed percentage of premiums.

method to use two additional types of bias function. Brown retained the balance principle and χ-squared functions from the Bailey and Simon papers. He added a least squares function (similar to the χ-squared function) and a maximum likelihood function, which assumes certain distributions of claim frequency or claim severity in the insured population. Brown also examined generalized linear models (GLM), which have potential statistical advantages and may accomplish the same objectives as the minimum bias procedures, though he did not find that they produced more accurate results.[3] For the Canadian private passenger automobile business, Brown found the multiplicative model superior to the additive model.

In 1990, Gary Venter introduced several extensions of the existing procedures in a discussion of Brown's paper, along with an analysis of credibility consideration and other modeling issues. Brown's *Proceedings* paper, along with Venter's discussion, was placed on the CAS actuarial syllabus in the mid-1990s.

These papers have proved difficult for practicing actuaries to understand and for actuarial candidates to master. The authors wrote for experienced actuaries who were familiar with the ratemaking issues and proficient with the statistical models.

This paper combines the theory of the original actuarial papers with the teaching material prepared by the authors and used to teach the minimum bias procedure to several hundred actuarial candidates since 1995. It explains the rationale for the procedure and shows its applications. It presents the method to new actuaries and gives them the background to read the original *Proceedings* papers.

The title of this paper is the "Minimum Bias Procedure," since that name is now common in the U.S. actuarial profession. The subject of this paper should more properly be described as the development of multidimensional classification systems. This sub-

[3]On generalized linear models, see Feldblum et al. [forthcoming].

ject is broad. The paper covers part of this subject, of which one component is the minimum bias procedure and the alternative methods discussed here.

This paper does not cover generalized linear models, which are commonly used in the United Kingdom and in continental Europe for multidimensional classification ratemaking; see instead the companion paper by S. Feldblum, D. Anderson, E. Schirmacher, D. Schirmacher, and N. Thandi [forthcoming], "Generalized Linear Models: A Practitioner's Guide."

2. CLASSIFICATION MODELS

Introduction

Before Bailey and Simon introduced the minimum bias procedure, classification relativities were determined one dimension at a time. This is suitable for a single-dimensional classification system. Workers compensation, for example, uses industry as the only classification dimension within each state. Insurers are now examining other classification dimensions for workers compensation; the minimum bias procedure and generalized linear models may prove valuable in this analysis.

The minimum bias procedure becomes useful when the classification system has multiple dimensions. In this paper, we use examples with two dimensions; the extension to three or more dimensions is straightforward, but the arithmetic and display are cumbersome.

We define the minimum bias terms, explain the statistical procedures, and review the intuition underlying each method. It is hard to grasp the intuition until one has a working knowledge of the methods. We provide the explanations alongside a series of heuristic illustrations.

The illustrations form the backbone of this paper. The basic illustration has two dimensions with two values in each dimension. This prevents the intuition from getting submerged under tedious mathematics. In practice, the minimum bias procedure is most useful for multidimensional classification systems with many entries in each dimension.

We show the computations for one iteration in each illustration, followed by the series of values until convergence. The illustrations here converge in a few steps. In practice, more iterations are needed for convergence of larger models. The work is tedious by hand but elementary with current spreadsheet applications. Some spreadsheets have built-in iterative functions, such as "goal-seek" and "solver" in Excel. Some software packages, such as SAS, have built-in routines for GLMs. Once the intuition is clear, the programming is not difficult.

The Multiplicative Model

We are setting pure premiums; we do not deal with expenses or profit by classification or with gross premiums. We base the pure premiums upon the empirical observations in each cell of an array. For a two-dimensional classification system, this means each cell in a matrix. The observations can be average loss costs, loss frequencies, or loss ratios. In practice, the data would consist of losses and exposures (for loss costs), claim counts and exposures (for loss frequencies), or losses and premiums (for loss ratios).

Illustration 1: A classification system for private passenger automobile insurance has two dimensions: (i) urban versus rural and (ii) male versus female. A company insures exactly four drivers, one in each cell, with the following observed loss costs:[4]

	Urban	Rural
Male	\$600	\$200
Female	\$300	\$100

We determine pure premium relativities. We first compare all males with all females, or \$800 for two exposures compared to

[4]We deal with unequal cell populations later in the paper.

\$400 for two exposures. This gives a pure premium relativity of male to female = 2 to 1.

We do the same for urban versus rural, and we get a relativity of 3 to 1. We arbitrarily choose "rural female" as the base class; by convention, the lowest cost class or the class with the largest number of exposures is often chosen as the base class. We get the following relativities:

Male:	$2.00 = s_1$	Urban:	$3.00 = t_1$
Female:	$1.00 = s_2$	Rural:	$1.00 = t_2$.

The indicated pure premium for a male urban driver is the base pure premium times the urban relativity times the male relativity, or $\$100 \times 2.00 \times 3.00 = \$600$. More generally, the pure premium in cell (i, j) is $\$100 \times s_i \times t_j$.

In this illustration, the indicated pure premiums exactly match the observed loss costs. The minimum bias method is not needed for this case.

The Additive Model

The indicated pure premiums may differ from the observed loss costs because the model structure may be incorrect or because random loss fluctuation may affect the observed loss costs. We treat the first reason, the model structure, in this section.

Illustration 2: The observed loss costs for four drivers are shown below.

	Urban	Rural
Male	\$700	\$500
Female	\$400	\$200

We begin in the same fashion as before, using rural females as the base class. We compare all males to all females, giving a pure premium relativity of \$1,200 to \$600, or 2 to 1. We compare all urban to all rural, giving a pure premium relativity of \$1,100 to \$700, or 1.571 to 1.

The indicated pure premium relativities no longer match the observed loss costs. The indicated pure premium for rural males is $\$200 \times 2.000 = \$400$, but the observed loss cost is \$500. The indicated pure premium for urban females is $\$200 \times 1.571 = \$314$, but the observed loss cost is \$400. The differences are significant.

No multiplicative factors work perfectly. In urban territories, the relationship of male to female is \$700 to \$400, or 1.75 to 1. In rural territories, the relationship of male to female is \$500 to \$200, or 2.50 to 1. A male-to-female relativity appropriate for the urban territories is not optimal for the rural territories.

Similarly, the urban-to-rural relativity is \$700 to \$500, or 1.4 to 1, for male drivers, and \$400 to \$200, or 2 to 1, for female drivers.

The discussion in the paragraphs above assumes that the rating model is multiplicative; in this illustration, an additive model works better. We add or subtract a dollar amount for each class instead of multiplying by a factor. We choose rural females as the base class, and we use the relativities below:

Male:	+\$300	Urban:	+\$200
Female:	+\$0	Rural:	+\$0.

The pure premium for any cell is the base pure premium plus the male/female relativity plus the urban/rural relativity. The indicated pure premiums now match the observed loss costs. Rural male = \$200 + \$300 + \$0 = \$500; urban male = \$200 + \$300 + \$200 = \$700; urban female = \$200 + \$0 + \$200 = \$400. The additive method provides an exact match to the observed loss costs because the dollar differences are the same in each row (\$200) and in each column (\$300).

Additive and Multiplicative Intuition

Some actuaries implicitly assume that pure premium relativities should be multiplicative, not additive. If urban male drivers have twice the accident frequency that rural male drivers have, urban female drivers should have twice the accident frequency that rural female drivers have. This assumption is most persuasive when class dimensions are independent, that is, when the high accident frequency of urban drivers is not correlated with the high accident frequency of male drivers.[5] Most multidimensional class systems for the casualty lines of business use multiplicative factors.

Regulators sometimes harshly criticize insurers for using multiplicative factors that compound increases in the rates for high-risk insureds. This criticism is often—but not always—political. When two or more dimensions of the classification system are correlated, multiplicative systems are often biased. For some types of insurance, multiplicative systems may be biased even when classification dimensions are not correlated.[6]

Illustration 3: The 1960 Bailey and Simon paper discusses two rating dimensions: (i) class group and (ii) merit rating class.

1. Class group refers to the driver characteristics, such as age, sex, and marital status, and use of the vehicle, such as pleasure use or business use.
2. Merit rating class refers to the number of immediately preceding accident-free years, ranging from 0 to 3 or more.

---

[5]This assumption is rarely tested, and the independence of classification dimensions does not necessarily imply a multiplicative model. The authors' impressions from private passenger automobile loss costs is that neither the additive nor the multiplicative model is perfect, but the multiplicative model is usually better.

[6]Life insurance rating systems provide an example. If smokers have twice the mortality of non-smokers, and persons with high-blood pressure have twice the mortality of persons with average blood pressure, should high-blood pressure smokers have four times the mortality of average blood pressure non-smokers? Life insurance underwriters employ judgment to assess the rating for applicants with multiple causes of high mortality. A pure multiplicative rating system would not be appropriate.

These two rating dimensions are correlated. For example, young, unmarried male drivers have a high average class relativity. Because these drivers either are new drivers or (if not new) are more likely to have had an accident in the past year, they have relatively few accident-free years, and a multiplicative model would penalize many young male drivers twice for the same risk factor.

3. BIAS FUNCTIONS

In practice, the indicated pure premiums do not perfectly match the observed loss costs for either an additive model or a multiplicative model. We illustrate with the same 2-by-2 classification system. The observed loss costs are shown in the table below:

	Urban	Rural
Male	$800	$500
Female	$400	$200

Neither an additive model nor a multiplicative model provides a perfect match. If we use a model that does not perfectly match the observed data, we must determine how to minimize the mismatch between the observed loss costs and the indicated pure premiums. A "bias function" is a means of comparing two or more models to see which fits the data with the smallest degree of mismatch.[7] To choose the optimal model, we proceed along three steps:

1. We choose a rating method, such as an additive model or a multiplicative model.

[7] The bias function is not a standard statistical term, and the balance principle is not a standard principle. As used here, the bias function determines how "close" the indicated pure premiums are to the observed loss costs or how great the mismatch is between these two sets of data. The sum of the squared deviations and the χ-squared deviation are common statistical bias functions. The balance principle, introduced by Bailey and Simon in 1960 and endorsed again by Bailey in 1963, minimizes the bias along the dimensions of the classification system, thereby leading to the term "minimum bias."

2. We select a bias function and use it to optimize the rating method. This paper discusses the balance principle, least squares, χ-squared, and maximum likelihood bias functions. For models using a maximum likelihood bias function, we must also choose a probability density function for losses within each cell.

3. For each optimized rating method, we examine the goodness-of-fit of the indicated pure premiums to the observed loss costs.

We begin with the balance principle, since it is the bias function most commonly used.

The Balance Principle

The balance principle requires that (after optimizing the relativities) the sum of the indicated pure premiums equals the sum of the observed loss costs along every row and every column.

Illustration 4: We examine the balance principle for both the additive and the multiplicative models. There is one exposure in each cell. On the left are the observed loss costs; on the right are the indicated pure premiums. We begin with the multiplicative model.[8]

	Urban	Rural		terr_1	terr_2
Male	\$800	\$500	sex_1	$200 \times s_1 \times t_1$	$200 \times s_1 \times t_2$
Female	\$400	\$200	sex_2	$200 \times s_2 \times t_1$	$200 \times s_2 \times t_2$

To balance along the first row (the "male" row), we must have

$$800 + 500 = 200 \times s_1 \times t_1 + 200 \times s_1 \times t_2.$$

[8]To keep the notation simple, we use rating dimensions of male versus female and urban versus rural throughout this paper. For the formulas in the illustrations, we use $\text{sex}_1 = s_1 =$ male, $\text{sex}_2 = s_2 =$ female, $\text{terr}_1 = t_1 =$ urban, and $\text{terr}_2 = t_2 =$ rural. The recursive equations use variable names of x, y, and z, and rating dimensions of i and j.

To balance along the second row (the "female" row), we must have

$$400 + 200 = 200 \times s_2 \times t_1 + 200 \times s_2 \times t_2.$$

To balance along the first column (the "urban" column), we must have

$$800 + 400 = 200 \times s_1 \times t_1 + 200 \times s_2 \times t_1.$$

To balance along the second column (the "rural" column), we must have

$$500 + 200 = 200 \times s_1 \times t_2 + 200 \times s_2 \times t_2.$$

Although we have four equations in four unknowns, we do not have a unique solution for the classification relativities. There are two special considerations we must be aware of. These two considerations offset each other so as to yield a unique set of indicated pure premiums for each cell of the matrix (see below).

Dependence among the equations: These equations are related by a totality constraint—using any three equations we can derive the fourth, since the sum of the rows equals the sum of the columns. For instance, the fourth equation equals the first equation plus the second equation minus the third equation.

More generally, the equation for any column equals the sum of the equations for the rows minus the sum of the equations for the other columns, and likewise for the equation in any row.

Invariance under reciprocal scalar multiplication: We can set one of the variables arbitrarily, and we can still solve the system of equations. To see this most clearly, suppose that we have solved these equations for values of the four variables s_1, s_2, t_1, and t_2. Another solution is $2s_1$, $2s_2$, $\frac{1}{2}t_1$, and $\frac{1}{2}t_2$. We could use any constant in place of 2. But no matter which set of relativities we pick, the values in the cells remain the same. The values in the cells are the product of an s relativity and a t relativity, so the additional constant cancels out.

We have an additional variable. The pure premium in each cell depends on the base pure premium. If the relativities s_1, s_2, t_1, and t_2 optimize the rating model for a base pure premium of \$200, the relativities $2s_1$, $2s_2$, t_1, and t_2 optimize the rating model for a base pure premium of \$100.[9]

The minimum bias procedure makes the relationship of the rating variables along each dimension of the classification system constant. If $s_1 = 2 \times s_2$ for a given base pure premium and a given set of territorial relativities, then $s_1 = 2 \times s_2$ for any other base pure premium and for any constant multiple of the territorial relativities.

We choose a base class in each classification dimension. This is often the largest class or the lowest-cost class, though any class may be used. The base class in each classification dimension is given a relativity of 1. This determines the values of the base pure premium and of all other rating variables.

Solving the Equations

The equations are not linear, so there is no closed-form solution. We begin with an arbitrary (but reasonable) set of relativities for one dimension, and we solve the equations iteratively.

Illustration 5: We choose an urban relativity of 2.00 and rural relativity of 1.00; this choice does not affect the final pure premiums.

	Urban	Rural		$\text{terr}_1 = 2$	$\text{terr}_2 = 1$
Male	\$800	\$500	sex_1	$200 \times s_1 \times 2$	$200 \times s_1 \times 1$
Female	\$400	\$200	sex_2	$200 \times s_2 \times 2$	$200 \times s_2 \times 1$

[9]With so much leeway in choosing the classification relativities, one might ask what we are "optimizing." We are optimizing the indicated pure premiums. Each set of classification relativities give the same indicated pure premiums. The optimization is relative to the bias function. For example, the optimal pure premiums have the least bias or the least χ-squared value.

The balance equation for the first row (the “male” row) says that

$$800 + 500 = 200 \times s_1 \times 2 + 200 \times s_1 \times 1,$$
$$\text{or} \quad s_1 = 1{,}300/600 = 13/6.$$

Balancing along the second row (the “female” row) gives

$$400 + 200 = 200 \times s_2 \times 2 + 200 \times s_2 \times 1,$$
$$\text{or} \quad s_2 = 600/600 = 1.$$

We now have intermediate values for the male and female relativities of 13/6 and 1. We discard the initial values for the urban and rural relativities of 2.00 and 1.00, and we solve for new intermediate values by balancing along the columns. The balance equation for the first column (the “urban” column) says that

$$800 + 400 = 200 \times (13/6) \times t_1 + 200 \times 1 \times t_1,$$
$$\text{or} \quad t_1 = 1{,}200/633.33 = 1.895.$$

Balancing along the second column (the “rural” column) gives

$$500 + 200 = 200 \times (13/6) \times t_2 + 200 \times 1 \times t_2,$$
$$\text{or} \quad t_2 = 1.105.$$

We continue in this fashion. We discard the previous male and female relativities, and we solve for new ones. Balancing along the first row (the “male” row) gives

$$800 + 500 = 200 \times s_1 \times 1.895 + 200 \times s_1 \times 1.105,$$

and balancing along the second row (the “female” row) gives

$$400 + 200 = 200 \times s_2 \times 1.895 + 200 \times s_2 \times 1.105.$$

We solve these two equations for new values of the male and female relativities, we discard the previous values of the urban and rural relativities, and we balance along the columns for new values of the urban and rural relativities.

We continue in this fashion until the relativities converge, i.e., the change in the relativities from an additional iteration is not material. Calculating minimum bias relativities is tedious by hand but easy with a spreadsheet. In this case, convergence is rapid, since there are only four cells. Once the series converges, common practice is to normalize the base class relativities to unity and change the base pure premium (to \$221.05), as we do below:

Iteration	Urban	Rural	Male	Female
Initial	2.0000	1.0000		
1-a			2.1667	1.0000
1-b	1.8947	1.1053		
2-a			2.1667	1.0000
Final	1.8947	1.1053		
Normalized	1.7143	1.0000	2.1667	1.0000
Normalized Base Pure Premium:		$\$200 \times 1.1053 = \$221.05$		

The initial territorial relativities of 2.00 and 1.00 were arbitrary; we generally begin with starting values determined by a one-way relativities procedure. The starting values have no effect on the final rates in each cell, though better starting values reduce the iterations required to reach convergence. In this illustration, the urban to rural relativity is 12 to 7. If we choose a pure premium relativity of 1.000 as the starting value for the rural class, we would choose a starting value of $12 \div 7 = 1.714$ for the urban class. With a starting value of $t_1 = 1.714$, the series converges immediately. We used a different starting value to show the steps of the procedure.

The Additive Model

There are several equivalent formulas for the additive model. The pure premium in cell (i, j), or row i and column j, is

A. Base pure premium $+ x_i + y_j$,

B. Base pure premium $\times (1 + u_i + v_j)$, or

C. Base pure premium $\times (p_i + q_j)$.

To see the equivalence of these formulas, suppose the base pure premium in formula A is \$10.

- In formula B, the base pure premium is also \$10, each u value is one tenth the corresponding x value in formula A, and each v value in formula B is one tenth the corresponding y value in formula A: $u_i = 0.1 \times x_i$ and $v_j = 0.1 \times y_j$.
- Formula C is equivalent to formula B, except that either the p values are all increased by 1, the q values are all increased by 1, or the p values are increased by a constant (c) and the q values are increased by the complement of that constant $(1 - c)$: $p_i = 1 + u_i$ or $q_j = 1 + v_j$ (but not both) or $p_i = c + u_i$ and $q_j = 1 - c + v_j$.

We use the first form—formula A—for our example, since it shows the method most clearly.[10]

Illustration 6: We choose initial values for urban and rural relativities of \$250 and \$0. These initial values are based on the traditional pure premium relativities procedure; the average differential between the urban and rural observed loss costs is $\frac{1}{2} \times [(800 - 500) + (400 - 200)] = \$250$.

	Urban	Rural		$\text{terr}_1 = 250$	$\text{terr}_2 = 0$
Male	\$800	\$500	sex_1	$200 + s_1 + 250$	$200 + s_1 + 0$
Female	\$400	\$200	sex_2	$200 + s_2 + 250$	$200 + s_2 + 0$

Balancing along the first row (the "male" row) gives

$$800 + 500 = 200 + s_1 + 250 + 200 + s_1 + 0,$$

$$\text{or} \quad s_1 = 650/2 = 325.$$

[10]In practice, formulas B or C might be preferred, since only the base pure premium need be increased for inflation. In formula A, the base pure premium and all the relativities must be increased for inflation.

Balancing along the second row (the "female" row) gives

$$400 + 200 = 200 + s_2 + 250 + 200 + s_2 + 0,$$
$$\text{or} \quad s_2 = -50/2 = -25.$$

We discard the initial values for the urban and rural relativities, and we balance along the columns. We use the intermediate values of the male and female relativities to get new values for the urban and rural relativities. We continue this iterative process until the series converges.

The relativity of –$25 for females seems odd at first. In truth, the relativity for female drivers is not inherently negative; this is an artifact of the base pure premium and the starting values. We could make the relativity for females positive by adding a constant to the male and female relativities and subtracting the same constant from the rural and urban relativities. For instance, we could add $75 to the male and female relativities to get relativities of $400 and $50, and we would subtract $75 from the rural and urban relativities.

Iteration	Urban	Rural	Male	Female
Initial	$250.00	$0.00		
1-a			$325.00	($25.00)
1-b	$250.00	$0.00		
Normalized	$250.00	$0.00	$350.00	$0.00
Normalized Base Pure Premium:	$200.00 – $25.00 = $175.00			

We can even make all the relativities negative by adjusting the base pure premium. For instance, by choosing a base pure premium of $1,000, we obtain negative relativities for all classes.[11] In this illustration, we added dollar amounts to make the base class relativities equal to zero.

[11]Companies may do this for marketing reasons. All drivers get discounts from the base pure premium, so all drivers feel they are gaining from the classification system.

Exposures

The illustrations above assume one driver in each cell or the same number of drivers in each cell. In practice, there are generally different numbers of risks in each cell. Two adjustments are needed, one to the bias function and another for credibility:

- We adjust the bias function for the relative volume of business in each cell.
- We may make a credibility adjustment based on the absolute volume of business in a cell.

Illustration 7—Credibility: We note the credibility issue, but we defer the possible adjustments until later. Suppose insurer A has 100 exposures per cell, and insurer B has 10,000 exposures per cell. Insurer A may rely more heavily on the minimum bias procedure. Insurer B may give greater weight to the empirical observations.

We deal here with the adjustment to the bias function. The balance principle requires that the sum of the observed loss costs in each row or column equal the sum of the indicated pure premiums in the corresponding row or column. If there are two drivers in a cell, we double both the observed loss cost and the indicated pure premium in that cell. If there are n drivers in a cell, we multiply both the observed loss cost and the indicated pure premium by n. When the number of drivers varies by cell, we need an additional matrix of the number of drivers in each cell.

Illustration 8: For the multiplicative model, suppose that the number of drivers is as follows:

Male urban:	1,200	Male rural:	600
Female urban:	1,000	Female rural:	800.

We include the number of drivers in the equations.

	Urban	Rural		terr_1	terr_2
Male	$1200 \times \$800$	$600 \times \$500$	sex_1	$1200 \times 200 \times s_1 \times t_1$	$600 \times 200 \times s_1 \times t_2$
Female	$1000 \times \$400$	$800 \times \$200$	sex_2	$1000 \times 200 \times s_2 \times t_1$	$800 \times 200 \times s_2 \times t_2$

To balance along the first row (the "male" row), we must have

$$1200 \times 800 + 600 \times 500$$
$$= 1200 \times 200 \times s_1 \times t_1 + 600 \times 200 \times s_1 \times t_2.$$

To balance along the second row (the "female" row), we must have

$$1000 \times 400 + 800 \times 200$$
$$= 1000 \times 200 \times s_2 \times t_1 + 800 \times 200 \times s_2 \times t_2.$$

To balance along the first column (the "urban" column), we must have

$$1200 \times 800 + 1000 \times 400$$
$$= 1200 \times 200 \times s_1 \times t_1 + 1000 \times 200 \times s_2 \times t_1.$$

To balance along the second column (the "rural" column), we must have

$$600 \times 500 + 800 \times 200$$
$$= 600 \times 200 \times s_1 \times t_2 + 800 \times 200 \times s_2 \times t_2.$$

Empirical Observations versus Modeled Pure Premiums

One might wonder: Why not use the observed loss costs, appropriately developed and trended, as the indicated pure premiums for the coming policy period? Instead of fitting either multiplicative or additive models to the observed data, let us use \$800 as the indicated pure premium for urban male drivers, \$400 for urban female drivers, \$500 for rural male drivers, and \$200 for rural female drivers.

The common answer is that the individual cells are "not fully credible." This answer is correct, though the terminology is not ideal. The term "credible" is vague. To understand the situation, we must be more precise.

Credibility is a relative concept. No cell is inherently credible or not credible. A cell's credibility depends on the reliability of its own experience in comparison with information in other cells. Consider our basic illustration with the following observed loss costs:

	Urban	Rural
Male	$800	$500
Female	$400	$200

The urban male observed pure premium of $800 represents a mixture of expected losses and random loss fluctuations. How might we judge whether this figure is higher or lower than the true expected loss costs?

Suppose that the rating values combine additively to generate the expected losses. The observed loss cost for urban males of $800 is $300 more than the observed loss cost of rural males of $500. This suggests that the urban attribute of the vehicle's location adds about $300 to the expected loss costs.

However, the urban female observed loss cost of $400 is only $200 more than the rural female observed loss cost of $200. This suggests that the extra cost associated with the urban attribute is only $200, not $300, and it implies that the observed urban male loss cost of $800 might be too high.

We perform a similar analysis for male versus female. Comparing urban drivers suggests that the male attribute adds about $400 to the expected loss costs, since male/urban = $800 and female/urban = $400. However, comparing rural drivers suggests that the extra cost associated with the male attribute is only $300, not $400, since male/rural = $500 and female/rural = $200. In other words, the urban male loss cost of $800 might be too high.

The \$800 observed loss cost in the urban male cell does not tell us what part of this observed loss cost is expected and what part is distorted by random loss fluctuations. If we know the mathematical function linking the cells—that is, if the pure premiums of the driver and the vehicle have some additive or multiplicative relationship—we can use additional cells to provide information about the true expected costs for urban male drivers, as we have here.

If we assume that the cells are linked multiplicatively, our inferences change. The urban male observed value of \$800 is 160% of the rural male observed value of \$500. This suggests that the urban attribute adds about 60% to the expected loss costs.

The urban female observed loss cost of \$400 is twice the rural female observed loss cost of \$200. This suggests that the extra cost associated with the urban attribute is +100%, not +60%. The urban male loss cost of \$800 might be too low.

Using a similar analysis for male versus female using the urban column suggests that the male attribute adds about 100% to the expected loss costs. The rural column suggests that the extra cost associated with the male attribute is +150%, not +100%. The urban male loss cost of \$800 might be too low.

If the cells are linked additively, we infer that the urban male observed loss costs of \$800 might be too high. If the cells are linked multiplicatively, we infer that the urban male observed loss costs of \$800 might be too low.[12]

If the exposures in a 2 by 2 matrix are evenly distributed among the cells, each cell has 25% of the total exposures, whether there is 1 car or 10,000 cars in each cell. We give much

[12]In most cases, the direction of the bias does not depend on the type of rating model. The more common scenario might show an observed loss cost of \$600, an additive model indicated pure premium of \$550, and a multiplicative model indicated pure premium of \$530. We might infer that the random loss fluctuations underlying these cell values have had a net positive effect. For very high rated or very low rated classifications, the multiplicative and additive models often give opposite results, as is the case here.

credence to the observed value in that cell compared to our inferences from other cells. With a larger array, such as a 10 by 10 by 10 array, there are many more cells. The average cell contains only 0.1% of the total exposures. We give less credence to the observed loss costs in that cell compared to our inferences from other cells.

This is the intuition for the minimum bias procedure. The rating model—such as additive, multiplicative, or combined—tells us the relationship joining the cells. The bias function—such as balance principle, χ-squared, least squared error, or maximum likelihood—provides a method of drawing inferences for one cell using the information in the other cells.

Credibility—Original Papers

The original papers on the minimum bias procedure differ regarding credibility.

The 1960 Bailey and Simon paper uses credibility considerations to pick a bias function. The authors' view that credibility should be inversely proportional to the variance of the observations led them to choose the χ-squared bias function over the balance principle.

The 1963 Bailey paper, which advocates the balance principle, has no explicit discussion of credibility. The balance principle has an implicit credibility component, since it weights the observed loss costs and pure premiums in each cell by the number of exposures in the cell.

This implicit credibility examines the relative weights of different cells in the minimum bias procedure. Venter looks at credibility from a different angle—the relative weight given to the indicated pure premium from the minimum bias procedure versus the observed loss costs from the experience. We said above that the $800 observed loss cost for urban male drivers might be overstated under an additive model or understated under a multiplicative model. The over- or understatement stems from random loss fluctuations. If there is a single exposure in each cell, an over- or understatement is likely. If there are 10,000 exposures in each cell, the degree of over- or understatement is likely to be smaller.

Iterative Formulas

We have so far presented simple illustrations. To program more complicated versions of this procedure, we need general formulas.

We derive the iterative formulas for the multiplicative balance principle model. We designate the base pure premium by b, the number of exposures in row i and column j by n_{ij}, and the observed pure premiums in row i and column j by r_{ij}. With a multiplicative model, the balancing equation for row i is

$$\sum_j (n_{ij}r_{ij}) = \sum_j (bn_{ij}x_i y_j).$$

Similarly, the balancing equation for column j is

$$\sum_i (n_{ij}r_{ij}) = \sum_i (bn_{ij}x_i y_j).$$

In these equations, x is a row relativity and y is a column relativity.[13] We solve these equations to get the indicated x and y relativities in each row and column:[14]

$$x_i = \frac{\sum_j n_{ij}r_{ij}}{\sum_j bn_{ij}y_j} \qquad \text{and} \qquad y_i = \frac{\sum_i n_{ij}r_{ij}}{\sum_i bn_{ij}x_i}.$$

When the series converges, we set the relativity for the base class in each classification dimension to unity, and we adjust the base pure premium to offset this.

[13]In the illustrations, we use s for the row relativity and t for the column relativity as abbreviations for the classification dimensions (sex and territory). The variables x and y are commonly used in the literature.

[14]We sum over the j subscript when we balance along the rows (the i subscripts). We do this separately for each i. When we balance along the columns, we sum over the i subscripts separately for each j.

We used two dimensions in this formula. One might assume that the two dimensions correspond to the two variables x and y. That is not correct. The two dimensions correspond to the two subscripts i and j. The x and y variables correspond to two sets of relativities. A model can have two or even more sets of relativities in a single dimension.

Illustration 9: The classification system has two dimensions: male versus female and territory A versus territory B. Territory A has more attorneys than territory B has, resulting in a higher propensity to sue and higher loss costs per claim. Territory B has several blind intersections, leading to more accidents. We might presume that the higher attorney involvement in territory A increases the cost of all claims, so a multiplicative factor is appropriate, whereas the blind intersections in territory B add additional hazards, so an additional factor is appropriate. The rating model might take the form

$$\text{indicated pure premium relativity} = x_i \times y_j + z_j,$$

where i represents the male/female classification dimension and j represents the territory dimension. The variable x is the relativity for the male/female dimension, and the variables y and z are the relativities for the urban/rural dimension. In this model, x and y are multiplicative factors, and z is an additive factor.[15]

The arithmetic is similar for any number of dimensions. The multiplicative model has one set of relativities for each dimension. With three dimensions, for example, the iterative formula for the i dimension is

$$x_i = \frac{\sum_{j,k} n_{ijk} r_{ijk}}{\sum_{j,k} b n_{ijk} y_j z_k}.$$

[15]To optimize this rating model, the balance principle is not sufficient; we would have to employ one of the other bias functions. The balance principle provides $i + j$ equations, but we have $i + 2j$ variables. The other bias functions discussed in this paper provide $i + 2j$ equations.

We develop the general formula for the balance principle additive model by assuming a base pure premium of \$0. The balance principle equation is

$$\sum_j (n_{ij} r_{ij}) = \sum_j n_{ij}(x_i + y_j),$$

and the iterative formula is

$$x_i = \frac{\sum_j n_{ij}(r_{ij} - y_j)}{\sum_j n_{ij}}.$$

Exercise: Multiplicative Model

Illustration 10: We use the balance principle to optimize a multiplicative rating model with two dimensions and two classes in each dimension. The observed loss costs and exposures in each class are shown below:

	Loss Costs		Exposures	
	y_1	y_2	y_1	y_2
x_1	300	300	100	150
x_2	200	400	100	100

We assume a base pure premium of \$100, so the indicated pure premiums are \$100$x_i y_j$. To simplify the mathematics, we compute all values in units of \$100. The indicated pure premiums are $x_i \times y_j$, and the observed loss costs are \$3, \$3, \$2, and \$4.

We form a matrix of observed loss costs and indicated pure premiums:

	y_1	y_2		y_1	y_2
x_1	3	3	x_1	$x_1 \times y_1$	$x_1 \times y_2$
x_2	2	4	x_2	$x_2 \times y_1$	$x_2 \times y_2$

We multiply each figure by the exposures in the cell:

	y_1	y_2		y_1	y_2
x_1	100×3	150×3	x_1	$100 \times x_1 \times y_1$	$150 \times x_1 \times y_2$
x_2	100×2	100×4	x_2	$100 \times x_2 \times y_1$	$100 \times x_2 \times y_2$

We choose 1.00 and 1.50 as the starting values for y_1 and y_2. We use the balance principle to obtain intermediate values for x_1 and x_2:

$$100 \times 3 + 150 \times 3 = 100 \times x_1 \times 1.00 + 150 \times x_1 \times 1.50,$$
$$\text{or} \quad 300 + 450 = 100 \times x_1 + 225 \times x_1,$$
$$\text{or} \quad x_1 = 2.308;$$

and

$$100 \times 2 + 100 \times 4 = 100 \times x_2 \times 1.00 + 100 \times x_2 \times 1.50,$$
$$\text{or} \quad 200 + 400 = 100 \times x_2 + 150 \times x_2,$$
$$\text{or} \quad x_2 = 2.400.$$

We now discard the initial values for y_1 and y_2, and we balance along the columns:

$$100 \times 3 + 100 \times 2 = 100 \times 2.308 \times y_1 + 100 \times 2.400 \times y_1,$$
$$\text{or} \quad 300 + 200 = 230.8 \times y_1 + 240 \times y_1,$$
$$\text{or} \quad y_1 = 1.062;$$

and

$$150 \times 3 + 100 \times 4 = 150 \times 2.308 \times y_2 + 100 \times 2.400 \times y_2,$$
$$\text{or} \quad 450 + 400 = 346.2 \times y_2 + 240 \times y_2,$$
$$\text{or} \quad y_2 = 1.450.$$

This completes one iteration. To solve for the optimal relativities, we continue in this fashion until convergence. We comment on several items in this exercise.

Data and Assumptions

The number of exposures in each cell is a credibility measure. We give 50% more credence to the observed loss costs in the $x_1 y_2$ cell than to the loss costs in the other cells.

- The observed loss costs in the x_1 row indicate that there is no difference between y_1 and y_2. The observed loss costs in the x_2 row indicate that the y_2 class should have twice the pure premium that y_1 has. We give more credence to the first of these two relationships.
- The observed loss costs in the y_1 column indicate that the x_2 class should have a pure premium 33% lower than the x_1 class. The observed loss costs in the y_2 column indicate that the x_2 class should have a pure premium 33% higher than the x_1 class. We give more credence to the second of these two relationships, so the x_2 relativity is slightly higher than the x_1 relativity.

Exercise: Additive Model

Illustration 11: An additive model with two dimensions has the observed loss costs shown below. Each cell has 1,000 exposures. The base loss cost is \$100. The formula for loss costs by cell is Loss Cost$_{ij}$ = (Base Loss Cost) $\times$ $(x_i + y_i)$. We use the starting values shown below to compute intermediate values for y_1 and y_2.

	Average Loss Costs per Exposure y_1	y_2	Starting Values
x_1	500	750	4.500
x_2	250	475	3.000
x_3	150	400	2.000

Since the number of exposures is the same in each cell, we may assume that there is a single exposure in each cell; the 1,000 cancels out of every equation.

The base pure premium is \$100. To simplify, we use units of \$100 and a base pure premium of unity. The matrix of observed loss costs and indicated pure premiums is shown below:

	Observed Values		Indicated Values	
	y_1	y_2	y_1	y_2
x_1	5	7.5	$x_1 + y_1$	$x_1 + y_2$
x_2	2.5	4.75	$x_2 + y_1$	$x_2 + y_2$
x_3	1.5	4	$x_3 + y_1$	$x_3 + y_2$

We balance along the columns. For the first column, we have

$$5.00 + 2.50 + 1.50 = (x_1 + y_1) + (x_2 + y_1) + (x_3 + y_1),$$

$$5.00 + 2.50 + 1.50 = (4.50 + y_1) + (3.00 + y_1) + (2.00 + y_1),$$

$$\text{or} \quad 3y_1 = 9.00 - 9.50, \quad \text{or} \quad y_1 = -0.167.$$

For the second column, we have

$$7.50 + 4.75 + 4.00 = (x_1 + y_2) + (x_2 + y_2) + (x_3 + y_2),$$

$$7.50 + 4.75 + 4.00 = (4.50 + y_2) + (3.00 + y_2) + (2.00 + y_2),$$

$$\text{or} \quad 3y_2 = 16.25 - 9.50, \quad \text{or} \quad y_2 = 2.25.$$

We have finished balancing along the columns. The next step is to balance along the rows. We take the new y values, $y_1 = -0.167$ and $y_2 = +2.25$, and we compute new values for x_1 and x_2 by balancing along each row. We continue this process—alternately balancing along rows and columns—until we reach convergence.

During the iterative process, the plan is alternately balanced along the rows or along the columns, but not along both. We have just balanced along the columns. To see that we are not yet balanced along the rows, we examine the first row:

$$5.00 + 7.50 = (x_1 + y_1) + (x_1 + y_2).$$

Substituting the starting values of the xs and the first iterative values of the ys, we get

$$12.50 = 4.50 + (-0.167) + 4.50 + 2.25 = 11.083.$$

The equality does not hold, since the plan is not yet balanced. The final values, after convergence, are shown below.

Iteration	x_1	x_2	x_3	y_1	y_2
Initial	4.50000	3.00000	2.00000		
1-a				−0.16667	2.25000
1-b	5.20833	2.58333	1.70833		
2-a				−0.16667	2.25000
Final	5.20833	2.58333	1.70833	−0.16667	2.25000

4. OTHER CLASSIFICATION DIMENSIONS

The basic illustrations use the minimum bias procedure to set pure premium relativities simultaneously for the male/female dimension and the urban/rural dimension. There may be other dimensions to the classification plan as well, such as age of driver, marital status, type of vehicle, use of the car, driver education, prior accident history, and so forth.

Suppose that we analyze the male/female dimension and the urban/rural dimension on a statewide basis, and we set relativities for other classification dimensions on a countrywide basis. We use a minimum bias method for the statewide analysis.

If all the classification dimensions are independent, the analysis should work well. If one or more of the other classification dimensions is correlated with the male/female or urban/rural dimensions, the rating analysis may be distorted.

Illustration 12: Suppose that young people migrate to urban areas for university education, work opportunities, and the glamor of urban social activities. Older people move to the suburbs and rural areas to buy homes and raise families away from the vices of urban areas. The age and marital status of the driver are correlated with the urban/rural garaging location.

The statewide analysis may indicate an urban to rural relativity of 2 to 1. The countrywide analysis, summing over all territories, may indicate a relativity for young unmarried male drivers of 3 to 1 when compared to adult drivers. The relativity for young unmarried urban male drivers is not 6 to 1, even if a multiplicative model is appropriate for automobile insurance.

Multiple Dimensions

Ideally, we would use a multidimensional minimum bias procedure to set all classification relativities simultaneously. In practice, this may not be possible. Some relativities may be analyzed each year, whereas other relativities may be analyzed every several years. Some relativities, such as territory, must be set on a statewide basis. Certain driver characteristics and vehicle characteristics may be analyzed on a countrywide basis, for two reasons:[16]

1. The relativities are not expected to vary by state, as long as the states use the same insurance compensation system.
2. Some classification cells would have few exposures in a state analysis, and the results may be distorted by random loss fluctuations. The countrywide analysis uses more data, providing more credible results. For example, we may wish to analyze driver age in yearly increments: age 17, age 18, age 19, and so forth. Single-state data may be too sparse to give credible results.

[16]The countrywide analysis may actually be done on all tort liability states or all no-fault states, since the bodily injury rate relativities may be higher for SUVs (sports utility vehicles) than for sedans in tort liability states, whereas the reverse may be true in no-fault states.

Some classification dimensions, such as driver education, have a minor effect on overall loss costs. We may analyze these classification dimensions every five years or so, not every year.

Loss Ratios

One method of dealing with an uneven distribution of business along other classification dimensions is to use loss ratios instead of loss costs in the minimum bias procedure.[17,18]

Suppose the empirical experience consists of the following loss ratios by classification.

	Urban	Rural
Male	75%	85%
Female	90%	80%

We could take either of two approaches:

First Approach: We treat the unadjusted loss ratios as though they were loss costs. Instead of using pure premium relativities, we develop loss ratio relativities. These relativities are adjustments to whatever pure premium relativities are embedded in these loss ratios.

In this scenario, the minimum bias procedure will indicate a loss ratio relativity close to 1.000 for urban versus rural and a relativity slightly higher than 1.000 for females versus males. This does not mean that urban risks are similar to rural risks, or that female drivers have more accidents than male drivers have. If the current rate relativities are reasonable, we would expect the loss ratios in all cells to be about equal. Suppose that

[17] In practice, we use loss ratios adjusted to the base rates for the classification dimensions included in the minimum bias analysis, though this is not shown in the illustration.

[18] This section assumes that the pure premium relativities are the same as the rate relativities.

the current male-to-female rate relativity is 2.4 to 1. Since the average female loss ratio of 85% is higher than the average male loss ratio of 80%, the loss ratio relativities would indicate that we should slightly reduce the male-to-female rate relativity.

Second Approach: We convert the raw loss ratios to base class loss ratios. Suppose the current rate relativities are 2.4 to 1 for male to female and 1.8 to 1 for urban to rural. We must divide the male premiums by 2.4 and the urban premiums by 1.8. This is equivalent to multiplying the male loss ratios by 2.4 and the urban loss ratios by 1.8. We multiply the raw loss ratios by the current classification relativities, as shown in the table below.

	Urban	Rural
Male	$75\% \times 2.4 \times 1.8 = 324\%$	$85\% \times 2.4 \times 1.0 = 204\%$
Female	$90\% \times 1.0 \times 1.8 = 162\%$	$80\% \times 1.0 \times 1.0 = 80\%$

We apply the minimum bias procedure to the adjusted loss ratios. The resulting loss ratio relativities would be the same as the indicated rate relativities.

To see this, suppose that the base rate is \$100. For the male urban cell, the premium is $\$100 \times 2.4 \times 1.8 = \$432$. The observed loss ratio is 75%, so the loss cost is $75\% \times \$432 = \$324$. We may verify this for the other three cells in the same fashion.

To set the rate relativity to unity for the base class in each dimension, we divide each adjusted loss ratio in the matrix by the adjusted loss ratio for the base class.

	Urban	Rural
Male	$324\%/80\% = 405.0\%$	$204\%/80\% = 255.0\%$
Female	$162\%/80\% = 202.5\%$	$80.0\%/80\% = 100.0\%$

Loss Ratio Intuition

We have shown how to convert loss ratios to reflect the loss costs in each cell. This might be useful if the observed data were loss ratios and we wanted to use loss costs for the minimum bias procedure. But if observed data are loss costs, we must convert the observed loss costs to loss ratios before converting back to loss costs. The purpose of this conversion from loss costs to loss ratios and then back to loss costs is to eliminate the potentially distorting effects of other classification dimensions that are not being analyzed in the minimum bias procedure.

Illustration 13: We explain by illustration. We have average observed bodily injury loss costs for four groups of drivers, with 1,000 drivers in each cell.

	Urban	Rural
Male	$800	$500
Female	$400	$200

Other dimensions in the classification system are correlated with the two dimensions above.

Type of Vehicle: For bodily injury rating, cars are subdivided between (a) large cars, such as sports utility vehicles (SUVs), station wagons, and light trucks, and (b) small cars, such as sedans. The large vehicles provide better protection for their occupants, but they cause greater damage to others. Smaller vehicles cause less damage to others, but they provide less protection for their occupants. Sedans and other small cars are more common in urban areas; SUVs and light trucks are more common in rural areas. The distribution of vehicle types between urban and rural areas, along with the appropriate surcharge or discount for each type of vehicle, affect the observed loss costs.

Suppose that SUVs and other large vehicles receive a 20% surcharge for bodily injury. In this state, SUVs comprise 40%

of the rural vehicles and 10% of the urban vehicles. The pricing actuary may not actually have this distribution for the state under review. This is not necessary; the use of loss ratios instead of loss costs corrects for the effects of vehicle type.

Age of Driver: The male/female rate relativity applies to all male and female drivers. Unmarried male drivers under the age of 21 receive additional surcharges, ranging from 25% for 20-year-old drivers to 125% for 16-year-old drivers. In this state, 10% of male drivers are unmarried and under the age of 21. The average surcharge for these drivers is 50%. (For this illustration, there is no corresponding surcharge for unmarried female drivers under the age of 21.) The pricing actuary may not actually have a distribution of male drivers by age and marital status. Again, this is not necessary; the loss ratios are sufficient.

Double Counting and Offsetting

If we do not take vehicle type and driver age into account, we overcharge male drivers and rural drivers.

Male Drivers: The male/female relativity is based on the statewide analysis. The surcharges for young unmarried male drivers are determined from a separate countrywide analysis. The poor driving experience of young unmarried male drivers is counted twice: once at the countrywide level for the surcharges and once at the state level for the male/female relativity. To determine accurately the male/female relativity, we must remove the hazardous effects of being young and unmarried from the male driver classification.

Rural Drivers: Rural drivers are less hazardous than urban drivers, but they drive vehicles more dangerous to others. The vehicle surcharge is determined in a countrywide analysis. To properly determine the urban/rural relativity, we must remove the effects of vehicle type from the statewide experience.

To remove the effects of vehicle type and driver age from the statewide analysis, we assume that the countrywide relativities are accurate. We examine each risk in the minimum bias procedure. We divide the actual loss costs by the vehicle type relativity and by the driver age relativity. This gives the relative loss costs that we would have expected to see were the vehicle types and driver ages evenly distributed over all other rating dimensions.

Illustration 13, continued: A four-door sedan is the base vehicle type and age 21+ is the base age. A two-door compact has a bodily injury discount of 10%, and an unmarried 20-year-old male driver has a surcharge of 25%. Suppose the observed loss costs for a 20-year-old unmarried male driver of a two-door compact car are \$450. The loss costs adjusted for driver age and vehicle type are $\$450/(0.90 \times 1.25) = \$400$.

It is not practical to make these adjustments car by car. Using loss ratios adjusts for all classification dimensions simultaneously. Using observed loss ratios instead of observed loss costs adjusts for driver age, driver sex, territory, vehicle types, and all other rating dimensions. We then restore the current rating relativities for the classification dimensions that we are analyzing—male/female and urban/rural in this illustration.

The average observed loss costs for the 1,000 drivers in each of four classes are displayed in the table after illustration 13. The current relativities are 2.4 to 1 for male to female and 1.8 to 1 for urban to rural. The average SUV-to-sedan relativity is 1.2 to 1. SUVs account for 40% of rural cars and 10% of urban cars. Unmarried males under the age of 21 make up 10% of male drivers, and their average surcharge is 50%. Ideally, we would convert the observed loss costs to adjusted loss costs for the minimum bias analysis in the following manner.

- *Rural female:* SUVs are 40% of rural cars, increasing the loss costs by a factor of $1 + (20\% \times 40\%) = 1.08$. Were the cars all sedans, the observed loss costs would be reduced by a factor of $1/1.08 = 92.59\%$.

- *Urban female:* The vehicle type factor is $1 + (20\% \times 10\%) = 1.02$. Were the cars all sedans, the observed loss costs would be reduced by a factor of $1/1.02 = 98.04\%$.

- *Rural male:* The vehicle type factor is $1 + (20\% \times 40\%) = 1.08$ and the driver age factor is $1 + (10\% \times 50\%) = 1.05$. Were the cars all sedans driven by adult drivers, the observed loss costs would be reduced by a factor of $1/(1.08 \times 1.05) = 88.18\%$.

- *Urban male:* The vehicle type factor is $1 + (20\% \times 10\%) = 1.02$ and the driver age factor is $1 + (10\% \times 50\%) = 1.05$. Were the cars all sedans driven by adult drivers, the observed loss costs would be reduced by a factor of $1/(1.02 \times 1.05) = 93.37\%$.

We have made all the adjustments by our knowledge of the distribution of other classification dimensions in the four cells of the matrix. This information is generally not available, and the procedure is complex when there are several classification dimensions. A simple alternative is to divide the losses by the premium charged in each cell, and then multiply by the base rate times the current relativities for the two classification dimensions which we are examining.

For each vehicle, we divide the losses by the premium, which is the base rate times the classification relativities for all classification dimensions. We multiply the result by the base rate times the classification relativities for male/female and urban/rural. This is equivalent to dividing by the classification relativities for the remaining dimensions.

Exercise: Loss Ratio Method

The incurred losses and earned premium in each cell are shown in the following table.

	Incurred Losses		Earned Premium	
	Urban	Rural	Urban	Rural
Male	\$2,700	\$2,000	\$3,000	\$4,000
Female	\$1,500	\$1,200	\$2,400	\$1,600

The current relativities by sex of driver and by garaging location are

Male:	1.50	Urban:	1.20
Female:	1.00	Rural:	1.00.

Causes of Unequal Loss Ratios

To correct for potential distortions caused by an uneven distribution of insureds by other classification dimensions, we use loss ratios instead of loss costs. If the rate relativities match the loss cost differences, the loss ratios should be equal in all cells, except for random loss fluctuations. In this example, the loss ratios are not all equal.

	Loss Ratios	
	Urban	Rural
Male	90.00%	50.00%
Female	62.50%	75.00%

There are several possible causes for the unequal loss ratios.

Cause 1—Random loss fluctuations: Random loss fluctuations are a credibility issue. This paper assumes either that the data are fully credible or that the pricing actuary has already made (or will make) whatever adjustments are warranted by credibility considerations. Credibility adjustments for sparse data are an important actuarial issue, though they are beyond the scope of this paper.

Cause 2—Improper rate relativities in other classification dimensions combined with an uneven distribution of insureds by these *other classification dimensions:* For example, perhaps the rates for a certain type of vehicle are too low, and the proportion of urban males driving that type of vehicle is greater than the proportions of the insureds in the other cells driving that type of vehicle.

If this is the cause of the differences, there is no perfect solution.[19] But if the distribution of insureds by the other classification dimension is not too uneven, an inaccuracy in the rates will not distort our analysis too much. We may restate our assumption as follows:

> For other classification dimensions, either the current rate relativities are accurate or the mix of insureds is relatively even across these other dimensions.

In many instances, this assumption is not perfect. Nevertheless, even if the use of loss ratios does not perfectly correct for distortions caused by an uneven distribution of insureds along other classification dimensions, it provides a partial correction.

Cause 3—Inaccuracies in the rate relativities for the two classification dimensions that we are examining (sex and territory): This is corrected by the minimum bias procedure, since the loss ratios by cell times the current relativities by cell equal the relative loss costs by cell.

Illustration 14: Suppose the loss ratio for male drivers is 90% and the loss ratio for female drivers is 62.5%. If the current male-to-female rate relativity is 1.5 to 1, the male-to-female loss cost relativity is $1.5 \times 90\%$ to $1 \times 62.5\% = 2.16$ to 1.

For the illustration in this section, we form a matrix of relativities by sex and territory:

[19]Without information about the other classification dimensions, we cannot optimize the class system.

	Current Rate Relativities	
	Urban	Rural
Male	1.80	1.50
Female	1.20	1.00

The relative loss costs by sex and territory are the product of the relativities and the loss ratios:

	Loss Cost Relativities	
	Urban	Rural
Male	1.62	0.75
Female	0.75	0.75

We can now determine optimal rate relativities by any of the minimum bias models discussed in this paper.

Cause 4—Improper model specification: We may be using a multiplicative model when an additive model would be more proper (or vice versa). Sometimes neither a multiplicative nor an additive model is ideal. We discuss the choice of model further below.

5. THE SQUARED ERROR BIAS FUNCTION

In this section, we examine other bias functions, beginning with the squared error function and the χ-squared function. We continue with our simple 2 by 2 illustration for both additive and multiplicative models using these bias functions. We review arguments for and against specific bias functions. We examine two goodness-of-fit tests—average absolute error and χ-squared—and we consider the relationship between the bias function chosen and the goodness-of-fit test.

We review also the maximum likelihood bias function and the distributions commonly used with it. We discuss the potential advantages and drawbacks of the more sophisticated bias functions compared to the balance principle.

Illustration 15: We return to the simple illustration with which we began, as reproduced below.

	Urban	Rural		terr_1	terr_2
Male	\$800	\$500	sex_1	$200 \times s_1 \times t_1$	$200 \times s_1 \times t_2$
Female	\$400	\$200	sex_2	$200 \times s_2 \times t_1$	$200 \times s_2 \times t_2$

The left-hand side of the matrix shows the observed loss costs; the right-hand side shows the indicated pure premiums. Our objective is to pick classification relativities such that the indicated pure premiums are "as close as possible" to the observed loss costs.

Statisticians would fit the classification relativities using one of the methods below:

1. Minimize the average absolute error between the indicated and observed figures.
2. Minimize the sum of the squared errors between the indicated and observed figures (i.e., the least squares bias function).
3. Minimize the sum of the relative squared errors between the indicated and observed figures (i.e., minimize the χ-squared error).
4. Maximize the likelihood of obtaining the observations given the classification relativities.

Although minimizing the average absolute error makes sense to practitioners, it is rarely used in statistics, because it is less mathematically tractable.[20] We use the average absolute error

[20]See, however, Cook [1967], p. 200: "Why then do we use the method of least squares? Simply because absolute values are alleged to be mathematically inconvenient." Cook provides an algorithm for minimizing the average absolute error, which is simple to compute and even easier to program.

as one of the goodness-of-fit tests. Given a set of classification relativities, it is easy to calculate the average absolute error. (It is less easy to determine the set of classification relativities that minimize the average absolute error.)

The three other methods result in relatively simple iterative equations for the minimum bias procedure. We first show the procedures, and then we discuss the intuition for each.

The squared error for each cell is the square of the difference between the observed loss costs and the indicated pure premium. For urban male drivers, this is $(\$800 - \$200 \times s_1 \times t_1)^2$.

We sum the squared errors for the four cells to get the sum of squared errors (SSE):

$$\begin{aligned}\text{SSE} = {} & (\$800 - \$200 \times s_1 \times t_1)^2 && \text{urban male}\\ & + (\$500 - \$200 \times s_1 \times t_2)^2 && \text{rural male}\\ & + (\$400 - \$200 \times s_2 \times t_1)^2 && \text{urban female}\\ & + (\$200 - \$200 \times s_2 \times t_2)^2 && \text{rural female.}\end{aligned}$$

To minimize the sum of the squared errors, we set the partial derivatives with respect to each variable equal to zero. For the "male" classification relativity (s_1), we have

$$\begin{aligned}0 = \partial\text{SSE}/\partial s_1 = {} & 2 \times (\$800 - \$200 \times s_1 \times t_1) \times (-\$200 \times t_1)\\ & + 2 \times (\$500 - \$200 \times s_1 \times t_2) \times (-\$200 \times t_2).\end{aligned}$$

We need to consider the cells only in the male (s_1) row. The other cells do not have an s_1 term in the squared error, so the partial derivative with respect to s_1 is zero.

Taking partial derivatives with respect to each of the classification relativities gives four equations in four unknowns. The equations are not linear, so we solve them iteratively.

Let us choose the same starting values for the squared error bias function as we chose for the balance principle (namely $t_1 = 2$ and $t_2 = 1$):

	Urban	Rural		$\text{terr}_1 = 2$	$\text{terr}_2 = 1$
Male	\$800	\$500	sex_1	$200 \times s_1 \times 2$	$200 \times s_1 \times 1$
Female	\$400	\$200	sex_2	$200 \times s_2 \times 2$	$200 \times s_2 \times 1$

Using the squared error bias function, we solve for the male relativity s_1. To avoid dealing with multiples of 100, we choose a base pure premium of \$2 and we evaluate the observed pure premiums in multiples of \$100.

$$0 = \partial\text{SSE}/\partial s_1 = 2 \times (\$8 - \$2 \times s_1 \times 2) \times (-\$2 \times 2)$$
$$+ 2 \times (\$5 - \$2 \times s_1 \times 1) \times (-\$2 \times 1)$$

$$-64 + 32s_1 - 20 + 8s_1 = 0$$
$$40s_1 = 84$$
$$s_1 = 2.1.$$

Similarly, we solve for the female relativity (s_2):

$$0 = \partial\text{SSE}/\partial s_2 = 2 \times (\$4 - \$2 \times s_2 \times 2) \times (-\$2 \times 2)$$
$$+ 2 \times (\$2 - \$2 \times s_2 \times 1) \times (-\$2 \times 1)$$

$$-32 + 32s_2 - 8 + 8s_2 = 0$$
$$-40 + 40s_2 = 0$$
$$s_2 = 1.$$

We now discard the starting values of $t_1 = 2$ and $t_2 = 1$. Using the intermediate values of $s_1 = 2.1$ and $s_2 = 1$, we set the partial derivatives of the sum of the squared errors with respect to t_1 and t_2 equal to zero and we solve for new values of t_1 and t_2. We continue in this fashion until the series converges.

Iteration	Urban	Rural	Male	Female
Initial	2.0000	1.0000		
1-a			2.1000	1.0000
1-b	1.9224	1.1553		
2-a			2.1029	0.9940
2-b	1.9223	1.1555		
Normalized	1.6636	1.0000	2.1155	1.0000
Normalized Base Pure Premium	$\$200.00 \times 1.1555 \times 0.9940 = \$229.71$			

Squared Error and χ-Squared Intuition

The properties of squared-error minimization in the minimum bias procedure are unlike the properties of squared-error minimization in other statistical problems, as explained below. We note first that the bias function makes a difference, even in this simple illustration.

A. The balance principle bias function gives $s_1 = 2.1667$ and $s_2 = 1$.

B. The squared error bias function gives $s_1 = 2.1155$ and $s_2 = 1$.

The balance principle ensures that the total error in each classification dimension is zero. The squared-error bias function minimizes the aggregate squared error, and the χ-squared bias function minimizes the aggregate squared error as percentages of the expected values. The squared-error and χ-squared bias functions place more weight on outlying cells, where the squares of the errors are large. The balance principle and the squared-error bias function place more weight on the cells with large dollar values.

Illustration 16: A classification system with two dimensions has male versus female in one dimension and territories 1, 2, and 3 in the other dimension. The starting relativities are 1.00, 2.00, and 3.00 for territories 1, 2, and 3. The observed loss costs for the three territories in the male row are \$2, \$4, and \$12, with equal exposures in each cell. We assume a base pure premium of \$1.00.

	Territory 1 (1.00)	Territory 2 (2.00)	Territory 3 (3.00)
Male	\$2.00	\$4.00	\$12.00
Female	—	—	—

We want to determine the indicated relativity for males. Our concern here is not to solve this problem but to understand the effects of the different bias functions.

- If the male relativity is 2.00, the indicated pure premiums are \$2, \$4, and \$6. The first two cells have a perfect fit, and the third cell is too low by \$6.
- If the male relativity is 4.00, the indicated pure premiums are \$4, \$8, and \$12. The first two cells are too high by a total of \$6, and the third cell has a perfect fit.

The balance principle considers the first power of the errors. The average observed loss cost is (\$2 + \$4 + \$12)/3 = \$6.00. The average territory relativity is 2.00. To achieve balance, we choose a male relativity of 3.00. The indicated pure premiums are \$3, \$6, and \$9. The first two cells are too high by a total of \$3, and the third cell is too low by \$3. The indicated male/female relativity is \$6/\$2 = 3.00.

If we optimize with the balance principle, the sum of the squared errors is $(3-2)^2 + (6-4)^2 + (9-12)^2 = 14$. We compare this figure with the result of the least squares bias function.

The squared error bias function is more concerned with the large error in territory 3 than with the small errors in territories 1 and 2. To minimize the sum of squared errors, we increase the male relativity slightly, reducing the error in territory 3 and increasing the errors in territories 1 and 2.

Squared Error Minimization

To solve the problem using a squared error bias function, we minimize the sum of squared errors:

$$\text{SSE} = (2 - x)^2 + (4 - 2x)^2 + (12 - 3x)^2.$$

Taking the partial derivative with respect to x and setting it equal to zero gives

$$\partial\text{SSE}/\partial x = 2(2 - x)(-1) + 2(4 - 2x)(-2) + 2(12 - 3x)(-3) = 0$$

$$4 + 16 + 72 = 2x + 8x + 18x$$

$$92 = 28x$$

$$x = 92/28 = 3.286.$$

The sum of the squared errors is $(3.286 - 2)^2 + (6.571 - 4)^2 + (9.857 - 12)^2 = 12.857$, which is less than the squared error of 14 under the balance principle. Minimizing the sum of the squared errors yields 3.286, not the average, which is 3.00.

Squared Error Minimization

The illustration above seems odd to some statisticians. We are choosing a value to minimize the squared error among a series of observations. An elementary statistical theorem is that the average minimizes the sum of the squared errors. This seems inconsistent with the comments above.

When we set rates in a single dimension class system, minimizing the squared error produces the arithmetic average. The following illustration explains this.

Illustration 17: We are measuring a patient's fever with an old, imprecise thermometer. The thermometer is unbiased, but the observed readings are distorted by sampling error. We perform nine trials, and we observe readings of

$(100.1, 100.2, \ldots, 100.9)$. (The readings were not in this order, so there is no trend; we have simply arranged them in ascending numerical order.) Using the least squared error function, we determine the best estimate of the patient's temperature.

We rephrase the illustration mathematically. We have observed values of $z_1, z_2, \ldots, z_n$, and we must choose a single value—call it z^*—to minimize the squared error.

The sum of the squared errors is $\sum(z_i - z^*)^2$. The partial derivative of this sum with respect to z^* is $\sum 2(z_i - z^*)(-1)$. Setting this equal to zero gives $z^* = (\sum z_i)/n$. The indicated z^* is the average of the z_is.

In the temperature measurement illustration, the average of the nine observations is 100.5. This is the solution using the squared error bias function.

If we had chosen instead some other value, such as 100.3, we could correct this estimate by the average of the errors. The error in each observation is the observation minus 100.3. This is the series $(-0.02, -0.01, 0, +0.01, \ldots, +0.06)$. The average is $+0.02$. The corrected estimate is $100.3 + 0.02 = 100.5$.

This is not true for multidimensional systems. In a multiplicative model with two dimensions, the z_is are the observed values. The z^* is the indicated relativity for one of the two dimensions. The other dimension has relativities of $y_1, y_2, \ldots, y_m$.

The sum of the squared errors is $\sum\sum(z_i - y_j \times z^*)^2$. The partial derivative of this sum with respect to z^* is $\sum\sum 2(z_i - y_j \times z^*)(-y_j)$. Setting this equal to zero gives $z^* = (\sum\sum z_i)/\sum y_j^2$.

The indicated z^* is no longer the average of the z_is. Rather, this result is the solution to the minimum bias procedure using the squared error bias function, as we show next.

Balance Principle Optimization

When we seek a pure premium for one dimension, minimizing the squared error produces the arithmetic average. With two or more dimensions, the balance principle selects the multidimensional equivalent to the mean of each class across the other dimension(s).

The balance principle provides an unbiased solution; Bailey [1963] considers it the only unbiased solution (see below). Some actuaries believe that an unbiased solution is more likely to maximize the firm's profitability than a biased solution.[21]

General Squared Error Minimization, Multiplicative Model

We consider a more general two dimensional classification system. The base pure premium is B. We again assume one exposure per cell (or the same number of exposures per cell) to keep the equations simple. In practice, one must multiply all terms by the number of exposures.

Suppose we have two dimensions, age of driver and territory, with n age classes and m territories. The observed loss cost in the ith age class and the jth territory is r_{ij}. The indicated pure premium in the ith age class and the jth territory is $B \times x_i \times y_j$.

The squared error in any cell is $(r_{ij} - Bx_iy_j)^2$. The sum of the squared errors is

$$Q = \sum_{i=1}^{n}\sum_{j=1}^{m}(r_{ij} - Bx_iy_j)^2.$$

We take partial derivatives with respect to each variable and set them equal to zero. The $(n+m)$ equations are not linear, so we must search for a solution by numerical methods. We choose

[21]There are exceptional scenarios when a different bias function may be better. In a jurisdiction that places restrictions on risk classification, the bias function may have to be changed to accommodate these restrictions. If the insurer seeks to expand in certain classifications for competitive or marketing reasons, the minimum bias procedure may not accommodate the insurer's strategy. In most scenarios, however, the balance principle serves the economic interests of the firm.

starting values for one dimension—say, the y_j. To solve for the value of x_i, we take the partial derivative with respect to x_i and set it equal to zero:

$$\sum_{j=1}^{m} 2(r_{ij} - Bx_i y_j)(-By_j) = 0.$$

This gives

$$x_i = \sum_{j=1}^{m}(r_{ij} \times y_j) \Big/ \sum_{j=1}^{m} By_j^2.$$

The x_i is a variable. The y values are fixed; they are not variables once we have assigned starting values to the y values.

We use this procedure to solve for $x_1, x_2, \ldots, x_n$. We then discard the starting y values and solve for new values of the y variables using the same procedure as we used for the x variables.

We have $(n+m)$ variables, and we have $(n+m)$ equations. The constraints for least squares minimization are the same as the constraints for the balance principle. There is one totality constraint, since taking the sum of the squared errors along the rows is the same as taking the sum of the squared errors along the columns. This means that we have only $(n+m-1)$ independent equations. In addition, we could multiply all the relativities along any dimension by a constant and divide the base pure premium by the same constant.

Squared Error Minimization, Additive Model

We can also use an additive model with the least squares bias function. We first show the results for the elementary 2 by 2 illustration. Below are the same observed loss costs and indicated pure premiums we have been using.

	Urban	Rural		terr_1	terr_2
Male	\$800	\$500	sex_1	$200 + s_1 + t_1$	$200 + s_1 + t_2$
Female	\$400	\$200	sex_2	$200 + s_2 + t_1$	$200 + s_2 + t_2$

As mentioned earlier, there are three mathematically equivalent ways of defining the additive model; the solution method is the same for each of them. The pure premium in cell $x_i y_j$ is

A. Base pure premium $+ x_i + y_j$,

B. Base pure premium $\times (1 + u_i + v_j)$, or

C. Base pure premium $\times (p_i + q_j)$.

We use the first of these three equations here for its intuitive simplicity. Note that a multiplicative relationship between the base pure premium and the relativities does not make the model multiplicative. If the relationship among the factors is additive, the model is additive. A combined multiplicative and additive model has relationships among the relativities that are both multiplicative and additive.

The method used here is the same as the method used for the multiplicative model above. For the male urban cell, the squared error is $(\$800 - \$200 - s_1 - t_1)^2$. The sum of the squared errors for all four cells is

$$Q = (\$800 - \$200 - s_1 - t_1)^2 + (\$500 - \$200 - s_1 - t_2)^2 + (\$400 - \$200 - s_2 - t_1)^2 + (\$200 - \$200 - s_2 - t_2)^2.$$

We take partial derivatives with respect to each variable and set them equal to zero. The partial derivative with respect to s_1 is

$$\partial Q/\partial s_1 = 2(\$800 - \$200 - s_1 - t_1)(-1) + 2(\$500 - \$200 - s_1 - t_2)(-1) = 0,$$

or

$$s_1 = (\$900 - t_1 - t_2)/2.$$

For the additive model with the least squares bias function, the simultaneous equations are linear, and we can solve them directly. Nevertheless, it is easier to program the solution using numerical methods. If we choose starting values of $t_1 = \$250$

and t_2 = \$0, we get s_1 = \$325. We leave it to the reader to verify that the relativities converge to the same figures as the additive model with the balance principle.

General Squared Error Minimization, Additive Model

For the general formula, we let B = the base pure premium. The sum of the squared errors is

$$\mathrm{SSE} = \sum_{i=1}^{n}\sum_{j=1}^{m}(r_{ij} - B - x_i - y_j)^2.$$

We take the partial derivative with respect to x_i and set it equal to zero:

$$\frac{\partial \mathrm{SSE}}{\partial x_i} = \sum_{j=1}^{m} 2(r_{ij} - B - x_i - y_j)(-1) = 0,$$

or

$$x_i = \sum_{j=1}^{m}(r_{ij} - y_j)/m - B.$$

6. THE χ-SQUARED BIAS FUNCTION

The χ-squared bias function is similar to the squared error bias function, except that each squared error is divided by the expected value in that cell. We define the χ-squared bias function, and then we apply it to the minimum bias procedure.

Illustration 18: Suppose the expected distribution in two cells is 40%–60% and the observed distribution is 30%–70%. The squared error is $(40\% - 30\%)^2 + (60\% - 70\%)^2 = 2.00\%$; the χ-squared error is $(40\% - 30\%)^2/40\% + (60\% - 70\%)^2/60\% = 4.17\%$.

We show the application of the χ-squared bias function to the multiplicative illustration.

	Urban	Rural		terr_1	terr_2
Male	\$800	\$500	sex_1	$200 \times s_1 \times t_1$	$200 \times s_1 \times t_2$
Female	\$400	\$200	sex_2	$200 \times s_2 \times t_1$	$200 \times s_2 \times t_2$

The χ-squared value for each cell is (observed value – expected value)^2/expected value. For urban male drivers in our basic illustration, this number is

$$(\$800 - \$200 \times s_1 \times t_1)^2/(\$200 \times s_1 \times t_1).$$

We sum the squared errors for the four cells to get the sum of χ-squared values:

$$\begin{aligned}
\text{SSE} = {} & (\$800 - \$200 \times s_1 \times t_1)^2/(\$200 \times s_1 \times t_1) \quad \text{urban male} \\
& + (\$500 - \$200 \times s_1 \times t_2)^2/(\$200 \times s_1 \times t_2) \quad \text{rural male} \\
& + (\$400 - \$200 \times s_2 \times t_1)^2/(\$200 \times s_2 \times t_1) \quad \text{urban female} \\
& + (\$200 - \$200 \times s_2 \times t_2)^2/(\$200 \times s_2 \times t_2) \quad \text{rural female.}
\end{aligned}$$

To minimize the sum of the squared errors, we take partial derivatives with respect to each variable and set them to zero. For the male classification relativity (s_1), we have[22]

$$\begin{aligned}
0 &= \partial\text{SSE}/\partial s_1 \\
&= [(\$200 \times s_1 \times t_1) \times 2 \times (\$800 - \$200 \times s_1 \times t_1) \times (-\$200 \times t_1) \\
&\quad - (\$800 - \$200 \times s_1 \times t_1)^2 \times (\$200 \times t_1)]/(\$200 \times s_1 \times t_1)^2 \\
&\quad + [(\$200 \times s_1 \times t_2) \times 2 \times (\$500 - \$200 \times s_1 \times t_2) \times (-\$200 \times t_2) \\
&\quad - (\$500 - \$200 \times s_1 \times t_2)^2 \times (\$200 \times t_2)]/(\$200 \times s_1 \times t_2)^2 \\
&\quad + 0 + 0.
\end{aligned}$$

Although the arithmetic looks cumbersome, the equation can be simplified. To avoid needless arithmetic, we derive the general solution, and we then resume the illustration.

[22]We use the quotient rule that if $y(x) = f(x)/g(x)$, then $dy/dx = [g(x) \times df/dx - f(x) \times dg/dx]/g^2(x)$.

χ-Squared Recursive Equations

We show the general recursive equations for the χ-squared bias function with two classes in each of two dimensions; the extension to three or more dimensions is straightforward. To save space, we include the number of exposures and derive the final recursive equations.

	y_1	y_2		y_1	y_2
x_1	n_{11}	n_{12}	x_1	$B \times x_1 \times y_1$	$B \times x_1 \times y_2$
x_2	n_{21}	n_{22}	x_2	$B \times x_2 \times y_1$	$B \times x_2 \times y_2$

We form the χ-squared bias function as a double summation covering all the cells in the array.

$$\text{SSE} = \sum_{i=1}^{n}\sum_{j=1}^{m} \frac{(n_{ij}r_{ij} - n_{ij}Bx_i y_j)^2}{n_{ij}Bx_i y_j}.$$

We factor out the number of exposures in each cell from the summand to give

$$\text{SSE} = \sum_{i=1}^{n}\sum_{j=1}^{m} \frac{n_{ij}(r_{ij} - Bx_i y_j)^2}{n_{ij}Bx_i y_j}.$$

We seek to minimize the χ-squared value. To simplify the mathematics, we use a base pure premium of \$1.00. In practice, we would choose the base pure premium at the end of the procedure to set the base relativities in each dimension to unity.

As before, given starting values for either dimension, we determine the intermediate values for the other dimension. Assume we have chosen starting values for the y relativities and we are solving for an intermediate value of x_i. Only the cells in the ith row have terms with x_i in them. We take the partial derivative of this row with respect to x_i, and we set it equal to 0.

In the equation below, we have differentiated with respect to x_i, and the summation is over the j dimension. The value of i is fixed.

$$\sum_{j=1}^{m} \frac{n_{ij}[x_i y_j 2(r_{ij} - x_i y_j)(-y_j) - (r_{ij} - x_i y_j)^2 y_j]}{(x_i y_j)^2} = 0.$$

The value $x_i = 0$ does not minimize the equation (or all the indicated pure premiums would be zero), so we multiply both sides of the equation by $(x_i)^2$. Simplifying further,

$$\sum_{j=1}^{m} \left[-2n_{ij} x_i (r_{ij} - x_i y_j) - \frac{n_{ij}}{y_j}(r_{ij} - x_i y_j)^2 \right] = 0.$$

We expand the square and combine like terms:

$$\sum_{j=1}^{m} \left[-2n_{ij} x_i r_{ij} + 2n_{ij} x_i^2 y_j - \left(\frac{n_{ij}}{y_j}\right) r_{ij}^2 + 2n_{ij} x_i r_{ij} - n_{ij} x_i^2 y_j \right] = 0,$$

$$\sum_{j=1}^{m} \left[n_{ij} x_i^2 y_j - \left(\frac{n_{ij}}{y_j}\right) r_{ij}^2 \right] = 0.$$

This gives a relatively simple expression for each x_i in terms of the y_j values:

$$x_i = \left[\sum_{j=1}^{m} \left(\frac{n_{ij} r_{ij}^2}{y_j} \right) \Big/ \sum_{j=1}^{m} n_{ij} y_j \right]^{0.5}.$$

In the illustration used here, there is one exposure in each cell. The starting values are $t_1 = 2$ and $t_2 = 1$. We use a base pure premium of \$200, and we divide all cells by \$200.

	Urban	Rural		$\text{terr}_1 = 2$	$\text{terr}_2 = 1$
Male	\$4	\$2.5	sex_1	$s_1 \times 2$	$s_1 \times 1$
Female	\$2	\$1	sex_2	$s_2 \times 2$	$s_2 \times 1$

Using the χ-squared bias function along the first row, we get

$$s_1(\text{male relativity}) = [(4^2/2 + 2.5^2/1)/(2+1)]^{0.5} = 2.179.$$

Using the χ-squared bias function along the second row, we get

$$s_2(\text{female relativity}) = [(2^2/2 + 1^2/1)/(2+1)]^{0.5} = 1.000.$$

The male-to-female relativity is 2.179 to 1. The series converges.

Iteration	Urban	Rural	Male	Female
Initial	2.0000	1.0000		
1-a			2.1794	1.0000
1-b	1.8887	1.1029		
2-a			2.1739	1.0055
2-b	1.8884	1.1032		
3-a			2.1739	1.0055
Normalized	1.7118	1.0000	2.1620	1.0000
Normalized Base Pure Premium	$\$200.00 \times 1.1032 \times 1.0055 = \$221.85$			

The final relativities are 2.1620 for the χ-squared bias function and 2.1155 for the least squares bias function. The dollar values in the urban male cell are larger than the dollar values in the rural male cell, so the least squares bias function gives more weight to the urban male cell as compared to the rural male cell than the χ-squared bias function gives.

Additive Model with χ-Squared

The χ-squared bias function can be used with any type of model, whether multiplicative, additive, or combined. If an additive model is used, we minimize the following expression:

$$\text{SSE} = \sum_{i=1}^{n}\sum_{j=1}^{m} n_{ij}(r_{ij} - x_i - y_j)^2/(x_i + y_j).$$

We set the partial derivative with respect to each relativity equal to zero. It is easiest to solve the resulting set of simultaneous equations by iteration, solving for Δx_i rather than for x_i. Bailey and Simon [1960], followed by Brown [1988], give the recursive equations as

$$\Delta x_i = \frac{\sum_j n_{ij}\left(\frac{r_{ij}}{x_i + y_j}\right)^2 - \sum_j n_{ij}}{2\sum_j n_{ij}\left(\frac{r_{ij}}{x_i + y_j}\right)^2\left(\frac{1}{x_i + y_j}\right)}.$$

The series converges along the following path.

Iteration	Urban	Rural	Male	Female
Initial	\$250.00	\$0.00		
1			\$96.81	(\$15.53)
2	\$334.23	\$63.18		
3			\$193.97	(\$115.14)
4	\$349.61	\$97.35		
5			\$220.32	(\$112.08)
6	\$336.84	\$103.42		
7	\$224.64	(\$113.40)		
Final	\$233.43	\$0.00	\$338.04	\$0.00
Base Pure Premium				\$190.02

7. MAXIMUM LIKELIHOOD

Some statisticians prefer a maximum likelihood bias function to either a χ-squared or a least squares bias function when fitting a distribution to observed data. In his 1988 *Proceedings* paper, Rob Brown used a maximum likelihood bias function to optimize classification relativities. The maximum likelihood bias function is rarely used in practical work, and not all actuaries are familiar with it.

The maximum likelihood bias function requires an assumption about the distribution of values in each class. The appropriate distribution for loss costs is not evident. It probably is not a simple parametric distribution, such as an exponential distribution or a Poisson distribution. If there is not support for a specific distribution, the merits of a maximum likelihood bias function are less clear.

Likelihood and Probability

We use the term likelihood, not probability. For a continuous distribution, the probability of observing a specific value is zero. If the exponential distribution function has λ of 0.0001, the likelihood of a loss of size \$20,000 is $0.0001 \times e^{-2}$.

Illustration 19: We are fitting an exponential curve to a set of insurance losses. For the exponential distribution function, the likelihood of a loss of size x is $\lambda e^{-\lambda x}$. We use integration by parts to solve for the mean of the exponential distribution function:

$$\int_0^\infty x\lambda e^{-\lambda x}dx = \left(-x\lambda e^{-\lambda x} - \frac{1}{\lambda}e^{-\lambda x}\right)\Big|_0^\infty = \frac{1}{\lambda}.$$

To fit an exponential curve to a set of insurance losses, we must determine λ. After discussing two methods outlined above (for comparison), we show the maximum likelihood method.

Method of Moments

The mean of the exponential distribution is $1/\lambda$. We take the average of the observations, and we set λ equal to the reciprocal of this average.

Least Squares

We divide the loss sizes into ranges, such as \$0 to \$5,000, \$5,001 to \$25,000, \$25,001 to \$100,000, and so forth. We calculate the percentage of observed losses that fall into each range. For any given λ, we determine the percentage of theoretical losses that would fall into each range.

For each range, we calculate the squared difference between the observed percentage and the theoretical percentage. We sum the squared differences over all the ranges. The result is a function of λ. To minimize this squared difference, we set the derivative with respect to λ equal to zero.

χ-Squared

The χ-squared procedure is similar to the least squares procedure, but instead of taking the squared difference we take the χ-squared difference. For each range, we divide the squared difference by the expected value.

Maximum Likelihood

We explain the method by means of an illustration. Suppose we have observed five losses with sizes of \$3,000, \$5,000, \$15,000, \$20,000, and \$80,000. For a given value of λ, the likelihood of a loss equal to \$3,000 is $\lambda e^{-\lambda\times 3{,}000}$. The likelihood of five losses for the values listed above is the product of the likelihoods of each individual loss, or

$$L = \lambda e^{-\lambda\times 3{,}000} \times \lambda e^{-\lambda\times 5{,}000} \times \lambda e^{-\lambda\times 15{,}000} \times \lambda e^{-\lambda\times 20{,}000} \times \lambda e^{-\lambda\times 80{,}000}.$$

We simplify the likelihood to $\lambda^5 e^{-\lambda\times 123{,}000}$ (\$123,000 is the sum of the losses). To find the λ that gives the greatest likelihood, we set the derivative with respect to λ equal to zero.

Before taking the derivative, we make one simplification. Maximizing a strictly increasing function, like the likelihood function, is the same as maximizing its logarithm. The logarithm of the likelihood (the log-likelihood, or LL) is

$$\mathrm{LL} = \ln L = 5\ln\lambda - 123{,}000\times\lambda$$

$$d(\ln L)/d\lambda = 5/\lambda - 123{,}000 = 0, \quad \text{or} \quad \lambda = 5/123{,}000.$$

Maximum Likelihood and Minimum Bias Procedure

The rating model uses the classification relativities to determine the expected loss in each cell. The maximum likelihood test is most practicable as a bias function when a single parameter distribution is used and the mean of the distribution equals the parameter itself or some simple function of the parameter, such as its reciprocal. It is most valuable when the distribution is a reasonable reflection of the insurance process.

The exponential and Poisson distributions have these properties. We illustrate a multiplicative model with the exponential distribution function, using the same illustration as before.

	Urban	Rural		terr_1	terr_2
Male	\$800	\$500	sex_1	$200 \times s_1 \times t_1$	$200 \times s_1 \times t_2$
Female	\$400	\$200	sex_2	$200 \times s_2 \times t_1$	$200 \times s_2 \times t_2$

Each class has an assumed exponential distribution of loss costs. If the indicated pure premium is \$200, we expect the observed losses to follow an exponential distribution with a mean of \$200. The λ differs by cell. The indicated pure premium in each cell is $1/\lambda$.

Illustration 20: For the urban male cell, the loss costs have an exponential distribution with the parameter λ equal to $1/(\$200 \times s_1 \times t_1)$.

We choose starting values for $t_1 = 2.00$ and $t_2 = 1.00$. We determine the likelihood of the observed loss costs. The value of λ for the urban male cell is $1/(200 \times s_1 \times t_1) = 1/(400 \times s_1)$. The likelihood of the \$800 loss cost in the urban male cell is

$$\frac{1}{400 s_1} e^{-800/400 s_1} = \frac{1}{400 s_1} e^{-2/s_1}.$$

The likelihoods of the observed values in the other cells are determined in the same manner. To maximize the likelihood, we maximize the log-likelihood. To repeat,

- The likelihood of four observed values is the product of the four individual likelihoods.

- The log-likelihood of four observed values is the sum of the four individual log-likelihoods.

The partial derivative of the log-likelihood with respect to s_1 depends on the log-likelihoods in the male row only. This is the same simplification that we used for the least squares method and the χ-squared method.

The log-likelihood of the values in the male row is

$$\begin{aligned}\text{LL} = &-\ln(400s_1) - (800/400)\times 1/s_1 - \ln(200s_1)\\ &-(500/200)\times 1/s_1\end{aligned}$$

$$\begin{aligned}\partial\text{LL}/\partial s_1 &= -1/s_1 + 2s_1^{-2} - 1/s_1 + 2.5s_1^{-2} = 0\\ -s_1 + 2 - s_1 + 2.5 &= 0, \quad \text{because} \quad s_1 \neq 0\\ s_1 &= 2.25.\end{aligned}$$

The log-likelihood of the values in the female row is

$$\begin{aligned}\text{LL} = &-\ln(400s_2) - (400/400)\times 1/s_2 - \ln(200s_2)\\ &-(200/200)\times 1/s_2\end{aligned}$$

$$\begin{aligned}\partial\text{LL}/\partial s_2 &= -1/s_2 + 1s_2^{-2} - 1/s_2 + 1s_2^{-2} = 0\\ -s_2 + 1 - s_2 + 1 &= 0\\ s_2 &= 1.00.\end{aligned}$$

The series converges to the following relativities.

Iteration	Urban	Rural	Male	Female
Initial	2.0000	1.0000		
1-a			2.2500	1.0000
1-b	1.8889	1.0556		
2-a			2.2430	1.0031
2-b	1.8886	1.0557		
3-a			2.2430	1.0031
Normalized	1.7889	1.0000	2.2361	1.0000
Normalized Base Pure Premium	$200.00 × 1.0557 × 1.0031 = $211.80			

Justification

If the distribution of loss costs is a simple mathematical function, such as a Poisson distribution, a normal distribution, a lognormal distribution, or an exponential distribution, we can derive simple recursive equations; see Brown [1988]. In practice, we don't know the proper distributions. The distributions that have been suggested for use in the minimum bias procedure are not necessarily correct. They are simply tractable.

The Poisson distribution is a reasonable model for loss frequency distributions, though not for loss severity distributions. The normal, lognormal, and exponential distributions may not be ideal fits to the loss costs distribution. However, maximum likelihood estimation is particularly useful when examining loss frequency and loss severity with generalized linear models; see Feldblum et al. [forthcoming].

The Bias Function

The optimal class relativities for a given data set depend on the choice of the bias function. The choice of bias function can be viewed from three perspectives:

1. Mathematical tractability,
2. Social equity, or
3. Economic optimization.

Mathematical tractability was of concern when computational capacity was limited and some bias functions gave simple relationships while other bias functions gave intractable equations. The minimum bias procedure gives simple equations for the bias functions discussed in this paper. With modern spreadsheets, however, even the average absolute error does not pose tractability issues. Just as the solution for the balance principle is the mean, the solution for the average absolute error is the median. It is not uncommon for actuaries to use the median instead of the mean in practical problems.

Social equity is subjective, though it is vital to the success of a highly regulated industry like insurance. The balance principle sometimes results in large errors for outlying cells. The errors may be particularly large for high-rated cells. If a multiplicative model is used when an additive model is more appropriate, the errors for outlying cells are frequently overcharges.

Of the bias functions that we consider in this paper, the squared-error bias function is the best at reducing large overcharges for individual cells. Ferreira's critique of insurance industry classification systems in Massachusetts illustrates this social position.[23]

Economic optimization drives the behavior of firms in free markets. Firms seek to maximize profits and to minimize losses (among other firm objectives). Suppose an insurer issues three policies. It must choose between two rating systems:

A. Under rating system A, it expects to lose \$1.00 each on the first two policies and to break even on the third policy.

B. Under rating system B, it expects to break even on the first two policies and to lose \$1.50 on the third policy.

Rating system A is off by \$2.00 using the balance principle while rating system B is off by \$1.50. Using the squared error bias function, rating system A is off by 2.00 dollars-squared while rating system B is off by 2.25 dollars-squared. The balance principle says we should choose rating system B, and the squared error bias function says we should choose rating system A.

To maximize profits (or minimize losses), we would probably prefer rating system B, as the balance principle says. In practice,

[23] See Ferreira [1978], as well as Cummins et al., [1983] chapter 4. We are not endorsing Ferreira's views, which are inconsistent with competitive insurance markets; see the discussion in the text of this paper.

economic forces are more complex than short-term profit maximization. There are many reasons for avoiding errors, including consumer dissatisfaction, consumer switching, and public relations. In democratic systems where social opinion and political pressures are strong, firms may sacrifice short-term profit maximization to achieve other ends; objectives such as workforce diversity and environmental protection are examples. Furthermore, manager incentives may encourage the pursuit of other goals, such as corporate growth instead of profit maximization. Nevertheless, profit maximization remains the dominant corporate goal. The pricing actuary should keep these social and economic considerations in mind when choosing a bias function for the minimum bias procedure.

8. COMBINED MODELS

Throughout this paper, we have used simple multiplicative and additive models, not combined models. This reflects current insurance practice.

In truth, business practice reflects ratemaking capabilities. Actuaries have not had simple methods to optimize combined models, so these models have not gained wide acceptance.

The rationale for combined models is strong. Since the least squares and χ-squared bias functions provide simple recursive equations for many combined models, these models may become more popular in the future.[24]

Illustration 21: Rating territory may have a variety of effects on insurance loss costs.

1. High-crime areas may have a greater incidence of car theft and claim fraud. Thefts would raise comprehensive pure premiums, and fraud would raise liability pure premiums.

[24] Generalized linear models allow the optimization of even more complex rating models; see Feldblum et al. [forthcoming].

2. Areas with more sophisticated medical facilities may have higher loss costs for bodily injury claims.
3. Territories with a higher number of attorneys per capita may experience a higher incidence of bodily injury claims per physical accident.[25]

The first effect argues for an additive model; the third effect suggests a multiplicative model; and the second effect may have both additive and multiplicative components. The greater incidence of theft may be unrelated to other hazards, whereas a higher proportion of attorneys may affect claim filing for all hazards.[26] Intuition alone, though, is rarely sufficient to optimize a rating model. The minimum bias method allows the pricing actuary to determine the optimal rating structure from the observed loss costs.

Illustration 22: Combined Model: We keep the same male/female and urban/rural classification system. We assume now that the male/female rating dimension has a multiplicative effect on loss costs, and the rating territory dimension has both a multiplicative and an additive effect on loss costs. We show the structure of this rating model, and we explain how to optimize it.

For the male/female classification dimension, we use pure premium relativities of s_1 and s_2. For the urban/rural dimension, each class has two relativities: a multiplicative relativity denoted by t_1 and t_2, and an additive relativity denoted by z_1 and z_2. We denote the base pure premium as B.

The indicated pure premium for any class is $B \times (s_i \times t_j + z_j)$. The subscripts i and j denote the classification dimension. The indicated pure premiums are shown in the following table.

[25]See Conners and Feldblum [1998] for the effects of territory on private passenger automobile claim frequency.

[26]On reviewing this paper, Ginda Fisher suggested that rating variables such as fire protection, theft protection devices, and age of dwelling may be additively related for Homeowners insurance, though there may also be some multiplicative relationships among them.

	Observed Loss Costs Urban	Observed Loss Costs Rural		Indicated Pure Premiums Urban (t_1, z_1)	Indicated Pure Premiums Rural (t_2, z_2)
Male	$r_{11} = \$800$	$r_{12} = \$500$	Male (s_1)	$B \times (s_1 \times t_1 + z_1)$	$B \times (s_1 \times t_2 + z_2)$
Female	$r_{21} = \$400$	$r_{22} = \$200$	Female (s_2)	$B \times (s_2 \times t_1 + z_1)$	$B \times (s_2 \times t_2 + z_2)$

If we use the balance principle as the bias function, we balance along the two rows and the two columns. This gives four equations, of which only three are independent, since there is a totality constraint. We must solve for six classification relativities.

When there are more unknowns than equations, the iterations will not necessarily converge. If they do converge, the convergence is generally not unique. If the balance principle is used with a multidimensional combined multiplicative and additive model, there are more relativities than there are equations.

If we use a least squares or a χ-squared bias function, the combined model is not conceptually different from a simple multiplicative or additive model. We set the partial derivative with respect to each rating variable equal to zero. This guarantees the same number of equations as rating variables.

The use of the minimum bias procedure with combined models is a powerful rating tool. But as the rating models grow more complex, there are more classification dimensions, more cells, and fewer exposures in each cell. The potential rating errors become more serious as the effect of random loss fluctuations grows.

Outliers

The least squares and χ-squared bias functions are particularly sensitive to outliers. Outliers are observed values that differ substantially from their expected values because of random loss fluctuations. Distortions stemming from random loss fluctuations can be controlled in several ways:

- Losses can be capped at basic limits or similar retentions.
- Low-volume classes can be assigned limited credibility.
- The data in each cell can be examined for unusual values.

The use of low retentions or low credibility conflicts with the objective of basing rates on observed experience as much as possible. The examination of the observed data for unusual values is sometimes too time-consuming for practical work. In any case, one should choose a bias function that is not too sensitive to outliers.

Illustration 23: A classification system has two dimensions: male/female along one dimension and ten territories along the other dimension. The current driver relativities are 1.00 for female and 2.00 for male. The current territorial relativities are $1.00, 2.00, \ldots, 10.00$ for the ten territories, labeled $(01, 02, \ldots, 10)$. The base pure premium is \$100, and a multiplicative model is used.

Scenario A: The observed loss costs are shown below, in units of 100 dollars.

Territory:	01	02	03	04	05	06	07	08	09	10
Male	\$2	\$4	\$6	\$8	\$10	\$12	\$14	\$16	\$18	\$20
Female	\$1	\$2	\$3	\$4	\$5	\$6	\$7	\$8	\$9	\$10

The observed loss costs exactly match the indicated pure premiums in the current rating system. No change to the current relativities is indicated.

Scenario B: Because of a random large loss, the observed loss costs for the males in territory 10 are \$10,000 instead of \$2,000. The "territory 10 male" cell shows \$100 instead of \$20. This type of random loss fluctuation is common in classification analysis for small populations. We have starting values of $(1.000, 2.000, \ldots, 10.000)$ for the ten territories. We determine the intermediate value for the male relativity.

The balance principle selects the male relativity s_1 such that (on the first iteration)

$$(s_1 \times t_1) + (s_1 \times t_2) + \cdots + (s_1 \times t_{10}) = r_{1,1} + r_{1,2} + \cdots + r_{1,10}$$

$$s_1 \times \$55 = \$190$$

$$s_1 = 3.455.$$

The least squares bias function selects the male relativity to minimize the squared error:

$$\mathrm{SSE} = \sum_{j-1}^{m} (r_{1j} - s_1 \times t_j)^2$$

$$\partial \mathrm{SSE}/\partial s_1 = \sum_{j=1}^{m} 2(r_{1j} - s_1 \times t_j) \times (-t_j) = 0$$

$$s_1 = \frac{\sum_{j=1}^{m}(r_{1j} \times t_j)}{\sum_{j=1}^{m} t_j^2}$$

$$[(1 \times 2) + (2 \times 4) + (3 \times 6) + \cdots + (9 \times 18) + (10 \times 100)]/$$
$$[1^2 + 2^2 + 3^2 + \cdots 9^2 + 10^2] = 4.078.$$

Compared with the balance principle, the least squares bias function exacerbates the distortion caused by random loss fluctuations. In this instance, the χ-squared bias function magnifies the distortion less than the least squares bias function does. This is not always the case; in other instances, the χ-squared bias function magnifies the distortion more than the least squares bias function does. Since combined models are more sensitive to random loss fluctuations than simple models are, and since the least-squares or χ-squared bias function must be used, the pricing actuary must be particularly careful to exclude outliers from the data.

9. GOODNESS-OF-FIT

For a given rating model and bias function, the minimum bias procedure optimizes the relativities. We now wish to optimize the rating system by choosing the best rating model and bias function. The choice of rating model, such as multiplicative, additive, or combined, depends on the characteristics of the observed loss costs. The choice of the bias function depends on the objective:

- The statistician seeking the best fit might use a maximum likelihood function if a tractable distribution function is appropriate for this coverage, or a χ-squared function if the probability distribution function is not known or not tractable.
- The regulator seeking to avoid large dollar mismatches between observed loss costs and indicated pure premiums might use a least squares function.
- The insurer seeking to avoid monetary losses might use the balance principle.

The preferences listed above are examples; other preferences are also possible. In particular, a regulator might prefer the balance principle to provide the most efficient rating system.

Empirical Tests

We can test the choice of rating model empirically.

Illustration 24: We are using a χ-squared bias function to optimize classification relativities. We do not know whether a multiplicative model or an additive model is more appropriate.

We perform the minimum bias procedure twice with a χ-squared bias function: once with the multiplicative model and once with an additive model. After optimizing the relativities for each model, we compare the final χ-squared difference between the observed loss costs and the indicated pure premiums for each model. The model with the lower χ-squared is preferred.

Illustration 25: We are using the balance principle to optimize classification relativities. We do not know whether a multiplicative model or an additive model is more appropriate.

We perform the minimum bias procedure twice with a balance principle bias function: once with the multiplicative model and once with an additive model. After optimizing the relativities for each model, we compare the average absolute difference between the observed loss costs and the indicated pure premiums for each model. The model with the lower average absolute difference is preferred.

We cannot empirically test the suitability of the bias function, as explained below.

Illustration 26: We are using a multiplicative model, and we are deciding between the balance principle and the χ-squared function.

We perform the minimum bias procedure twice: once with the multiplicative model and the balance principle and once with the multiplicative model and a χ-squared bias function.

If we test the performance of the models by using a χ-squared test to measure the difference between the observed loss costs and the indicated pure premiums, the χ-squared bias function does better. This result is tautological, since the χ-squared bias function minimized the χ-squared difference between the observed loss costs and the indicated pure premiums.

If we test the performance of the two models by using the average absolute difference between the observed loss costs and the indicated pure premiums, the balance principle generally does better. The χ-squared bias function minimizes large percentage errors. The balance principle and the average absolute difference minimize dollar differences.

The choice of bias function is a qualitative choice, depending on the objectives of the rating system. It is not subject to a quantitative test of suitability.

Squared Error versus χ-Squared

The squared error bias function is similar to the χ-squared bias function, but whereas the squared error test looks at absolute differences, the χ-squared test looks at percentage differences. Some statisticians prefer the χ-squared test to a least squares test.

Illustration 27: We are fitting a distribution to two empirical data points:

- Point A has an observed value of \$101 and a fitted value of \$100.
- Point B has an observed value of \$1.50 and a fitted value of \$1.00.

We examine the errors for each point:

- The squared error is $(101 - 100)^2 = 1.00$ for point A and $(1.50 - 1.00)^2 = 0.25$ for point B. This distribution fits point B better.
- The χ-squared value is $(101 - 100)^2/100 = 0.01$ for point A and $(1.50 - 1.00)^2/1.00 = 0.25$ for point B. This distribution fits point A better.

The statistician might prefer the χ-squared test to the squared error test. The practical businessperson might argue that the insurance enterprise is not concerned with optimizing a statistical fit. It is concerned with optimizing net income. At point A, the insurer has a gain or loss of \$1.00. At point B, the gain or loss is \$0.50. The squared error test is preferred.

This argument does not fully reflect the purpose of the minimum bias procedure. The argument would be correct if we fully believed the observed values—that is, if the observed values were fully credible. But if the observed values were fully credible, we would have no need to use the minimum bias procedure; we would just use the rates indicated by the observed loss costs in each cell.

We are using the minimum bias procedure because the individual observed values are not fully credible, and we believe that the relationships among all the cells in the observed matrix provide useful information for choosing the true expected values. When we say that a particular fit X has less of an error than another fit Y, we do not mean that we know the true values and that model X is closer to these true values. We generally do not know the true values, but we presume that these true values might be represented by a mathematical function. When we say that fit X is better, we mean that model X is more likely to be a better model. The χ-squared bias function perhaps does a better job of choosing the better model. If so, the businessperson might also prefer the χ-squared bias function.

Balance Principle versus χ-Squared

The 1960 Bailey and Simon paper prefers the χ-squared bias function to the balance principle, whereas the 1963 Bailey paper argues for the balance principle. In defense of the χ-squared bias function, the 1960 Bailey and Simon paper says (p. 10):

> …the indication of each group should be given a weight inversely proportional to the standard deviation of the indication.

This is a traditional justification for classical credibility, as Bailey and Simon continue:

> The standard deviation of the indication is inversely proportional to the square root of the expected number of losses for the group.[27]

[27]Bailey and Simon [1960] assume that if all claims are independent, the variance is proportional to the number of claims, so the standard deviation is proportional to the square root of the number of claims (see also Longley-Cook [1962]). After the writings of Hans Bühlmann, Gary Venter, Howard Mahler, and others, this assumption is no longer the standard rationale for credibility.

The 1963 Bailey paper prefers the balance principle because it is unbiased, whereas the χ-squared bias function may be biased. By "unbiased," Bailey means that the balance principle constrains the relativities so that the total indicated pure premiums along any dimension equal the total observed loss costs along that dimension.

The balance principle uses the first-order departure, which is generally preferred by firms seeking to maximize profits. This is perhaps the strongest argument for the balance principle.

Common practice among casualty actuaries is to use the balance principle. One might presume that since more effective procedures drive out less effective procedures in a competitive market, the balance principle is perhaps the most effective bias function.

In truth, many ratemaking procedures are selected for ease of implementation, not necessarily for accuracy. The balance principle was easier to implement before the widespread use of desktop computers. Few actuaries have tried the χ-squared bias function or the least squares bias function. No conclusions should be drawn from the current practice.

10. CREDIBILITY

Many practitioners combine the minimum bias procedure with credibility weighting of the indicated pure premiums, either with the observed loss costs or with the underlying pure premiums. We show illustrations of each method.

The minimum bias procedure gives the indicated pure premiums for each class in an array. One may choose the pure premiums used for the final rates as a weighted average of the indicated pure premiums and the observed loss costs for that class. The credibility for the observed loss costs is a function of the volume of business in the class. Classes with greater volume place more weight on the observed loss costs; see Venter [1992].

Classical credibility formulas are the most commonly used. Classes with a certain volume of claims or of exposures are given full credibility. The square root rule is used for classes with lower volume of claims or exposures.

Illustration 28: Suppose that classes with exposure of 10,000 or more car-years are accorded full credibility. A class with 3,600 car-years of exposure has an \$800 observed loss cost. The minimum bias indicated pure premium for this class is \$700. The credibility assigned to the class is $(3{,}600/10{,}000)^{0.5} = 60\%$ credibility. The credibility weighted pure premium is

$$60\% \times \$800 + (1 - 60\%) \times \$700 = \$760.$$

Illustration 29: For premises and operations ratemaking, Insurance Services Office (ISO) uses a balance principle minimum bias procedure with observed loss ratios to determine the indicated changes to class group and type of policy relativities.[28]

- An indicated relativity change of 1.08 for type of policy 12 means that the existing relativity for type of policy 12 should be increased by 8%.
- The full credibility standard is based on the number of claims in the class during the experience period. These standards are 2,500 claims for OL&T BI, 3,000 claims for M&C BI, and 7,500 claims for M&C PD.
- Partial credibility is based on the square root rule. For example, 1,080 claims in M&C BI gives $(1{,}080/3{,}000)^{0.5} = 60\%$ credibility.
- The indicated relativity change for the class is raised to the power of the credibility. If the indicated relativity change is 1.08 and the credibility is 60%, the credibility weighted relativity change is $1.080^{0.6} = 1.047$.

[28]Type of policy refers to monoline versus multiline policies (and type of multiline policy). See Graves and Castillo [1990] for a more complete discussion of the ISO procedure.

These two illustration show different uses of credibility. ISO credibility weights the indicated classification relativities with the current classification relativities to dampen the changes from year to year. The observed loss costs in the first illustration are credibility weighted with the indicated pure premiums to increase the accuracy of the final pure premiums.[29]

Embedded Credibility

The minimum bias procedure has credibility embedded in the calculations, since each cell is weighted by the number of exposures in that cell.

A comparison with the single-dimensional classification ratemaking procedure should clarify this. Suppose there are three territories in a state with the experience shown below. The exposures are car-years, and the dollar figures are in thousands.

	Exposures	Claims	Premium	Losses	Loss Ratio	Indication
Terr 01	5,000	500	5,000	3,500	70.0%	0.972
Terr 02	10,000	1,000	15,000	10,800	72.0%	1.000
Terr 03	2,000	200	4,000	2,980	74.5%	1.035
Total	17,000	1,700	24,000	17,280	72.0%	

The unadjusted observed data suggest that

- Territory 01 should have a reduction of 2.8% in its base rate.
- Territory 02 should have no change in its base rate.
- Territory 03 should have an increase of 3.5% in its base rate.

The indications in the table do not consider the number of exposures or claims in each territory. Since territory 03 has only 200 claims in the experience period, the +3.5% indication may

[29]See Venter's [1992] distinction between classical credibility used to minimize rate fluctuations from year to year and Bayesian-Bühlmann credibility used to increase the accuracy of the estimate.

be distorted by random loss fluctuations. To adjust for the volume of business in each territory, the raw indications may be credibility weighted with the overall average of unity, where the credibility depends on the number of exposures or claims.

In the minimum bias procedure, the number of exposures in each cell affects the computation. The weight accorded to the observed loss costs in the cell is proportional to the number of exposures in the cell. From this perspective, credibility weighting the observed loss costs by the number of exposures would be applying credibility twice.

Nevertheless, some justification remains for a credibility adjustment. To determine the indicated pure premium for a cell, the minimum bias procedure uses the type of rating model along with all the cells in the array. The credibility embedded in the minimum bias procedure deals with random loss fluctuations. A second credibility adjustment deals with model specification risk. We explain these concepts with an illustration.

Illustration 30: The observed loss cost for young unmarried urban male drivers is $2,500 per car. After applying a minimum bias procedure, the indicated pure premium for these drivers is $3,000 per car. There are two explanations for the difference.

1. Random loss fluctuations account for the difference. The credibility embedded in the minimum bias procedure is sufficient. No additional credibility adjustment should be used.
2. The rating model is not correct. For example, the minimum bias procedure may be using a multiplicative model when an additive model is proper. This is model specification risk, and a second credibility adjustment is warranted.

Classical credibility procedures are not an ideal compensation for model specification risk. The ideal approach is to use several models, such as multiplicative, additive, and combined models, and to test the goodness-of-fit for each model. Time constraints preclude this ideal approach in many cases, and a credibility adjustment may be a practical alternative.

Rate Fluctuations

When rating bureaus made advisory rates, they had more incentive to temper rate fluctuations from year to year than private insurers have. ISO's credibility procedure may not have firm statistical justification, but it fulfills the objective of tempering the requested rate changes.

The use of credibility to temper rate fluctuations from year to year is a dubious practice. In practice, most actuaries conceive of credibility as a means to price more accurately. Although Venter correctly notes that the stated rationale for classical credibility deals with tempering rate fluctuations, even classical credibility does serve the objective of increasing the accuracy of the rate indications.[30]

11. SUMMARY

For each model discussed in this paper, there are simple iterative functions. The task of the pricing actuary is to determine a rating function—such as multiplicative, additive, or combined—and a bias function (balance principle, least squares, χ-squared, or maximum likelihood). If the maximum likelihood bias function is used, the actuary must also select a probability distribution function for the loss costs (or other values) in each cell.

The type of data in each cell will generally be either loss costs or loss ratios. If the pricing actuary is using all the dimensions of the classification system in the minimum bias analysis, it is easier to use loss costs. If there are significant classification dimensions that are not included, and if there may be an uneven

[30]See Venter's chapter on "Credibility" in any of the first three editions of the CAS textbook, *Foundations of Casualty Actuarial Science*, and Mahler [1986]. As Mahler points out, tempering rate changes and aiming for rate accuracy are different purposes, but they usually have a similar result.

distribution of exposures along these other classification dimensions, the pricing actuary may prefer to use loss ratios.

We list here the models that have been proposed for insurance use, along with their recursive equations.

Multiplicative model, balance principle:

$$x_i = \frac{\sum_j n_{ij} r_{ij}}{\sum_j n_{ij} y_j}.$$

Additive model, balance principle:

$$x_i = \frac{\sum_j n_{ij}(r_{ij} - y_j)}{\sum_j n_{ij}}.$$

Multiplicative model, least squares:

$$x_i = \sum(n_{ij} \times r_{ij} \times y_j) \Big/ \sum(n_{ij} \times y_j^2).$$

Additive model, least squares:

$$x_i = \frac{\sum n_{ij} \times (r_{ij} - y_j)}{\sum n_{ij}} - B.$$

Multiplicative model, χ-squared:

$$x_i = \left[\sum(n_{ij} \times r_{ij}^2 / y_j) \Big/ \sum n_{ij} y_j\right]^{0.5}.$$

Additive model, χ-squared:

$$\Delta x_i = \frac{\sum_j n_{ij}\left(\frac{r_{ij}}{x_i + y_j}\right)^2 - \sum_j n_{ij}}{2\sum_j n_{ij}\left(\frac{r_{ij}}{x_i + y_j}\right)^2\left(\frac{1}{x_i + y_j}\right)}.$$

Multiplicative model, maximum likelihood, normal density function:

$$x_i = \frac{\sum_j n_{ij} r_{ij} y_j}{\sum_j n_{ij} y_j^2}.$$

Additive model, maximum likelihood, normal density function:

$$x_i = \frac{\sum_j n_{ij}(r_{ij} - y_j)}{\sum_j n_{ij}}.$$

Multiplicative model, maximum likelihood, exponential density function:

$$x_i = \frac{\sum_j \dfrac{r_{ij}}{y_j}}{k},$$

where k is the number of classes in the j dimension.

The recursive functions for a multiplicative model, maximum likelihood, Poisson distribution function are the same as those for the multiplicative model, balance principle.

Derivations of the formulas for the maximum likelihood models may be found in Brown [1988].

Accurate classification systems are the bedrock of insurance pricing. Accurate and unbiased rating systems enable insurers to attain competitive advantages over their peer companies. Inaccurate rating systems lead to unsatisfactory profits and to loss of market share.

As competition increases in the insurance industry, and as companies are forced to rely on their own pricing instead of bureau rates, the need for more accurate ratemaking increases. The minimum bias procedure can be used to optimize a variety of rating models.

REFERENCES

Bailey, Robert A., "Insurance Rates with Minimum Bias," *PCAS* L, 1963, pp. 4–11; discussion by James R. Berquist, pp. 11–13; Stephen S. Makgill, pp. 13–14.

Bailey, Robert A. and LeRoy J. Simon, "Two Studies in Automobile Insurance Ratemaking," *PCAS* XLVII, 1960, pp. 1–19; discussions by L. H. Roberts, pp. 200–203; D. B. Martin, pp. 203–206.

Brown, Robert L., "Minimum Bias with Generalized Linear Models," *PCAS* LXXV, 1988, pp. 187–218; discussion by Gary G. Venter, *PCAS* LXXVII, 1990, pp. 337–349.

Conners, John B. and Sholom Feldblum, "Personal Automobile Insurance: Cost Drivers, Pricing, and Public Policy," *PCAS* LXXXV, 1998, pp. 370–403.

Cook, Charles F., "The Minimum Absolute Deviation Trend Line," *PCAS* LIV, 1967, pp. 200–204.

Cummins, J. David, Barry D. Smith, R. Neil Vance and Jack L. VanDerhei (eds.), *Risk Classification in Life Insurance*, Boston: Kluwer Academic Publishing, 1983.

Feldblum, Sholom, "Personal Automobile Premiums: An Asset Share Pricing Approach for Property/Casualty Insurance," *PCAS* LXXXIII, 1996, pp. 190–296.

Feldblum, Sholom, Duncan Anderson, Ernesto Schirmacher, Doris Schirmacher, and Neeza Thandi, "Generalized Linear Models: A Practitioner's Guide," forthcoming.

Ferreira, Joseph Jr., "Identifying Equitable Insurance Premiums for Risk Classes: An Alternative to the Classical Approach," in Andrew F. Giffin, Vincent Travis, and William Owen (eds.), *Automobile Insurance Risk Classification: Equity and Accuracy* Boston: Massachusetts Division of Insurance, 1978, pp. 74–120.

Finger, Robert J., "Estimating Pure Premiums by Layer—An Approach," *PCAS* LXIII, 1976, pp. 34–52; discussion by Lee R. Steeneck, pp. 53–55.

Graves, Nancy A. and Richard J. Castillo, "Commercial General Liability Ratemaking for Premises and Operations," *Pricing*, Casualty Actuarial Society Discussion Paper Program, 1990, Volume II, pp. 631–696.

Longley-Cook, Laurence H., "An Introduction to Credibility Theory," *PCAS* XLIX, 1962, pp. 194–221.

Mahler, Howard C., "An Actuarial Note on Credibility Parameters," *PCAS* LXXIII, 1986, pp. 1–26.

Venter, Gary G., "Credibility," Chap. 7 in *Foundations of Casualty Actuarial Science*, Second Edition, Casualty Actuarial Society, 1992, pp. 375–483.