Something Old, Something New in Classification Ratemaking With a Novel Use of GLMs for Credit Insurance

Keith D. Holler, FCAS, MAAA, ASA, ARM,
David Sommer, FCAS, MAAA, and
Geoff Trahair, BEc(Hons), FIAA, FIA

Something Old, Something New in Classification Ratemaking With a Novel Use of GLMs for Credit Insurance

Abstract:

This paper discusses some methods that can be used to calculate classification relativities and reduce the error that would otherwise occur by using one-way analysis. Section 2 will discuss the problem of risk classification analysis from a mathematical and statistical viewpoint and show some of the implied solutions from these approaches. This exposition revisits the work pioneered in the USA by Bailey, Bailey and Simon, and Brown, which are the foundations of American casualty practice in the area of classification ratemaking. We will then revisit another technique based on Generalized Linear Modeling (GLM) in Section 3 and discuss the advantages of implementing this technique. For those who have a strong background in classification ratemaking and GLM, we recommend skipping to Sections 4 and 5, where we present an application of this technique to credit insurance and discuss the results.

By: Keith D. Holler, FCAS, MAAA, ASA, ARM

David Sommer, FCAS, MAAA, *and*

Geoff Trahair, BEc(Hons), FIAA, FIA

TABLE OF CONTENTS

33

TABLE OF CONTENTS

Section 1. Introduction

1.1 Description of the General Problem

A premium rating plan has two goals. First, it should ensure that the insurer receives premiums at a level which is expected to be adequate to cover losses and expenses, while providing a fair rate of return. Second, it should allocate those premiums fairly between insureds, where "fairly" means that higher premiums are paid by those insureds with greater risk of loss and vice-versa, while all insureds contribute consistently to profit and expense. While we recognize that there may be considerations in which an insurer chooses not to price a risk with respect to these goals (regulatory, competitive, etc.), we will assume, for the purposes of this paper, that these other considerations are addressed subsequent to determining the expected value premiums.

To meet these goals, most ratemaking consists of two aspects. The first is the determination of the overall rate level. This addresses the first goal mentioned above. The second aspect of ratemaking is the risk classification analysis. It is through the risk classification plan and its rate relativities that the second goal of equity is installed in the pricing process.

In determining classification relativities, it appears simple enough to analyze loss costs (loss per exposure) by variable to calculate the necessary factors. If married drivers have half of the loss cost of unmarried drivers, they should receive a relativity of 0.5 and so on. This single-variable analysis, however, makes an assumption that is generally not true - that the effects of a single variable are independent of all other rating variables. We introduce the

35

following example[1] which appears in the SAS/STAT manual [1] to show some of the difficulty with this assumption.

1.2 A Simple Example

Consider claim count data which are modeled using two classification variables: age group, with two levels; and car type, with three levels. The claim counts and exposures for each of the classes are as follows:

Claims

Age Group	Car Size		
	Large	Medium	Small
1	1	37	42
2	14	73	101

Exposures

Age Group	Car Size		
	Large	Medium	Small
1	100	1200	500
2	300	500	400

Actual Frequency

Age Group	Car Size		
	Large	Medium	Small
1	.010	.031	.084
2	.047	.146	.253

Frequency Relativities

Age Group	Car Size		
	Large	Medium	Small
1	1.0	3.1	8.4
2	4.7	14.6	25.3

The actual frequency for a class is computed as the number of claims divided by the number of exposures for that class. Each class is a combination of values for each classification variable (e.g. - age group 1 with a large car). The observed relativities in this example are

---

[1] Reprinted with permission: SAS Institute Inc., SAS® Technical Report P-243, SAS/STAT® Software: The GENMOD Procedure, Release 6.09, Cary, NC: 1993, Copyright® SAS Insitute Inc. 88 pp.

computed using claim frequency. (This approach assumes that the average claim size is the same for each class.) In addition, the large car size/age group one (L1) class is assumed to be the 'base class', which has a relativity of 1.0. The observed relativity of 25.3 for the small car size/age group two (S2) class means that for each S2 car, we observed 25.3 times as many claims on average than for each L1 car. If the base rate (i.e. the premium rate for a single L1 car) is $100, the premium charged for a single S2 car would be $2,530 or 25.3 x $100.

One Way Method

Class	Claims	Exposures	Frequency	Relativity
Large car size	15	400	.038	1.000
Medium car size	110	1700	.065	1.725
Small car size	143	900	.159	4.237
Age Group 1	80	1800	.044	1.000
Age Group 2	188	1200	.157	3.525

The one-way method computes a relativity separately for each value of the car size variable and the age group variable. For example, based on this method, the relativity for a medium sized car is .065/.038, or 1.725, where .038 is the total frequency for the base car size, large. Note that all of the data is used to determine the car size relativities and then used again to determine the age group relativities.

The final overall rating class (car size/age group) relativity is then the product of the individual car size relativity and the individual age group relativity. For example, the S2 relativity based on the one-way method would be 4.237 x 3.525, or 14.936. The table below summarizes the relativities based on the one-way method.

37

Age Group		Car Size		
		Large	Medium	Small
Level	Relativity	1.000	1.725	4.237
1	1.000	1.000	1.725	4.237
2	3.525	3.525	6.082	14.936

We can see that this method fails to make the relativities as steep as necessary to reflect the combined increased risk from both variables. For example, the S2 car would be charged a premium of $1,493.60 instead of the $2,530 premium indicated by the data. Because this simple method uses the data to derive the relativity for each class variable independently of the other class variables, it produces results which are inconsistent with the data.

This effect is not due to a quirky example. There are very strong practical reasons that would lead us to reject one-way analysis. Normally, we would expect to see some degree of association between rating factors. An insurer's portfolio of risks is unlikely to be a random sample from the entire population of insurance risks - the insurer's pricing structure may target specific segments of the market and so we would expect to see this reflected in the relative loss-costs. We therefore prefer modeling techniques that can deal with these exposure-related issues directly.

This paper discusses some methods that can be used to calculate classification relativities and reduce the error that would otherwise occur by using one-way analysis. Section 2 will discuss the problem of risk classification analysis from a mathematical and statistical viewpoint and show some of the implied solutions from these approaches. This exposition revisits the work

pioneered in the USA by Bailey [2], Bailey and Simon [3], and Brown [4], which are the foundations of American casualty practice in the area of classification ratemaking. We will then introduce another technique based on Generalized Linear Modeling (GLM) in Section 3 and discuss the advantages of implementing this technique. For those who have a strong background in classification ratemaking and GLM, we recommend skipping to Sections 4 and 5, where we present an application of this technique to credit insurance and discuss the results.

Section 2. Mathematical Formulation of Solution

2.1 Class Plan Objective - Minimum Bias Approach

To better understand the techniques being introduced, it will be useful to discuss the objectives of classification ratemaking and frame them in a mathematical context.

The objective of a classification plan is to replicate the actual loss cost relativities as closely as possible. Let's call the selected relativities x_i, y_j for the ith, j^{th}, (etc.) values of the respective rating variables.[2] Let's call r_{ij} the actual loss cost relativity for the set of exposures that have both of these variable values (for example - youthful driver and large car). The goal is then, for all i, j, to have $x_i y_j$ be as close to r_{ij} as possible (if we are designing an additive class plan, replace $x_i y_j$ with $1 + x_i + y_j$), where "close" is measured by some bias function $f(r_{ij}, x_i, y_j)$.

2.2 Example - Least Squares

For example, suppose we define a bias function as the weighted squared error:

$SSE = \sum_i \sum_j n_{ij} (r_{ij} - x_i y_j)^2$ where n_{ij} is the number of exposures in the ij^{th} cell.[3]

---

[2] While we are dealing with two variables in this example, we can generalize to n variables. Similarly, we can generalize to allow for interactions. If we know that two variables interact (e.g., age and sex) then we can create a new composite variable formed for each combination of the categories of the original variables.

[3] n_{ij} is used as a weight to reflect the relative exposure amount of the ij^{th} cell.

40

In minimizing SSE, we set $\partial \text{SSE}/\partial x_k = 0$, and solve for x_k in terms of y_j.

$$-2\sum_j y_j n_{kj}\left(r_{kj} - x_k y_j\right) = 0$$

$$\sum_j y_j n_{kj} r_{kj} = x_k \sum_j n_{kj} y_j^2$$

$$x_i = \frac{\sum_j n_{ij} r_{ij} y_j}{\sum_j n_{ij} y_j^2} \quad \text{and similarly}^4 \quad y_j = \frac{\sum_i n_{ij} r_{ij} x_i}{\sum_i n_{ij} x_i^2}$$

We will call this the least squares multiplicative model. For this model, the solution of the partial derivative equations leads to forms which can be solved iteratively. This approach proceeds by selecting initial values for each y_j and then using the model solutions to solve for each of the x_i's. The x_i's are then substituted into the equations for the y's to produce the next estimate of the y_j's. The process is repeated until the solutions at each iteration converge.

The indicated class relatives for the auto example, using the least squares multiplicative model, are as follows:

Age Group		Car Size		
		Large	Medium	Small
Level	Relativity	1.000	3.021	5.533
1	1.000	1.000	3.021	5.533
2	3.541	3.541	10.697	19.592

---

[4] For the final x_i solution, the previous subscript of k is simply replaced with i to enable us to continue with the notation.

A detailed example of the iterative calculations is presented in Section 2.4 for the Poisson maximum likelihood multiplicative model

The loss cost relativity, r_{ij}, is the loss cost for the ij[th] class divided by the loss cost for the base class, or by the total loss cost if there is no base class. For purposes of this paper, we will assume that there is a base class, unless specifically noted. The loss cost relativity can also be derived as the frequency relativity multiplied by the severity relativity. If each class has the same average claim size, then the severity relativity is unity for every class. In this case, the loss cost relativity r_{ij} is equal to the frequency relativity. The example in Section 1 assumes the same average claim size by class.

Allowing the subscript B to represent the base class, we can formalize this discussion as:

$$r_{ij} = \frac{L_{ij}}{L_B} \frac{n_{ij}}{n_B} = \frac{s_{ij} m_{ij}}{s_B m_B} \frac{n_{ij}}{n_B}$$

Where L_{ij} is the total loss in the ij[th] class, s_{ij} is the average claim in the ij[th] class, and m_{ij} is the number of claims in the ij[th] class.

If the classes have the same average claim size, i.e. s_{ij} equals s for all $_{ij}$, then:

$$r_{ij} = \frac{sm_{ij} / n_{ij}}{sm_B / n_B} = \frac{f_{ij}}{f_B}$$

which equals the frequency relativity. Here, f_{ij} is the frequency of the ij[th] class.

2.3   Class Plan Objective - Maximum Likelihood Approach

An alternate approach centers on answering the question "which x_i, y_j's are those that maximize the likelihood of the actual r_{ij}'s being generated?" This approach attempts to obtain the objective of the class plan via a firmer statistical setting, rather than minimizing a general subjective bias function. There are, of course, several items that could be considered random variables. For example, the class losses, L_{ij}, class claim counts m_{ij}, class severity s_{ij}, and class loss cost relativity r_{ij} can each be viewed as having underlying statistical distributions in which the x_i's and y_j's are parameters. In fact, the random variables could be placed at an individual exposure level, rather than a cell level.

If the random variable is r_{ij} at the individual class level and is drawn from the probability distribution g, then the likelihood function L, which is the product of the probabilities of independent observations, is $L = \prod_{i,j} g\left(r_{ij}; x_i, y_j\right)$ with the parameters x_i and y_j.

We can maximize the likelihood function by maximizing its logarithm, so

$$\ell n(L) = \sum_i \sum_j n_{ij} \ell n\left[g\left(r_{ij}; x_i, y_i\right)\right]$$

which we maximize by calculating the partial derivatives and setting them equal to zero.

2.4   Example - Possion Frequency

Let's work through the maximum likelihood estimate for a multiplicative model,

$r_{ij} = x_i y_j$. For this model, we will assume that the random variable is the number of claims per class, m_{ij}, and that each class has the same severity. The Poisson density would be:

$$g\left(m_{ij}, x_i, y_j\right) = \exp\left(-h\left(x_i, y_j\right)\right) h\left(x_i, y_j\right)^{m_{ij}} / m_{ij}!$$

Here, $h(x,y_j)$ takes the role of the familiar lamda parameter. The parameter is a function of x_i and y_j. In the multiplicative model, $h(x_i,y_j) = x_i y_j f_B n_{ij}$ where f_B is the observed frequency of the base class. Because of the additive property of the Poisson distribution, this model will also result if the random variable is the number of claims per exposure and the lambda function equals $x_i y_j f_B$.

Either way, the likelihood function is:

$$L = \prod_{i,j} \frac{e^{-x_i y_j f_B n_{ij}} \left(x_i y_j f_B n_{ij}\right)^{m_{ij}}}{m_{ij}!}$$

$$\ell n(L) = \sum_{i,j}\left[-x_i y_j f_B n_{ij} + m_{ij}\, \ell n\left(x_i y_j f_B n_{ij}\right) - \ell n\left(m_{ij}!\right)\right]$$

$$\frac{\partial \ell n(L)}{\partial x_k} = \sum_j\left[-y_j f_B n_{kj} + m_{kj}/x_k\right] = 0$$

$$f_B \sum_j y_j n_{kj} = \frac{1}{x_k}\sum_j m_{kj}$$

(replacing k with i)

$$x_i = \frac{\sum_j m_{ij}}{f_B \sum_j y_j n_{ij}} = \frac{\sum_j \dfrac{m_{ij}}{f_{ij}}\dfrac{f_{ij}}{f_B}}{\sum_j y_j n_{ij}}$$

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

which we will call the Poisson (multiplicative) model.

44

Let's illustrate the use of the Poisson model by applying it to the previously introduced example. x_i will be car size, y_j will be age group, n_{ij} will be the number of exposures, and r_{ij} will be the actual claim frequency relativity. The first iteration of calculation[5] would be:

x_1 $= (n_{11} r_{11} + n_{12} r_{12})/(n_{11} y_1 + n_{12} y_2)$ Assume $y_1 = 1$, $y_2 = 4$, initially.

$= (100 *1.0 + 300 * 4.7)/(100 * 1 + 300 * 4)$

$= 1500 / 1300$

$= 1.15$

x_2 $= (n_{21} r_{21} + n_{22} r_{22})/(n_{21} y_1 + n_{22} y_2)$

$= (1200 * 3.1 + 500 * 14.6)/(1300 * 1 + 500 * 4)$

$= 11000 / 3200$

$= 3.44$

x_3 $= (n_{31} r_{31} + n_{32} r_{32})/(n_{31} y_1 + n_{32} y_2)$

$= (500 * 8.4 + 400 * 25.3)/(500 * 1 + 400 * 4)$

$= 14300 / 2100$

$= 6.81$

y_1 $= (n_{11} r_{11} + n_{21} r_{21} + n_{31} r_{31})/(n_{11} x_1 + n_{21} x_2 + n_{31} x_3)$

$= (100*1.0 + 1200*3.1 + 500*8.4)/(100*1.15 + 1200*3.44 + 500*6.81)$

$= 8000 / 7645$

$= 1.05$

y_2 $= (n_{12} r_{12} + n_{22} r_{22} + n_{32} r_{32})/(n_{12} x_1 + n_{22} x_2 + n_{32} x_3)$

$= (300*4.7 + 500*14.6 + 400*25.3)/(300*1.15 + 500*3.44 + 400*6.81)$

$= 18800 / 4789$

$= 3.93$

---

[5]   While the steps are displayed with r_{ij}, x_i, and y_j rounded, the exact figures are used in each step of the calculations

45

After the first iteration, we would use the new y_j's to recalculate the x_i's and so on, until the results converged. The subsequent iterations are shown in the table below:

Parameter	First Iteration	Second Iteration	Converged Solution	Rebased Relativities
x_1 - Large Car Size	1.15	1.17	1.1703	1.000
x_2 - Medium Car Size	3.44	3.42	3.4169	2.920
x_3 - Small Car Size	6.81	6.83	6.8312	5.837
y_1 - Age Group 1	1.05	1.05	1.0481	1.000
y_2 - Age Group 2	3.93	3.92	3.9232	3.743

The rebased relativity for a specific class level is the converged solution divided by the base class level converged solution. For example the 2.920 relativity for the medium car size equals 3.417/1.170.

The resulting implied class relativities are as follows:

Age Group		Car Size		
		Large	Medium	Small
Level	Relativity	1.000	2.920	5.837
1	1.000	1.000	2.920	5.837
2	3.743	3.743	10.929	21.850

which is a significant improvement over the one-way relativity calculations. The improvement lies in the fact that the fitted class relativities for the Poisson model more "closely match" the relativities, r_{ij}, in the data.

46

2.5 Loss Ratio Relativities

Before proceeding, it is worth digressing to discuss the meaning of "actual losses." In standard ratemaking procedures, it is common to use loss ratios, rather than pure premiums, in a relativity analysis. However, loss ratios only give the required *change* in relativity, as the *existing relativities are embedded in the denominator. Therefore, one must adjust the loss ratios* to remove the effect of the existing relativities of any rating variables being analyzed in the study. This adjustment can be handled via the following steps:

1. Calculate a matrix of existing differentials, D_{ij}, where for a multiplicative model D_{ij} is the product of the current rate relativities for row i and column j. In the additive model, D_{ij} = 1 + the sum of the current rate relativities for row i and column j. The base class should have D_{ij} equal to 1.

2. Calculate the matrix of loss ratios, LR_{ij}.

3. Divide all of the loss ratios by the loss ratio for the base class. This will give "raw loss ratio relativities," W_{ij}.

4. Multiply each of the W_{ij}'s by D_{ij} to get the adjusted loss cost relativities, r_{ij}. This adjustment avoids double-correcting for the variables in the model.

Bailey [2], Bailey and Simon [3], and Brown [4] introduce a number of other models. In the Appendix to this paper, we will derive some of these additional models as well as show the solution to the above example (but not the calculations) for each of these models. While this set of models is not exhaustive, it gives the reader an indication of how to construct maximum likelihood estimates given an underlying distributional assumption, as well as other types of

47

constraints. Finally, it should be kept in mind that by using alternative notations, a single model may often be written in several different forms and may arise through the optimization of different criteria.

Section 3.   Introduction to GLMs

3.1   Introduction

This section provides a brief introduction to Generalized Linear Models (GLMs). Those who

are familiar with this theory may wish to skip ahead to Section 4, which contains an

application of GLMs for classification data. Several good introductory texts include those

by Aitkin [5] et al and the SAS[x] Institute Inc.[1] The standard, complete reference is by

McCullagh and Nelder [6].

3.2   Traditional Linear Models

Traditional linear models include the familiar simple and multiple regressions and Analysis of

Variance (ANOVA) models, among others   GLMs include all of these linear models and

extend well beyond the traditional frameworks by broadening most of the major assumptions.

This implies that the use of multiple regression for classification ratemaking is a specific, albeit

simpler, application of GLM.

Before proceeding to the general GLM framework, we will briefly recap the traditional

linear model in matrix form:

$\bar{y} = X\bar{\beta} + \bar{\varepsilon}$ where

$\bar{y}$ is the nx1 vector of actual observed values;

X is the nxp matrix of explanatory variables;

$\bar{\beta}$ is the px1 vector of unknown parameters; and

$\bar{\varepsilon}$ representing the 'error' term, is the nx1 vector of independent, identically

distributed (iid) normal random variables, with common variance, σ^2.

49

Note that a single observation, y_i, is modeled as $y_i = \bar{x}_i^T \bar{\beta} + \varepsilon_i$, where x_i is the ith row of the matrix X and $^{\cdot}$ is the matrix transpose operator. In the classification setting, the parameter vector, $\bar{\beta}$, contains parameters for <u>all</u> of the classification variables. The ith row of the matrix X would represent the actual risk characteristics of the ith insured .

Analysis generally proceeds by estimating $\bar{\beta}$ via least squares, which is equivalent to maximum likelihood estimation for these models. Confidence intervals, point estimates, and hypothesis tests can all be conducted using the estimated parameters, $\hat{\beta}$.

The assumptions are reviewed by analyzing the residuals, e_i, where

$$e_i = y_i - \hat{y}_i \, and \, \hat{y}_i = \bar{x}_i^{\cdot} \hat{\beta}.$$

A very thorough reference for the theory underlying linear models is by Searle [7]. Residual diagnostics is covered in Belsley et. al. [8].

Shortcomings of Traditional Linear Modeling

As GLM's encompass traditional linear models, GLM theory, model structure, and model diagnostics all have their impetus in the traditional models. One can view GLM theory positively as an extension of traditional linear model theory in which the traditional model assumptions are relaxed to include more real-life problems. Specifically, situations that GLM's can handle but traditional models cannot, without resulting to painful transformations, are:

50

1. Non-normal response variables (y) - for example, there is no reason to believe that claim count data which is discrete and non-negative can be modeled appropriately by a continuous distribution which includes negative values in its range.

2. Non-linear Structure. The traditional model is $\hat{y}_i = E\left[y_i | \bar{x}_i\right] = \bar{x}_i^T \hat{\beta}$ which is linear in $\hat{\beta}$. Note that this implies that there exists some $\bar{x}_i$ for which $\hat{y}_i$ is negative. If, again, the data is count data or loss data, the mean will usually not be negative.

3. Non-constant variance. Traditional linear models assume that the variance is the same for each class. However, the variance often fluctuates with the overall magnitude of the class mean. For example, in the Poisson case, the mean equals the variance. There is nothing constant about it.

3.4 GLM

The general discussion in this section will use the traditional notation of y for the response variable and x for the covariate vector. The x_i and y_j from Section 2 will appear in this section as well. However, in the latter occurrences, Section 2 will usually be referenced and hopefully the context of the discussion will remove any confusion as to which x and y are being referenced.

51

GLM theory is built for probability distributions from the exponential families of the form:

$$f(y) = \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi) \right)$$

where θ are the underlying parameters, whose value may vary by class, and ϕ represents a scale parameter.

Exponential families include the normal, Poisson, gamma, and binomial distributions. The mean and variance of the exponential family are:

$$E[Y] = b'(\theta), \text{ which we denote } \mu,$$
$$Var[Y] = b''(\theta)\phi/W = V(\mu)\phi/W$$

where ' and '' denote first and second derivatives with respect to θ, $V(\mu)$ is a one-to-one variance function relating the mean and the variance, and W is the weight assigned to each observation. The weight is embedded in $a(\phi)$ and $c(y,\phi)$.

Two additional items that tend to arise are the link[6] and offset functions. The link function is a one-to-one function of the mean, g, such that $g(\mu)$ is modeled as $\vec{x}'\vec{\beta}$. Hence, a function of the mean, and not the mean itself, is modeled in a linear fashion. The offset function is generally used with the Poisson distribution to account for the level of exposure in each class.

---

[6] In this paper, we will use only canonical link functions. Canonical links result in the linear predictor, $\vec{x}'\vec{\beta}$, equaling the natural exponential parameter θ.

For example, in the Poisson distribution: $f(y) = e^{-\lambda}\lambda^y/y! = \exp(y\log\lambda - \lambda)/y!$

and using the log link, $\quad \theta = Log\mu = Log\lambda = \bar{x}'\hat{\beta}$

$$f(y) = \exp(y\theta - e^{\theta})/y!$$
$$b(\theta) = e^{\theta}$$
$$b'(\theta) = e^{\theta} = \lambda = \mu = E[Y]$$
$$b''(\theta) = e^{\theta} = \lambda = \mu = Var[Y] = V(\mu)$$
$$\phi = 1 \ and \ w = 1.$$

The fitted parameters, $\hat{\beta}$, are obtained as in the traditional models, via maximum likelihood estimation. However, a closed form solution for the estimates does not usually exist, so an iterative process is used to obtain the estimates.

Typically, for count data, for each class, the exposure, n_i, and number of claims y, might be available. The Poisson model would become: $f(y_i) = e^{-\lambda_i n_i}(\lambda_i n_i)^{y_i}/y_i!$

The log-likelihood contribution of y_i is: $-\lambda_i n_i + y_i \log\lambda_i + y_i \log n_i$.

Further, $E[Y_i] = \lambda_i n_i$, which on the log scale becomes:

$$\log E[Y_i] = \log\mu_i = \log\lambda_i + \log n_i = \bar{x}_i\hat{\beta} + \log n_i.$$

The exposure, n_i, is usually handled via an offset. For the Poisson model, the offset is *log n_i* Once the parameters are fit, the estimated means are obtained as $\mu_i = g^{-1}(\theta_i)$. In a Poisson model with two variables and an intercept, $\quad_{ij} = \exp(\bar{x}'\bar{\beta}) = \exp(\text{Intercept} + \alpha_i + \delta_j)$

The x_i and y_j for a multiplicative Poisson model presented in Section Two could then be obtained as $x_i = e^{\alpha_i}/e^{\alpha_s}$ and $y_j = e^{\delta_j}/e^{\delta_s}$ The estimated mean can be thought of as the predicted or fitted value

3.5   The Poisson Example Revisited

We now show how our previously introduced example would be handled with this method.

The following SAS code generates the data set to be used for the analysis:

```
DATA insure;
INPUT n m car $ age;
lnoffset = LOG(n);
*exposure          counts      car size        age group;
*n        m        car         age;
CARDS;
500      42       small        1
1200     37       medium       1
100      1        large        1
400      101      small        2
500      73       medium       2
300      14       large        2
;
RUN;
```

So, for example, there are 500 small cars (exposures) in age group 1 and this class had 42 claims. The model could be written as $\log \lambda_{ij} = Intercept + \alpha_i + \delta_j$. where α_i is the fitted parameter for car size i and δ_j is the fitted parameter for age group j.

54

To fit the Poisson regression model in SAS, we use the GENMOD procedure in the SAS/STAT module. The SAS code for this analysis is:

```
PROC GENMOD DATA = insure;
        CLASS car age;
        MODEL m = car age /
                DIST = Poisson
                LINK = log
                OFFSET = lnoffset;
        RUN;
```

The parameter estimates along with their standard errors are displayed below:

Parameter	Estimate	Standard Error
Intercept	-1.3168	0.0903
Large car size	-1.7643	0.2724
Medium car size	-0.6928	0.1282
Small car size	0.0000	0.0000
Age Group 1	-1.3199	0.1359
Age Group 2	0.0000	0.0000

Like linear regression, the model can be fitted either with or without an intercept. The above model has assumed that the small car size for age group 2 is the base class. The base class will have log λ_{32} equal to the intercept. By taking the inverse link function, $g^{-1} = exp$, a table of fitted expected frequencies can be constructed:

55

$\lambda_{ij} = \exp\left(Intercept + \alpha_i + \delta_j\right)$		
	Car Size		
Age Group	**Large**	**Medium**	**Small**
1	.0123	.0358	.0716
2	.0459	.1340	.2680

Now if we wanted to use large cars and age group 1 as our base class for a rating plan, and if the severity for each class was the same, the class relativities could be obtained by dividing the previous table through by λ_{11}. On the other hand, multiplicative class factors could be obtained for each level within the variable as exp(level parameter - base level).

For example, if large cars for age group 1 are the base class, the medium car class relativity could be computed as exp(1.7643 - .6928) = 2.920 The resulting class relativities are displayed below:

Age Group		**Car Size**		
		Large	**Medium**	**Small**
Level	**Relativity**	1.000	2.920	5.837
1	1.000	1.000	2.920	5.837
2	3.743	3.743	10.929	21.850

More detailed examples of using GLM's in auto classification rate making in the United Kingdom are described in Reference 9

3.6 Model Validation

We will discuss two types of goodness of fit or validation tests. First, we will introduce some more technical tests which do not usually get mentioned with traditional models. Then we will discuss some analogs of the more traditional residual plots and other less objective tests.

The "more technical" tests center on two statistics, which often have asymptotic chi-square distributions. The first statistic centers on an item known as the deviance. For a fixed ϕ, the scaled deviance is defined as:

$$D^*(\bar{y}, \bar{\mu}) = 2(\ell nL(\bar{y}, \bar{y}) - \ell nL(\bar{y}, \bar{\mu}))$$

where $\log L$ is the log likelihood. This looks very much like the log of the likelihood ratio test statistic.

For the Poisson distribution (with weight one, as shown previously):

$$\ell nL(\bar{y}, \bar{\mu}) = -\sum \mu_i + \sum y_i \ell n \mu_i - \sum \ell n(y_i!)$$
$$\ell nL(\bar{y}, \bar{y}) = -\sum y_i + \sum y_i \ell n y_i - \sum \ell n(y_i!), \text{ and}$$
$$D^*(\bar{y}, \bar{\mu}) = 2\sum[y_i \ell n(y_i/\mu_i) + (\mu_i - y_i)]$$

The second statistic used is Pearson's chi-square statistic $Q = \sum w_i (y_i - \mu_i)^2 / V(\mu_i)$

This statistic should also have a somewhat familiar look. In fact, $y_i - \mu_i$ is the residual, or actual less expected amount. The scaled Pearson chi-square statistic is Q/ϕ, which is Q in this Poisson case. For the Poisson case, Q becomes the very familiar

$$\sum \frac{(Observed - Expected)^2}{Expected}$$

This form can be used to evaluate the types of models presented in Section 2.

Both of the scaled statistics have an asymptotic chi-square distribution under various general conditions. The degrees of freedom is equal to the number of observations less the number of estimated parameters.

The deviance lends itself readily to testing heirachical or nested model structures. For two given models, M1 and M2, where M2 contains all the predictors in M1 as well as some additional ones, then the difference of the deviances for Model 1 and Model 2 is equal to twice the difference in the log likelihoods under each model. Thus the deviance can be compared to the chi-square distribution to test the significance of adding the new variables, as noted in Hogg & Klugman [10]. The degrees of freedom for the statistic is equal to the number of new variables added.

As with traditional models, one may examine residual plots in an attempt to validate the model. Three simple types of plots may be used – quantile plots, burst plots, and predictor plots.

Quantile plots are used to check the underlying distributional assumption. The traditional analog is the normality plot of the residuals. As with the traditional plot, a theoretical quantile - observed quantile (Q-Q) plot that is linear supports the distributional assumption made.

Burst plots are used to test the randomness of the residuals. As in traditional models, the residuals are plotted against the fitted points. If the plot appears to be a random burst with no discernable pattern, then the model structure is supported.

Predictor plots are used to ensure that the variables used in the model have been properly reflected. In these plots, the residuals are plotted against each of the variables. A good model will not display any patterns in these plots. The presence of a pattern usually indicates some sort of bias in the fit and may point to a more complex breakdown of model assumptions. For example, in fitting models to claim severities, a common problem is increasing variability with increasing severity and would be reflected in these plots. This problem often leads to a situation of systematic under-prediction and over-prediction and can go unnoticed without these diagnostic procedures.

If there are points that prevent the plots from conforming to the above requirements (outliers), then corrective action is necessary. The most common course is to look at the specific data points concerned, exclude them from the data set (if they are relatively few in number), and refit the model. If there is a significant number of outliers, then this indicates a more serious problem, such as the one discussed above, and may indicate the need for reconsidering basic model assumptions.

It is very important to be aware of the model structure when reviewing the residual plots. Discrete data, or other model forms, may induce residual behavior which does not conform with the traditional expectations, but which is still acceptable. For example, consider a Bernoulli (binomial) model in which each observation is a claim (1), or not a claim (0), for each given exposure. Then the raw residual will either be 1-p or -p. This separation, which is not encountered in the traditional continuous normal models, leads to different expectations of what an acceptable burst or predictor plot would look like. For discrete data, it is often more useful to examine the ratios of fitted versus actual data, as we discuss next.

A practical model validation procedure is to examine tables of the ratios of fitted to actual (F:A) number of claims or total cost of claims. The aim of this analysis is to establish if there is any systematic bias in the model estimates. In general, for any subclass, we do not expect the F:A ratio to be 100%. It may be greater or lesser than 100% depending on what model constraints are in place. For example, claim severities below a set amount may be excluded for the reason of financial insignificance and hence the average claim cost will be higher. This would cause the fitted total claim cost to be higher and hence the F:A ratio to be greater than 100%. As long as the F:A ratio is reasonably consistent across all levels of the relativity factors, there is no cause for concern. However, if the F:A ratio declines as age of driver increases, for example, this would indicate a systematic bias in the model for age of driver.

Correcting systematic bias would require further investigation as to the source of the bias. It could be due to one or more variables being omitted from the final model or it may simply

be due to small amounts of exposure at young ages. Another possible cause of this bias may be a changing book of business over time.

One final principle of practical model validation is "the eyeball axiom". By graphing the indicated relativities for each variable, one can examine these estimates (and their confidence intervals) for reasonableness. These graphs can be telling in terms of data quality as well as implied relationships.

7   Why use GLM?

The astute reader may have noticed that the maximum likelihood example in Section 2 and the GLM example produce the same relativities. As the GLM estimates are also based on maximum likelihood, the solutions should be the same. This leads to the obvious question "Why bother with GLM if I can iterate?"

There are several reasons to implement a model using GLM. There are a number of statistical software packages available which handle GLM. GLM and these software packages have the following advantages:

1)   The software packages include a general fitting routine that is applicable to any GLM. Simple closed form iterative solutions may not be available for a specific GLM.

2)   Continuous rating variables, such as actual age, can be incorporated into a model.

3)   Most of the common model forms, such as poisson, binomial, normal, lognormal, and gamma, are already included as standard models. Non-standard exponential family

61

models can be included with a few lines of code. The package saves one the time of deriving, programming, and verifying iterative models

4) The process of exploring residual plots, goodness of fit statistics, variable groupings, and variable interactions is easier.

5) Most packages produce "standard errors" for each parameter. These can also be used to evaluate the model

6) Most of the packages are fairly efficient. For example, the model to be discussed in Section Four was fit to several hundred thousand records in a few minutes using SAS.

7) Finally, when viewed as an extension of traditional linear models, the whole GLM modeling process may seem more natural than an iterative formula, or at least less alien. This will certainly assist the actuary in relating the analysis to non-technical decision makers, who may be somewhat familiar with regression.

Section 4. Applications of GLMs

4.1 Introduction

GLM techniques are well established in rating for personal lines insurance in some areas of the world (auto and household). Typically, claim frequency and claim severity are modeled separately and the results combined to produce loss cost relativities. Claim frequency is often modeled using Poisson or negative binomial error structures, while claim severity is often modeled using gamma or log-normal error structures. Model structures are usually multiplicative, that is for a given cross-classification of risk-factors called the "base class," the product of the various loss cost relativities is unity. Relativities greater than one indicate increased risk while relativities less than one indicate reduced risk, relative to the base class. As mentioned in Section 1, a separate exercise is needed to establish the actual base premium for the base class.

The above description, although brief, summarizes the situation for many insurance applications. However there is ongoing debate on issues such as multiplicative versus additive model structures, whether frequency and severity should be modeled separately or jointly, the correct treatment of no-claim-bonus scales, etc. The interested reader should consult the literature for discussion of these and other issues [5], [9].

4.2   Credit Insurance

The example we present in this paper is based on analysis of a U.S. financial institution's claims experience. In particular, our aim in modeling terms is to improve their ability at the time of extending credit to correctly assess high- and low- risk applicants, using information collected at the time of loan application. By developing these models, the loan default performance of the outstanding balances should improve, increasing profitability.

A large amount of information is collected during the application process, including credit score, amount of the loan, type of collateral, income ratios, marital status, loan term, loan purpose, state, borrower age, gender, etc. Some of this information was not used because of insurance and lending nondiscrimination requirements.

Some of the information collected is naturally categorical in nature, such as type of collateral. Some of the information, like age of borrower, is naturally continuous. More generally, the categorical nature of many rating factors and the number of rating factors gives rise to the problem that there may be large number of cross-classified cells (classifications). However, the actual number of cells is usually much smaller and there is often a large number of cells with very small exposure.

4.3   GLM Model for Credit Insurance Claims

As a loan (single exposure) only has two possible outcomes - claim or no claim - we chose

to model claim frequency using a multiplicative model with a binomial error structure, using

a logit link function. This approach is identical to logistic regression. The regression model

equation is:

$$Logit(P_i) = Log\left(\frac{P_i}{1 - P_i}\right) = \vec{x}_i^T \vec{\beta}$$

where

P_i is the probability that the ith loan becomes a claim,

$\vec{x}_i^T$ is the vector of risk factors for the ith loan, and

$\vec{\beta}$ is the vector of risk factor relativities.

The model is fitted by maximum likelihood. For our work, we have used The SAS System,

in particular PROC GENMOD from the SAS/STAT module.

In the context of multiplicative relativities, the need for an interaction model means that

there are significant exposure-related differences for the particular factors in question.

This is analogous to the assumption about equal underlying exposure breaking down for

one-way analysis. In the GLM case, this can be corrected by fitting a model with terms

like x_j*x_k and excluding x_j and x_k . This is done even though testing for significance of

the interaction effect would include all of the terms.

Our model includes only main effects. We did not model any conditional relationships

between variables that would take into account interaction effects During the model

validation process, we did not see any sign of significant bias that suggested the need for these

interaction terms.

65

4.4   Model Validation

As discussed in Section 3, the usual statistical tool for model validation is residual analysis. This approach confirms that the underlying distributional assumptions have not been violated, as well as ensures that there is no systematic bias in the parameter estimates The first of these checks would often be conducted via two plots The first is a quantile plot of residuals versus quantiles from the assumed error distribution The second is a 'burst' plot of residuals versus actual values. Systematic bias would be explored with a series of plots of residuals versus the rating factors. Trends in the residuals would indicate a bias.

In the case of a binomial error structure with (0,1) outcomes, the residual plots as described above may not provide much added value. Due to the potential for many cells with small exposure, plots at a higher level of summarization may still not be much of an improvement. In this example, where the observed claim frequency is usually very low (generally less than 10%), these conditions are exacerbated

We have relied upon examination of tables of actual versus expected scaled claim frequency to provide validation. Since we fit models to loans originated in one year and validated them against loans originated in the following two years, it was necessary to scale the expected number of claims for latter two years to equal the observed number of claims. Any systematic departure from actual-to-expected ratios of 100% is evidence of bias. The results of such validation for the loan data indicate that the models fitted were robust with no significant bias.

4.5   Rating Factors

66

The five rating factors used for the models presented here are:

Credit Score: Of the primary borrower, as assessed by an external credit rating agency. Credit scores range from about 400 to 800, with higher scores indicating a better rating. We grouped credit score into 10 bands (low-648, 649-677, 678-697, 698-714, 715-728, 729-742, 743-755, 756-768, 769-782, 783-high) selected to evenly divide the exposures. The base class band is 715-728. Credit scores are whole numbers.

Loan Amount: In thousands of U.S. dollars, banded into seven groups (low-50, 50-75, 75-100, 100-125, 125-150, 150-175, 175-high). The base class is 75-100. Actual loan amounts are in dollars. The groups are formed such that the 50-75 group includes loans of at least $50,000, but less than $75,000.

Financial Commitment Ratio: Loan commitments as a percentage of salary, banded into 8 groups (low-18, 18-20, 20-22, 22-24, 24-26, 26-28, 28-30, 30-high). The base class is 20-22. As above, 20-22 means a commitment of at least 20%, but less than 22%.

Loan Term: The length of the loan payment schedule, presented in months and split into two groups (0-5 years, 5+ years). The base class is 5+ years.

Loan Purpose: Whether the loan is for a new venture or to refinance an existing loan. The base class is refinance.

Multiplying out the number of categories of rating factors gives a potential 2,240 cells for this particular model and requires 25 parameter estimates. The base class loan is for a borrower who rates a credit score between 715 and 728, has borrowed between $75,000 and $100,000, has a financial commitment ratio of between 20% and 22%, a loan term of more than 5 years, and is refinancing an existing loan.

In general, the variable groupings proceeded along natural boundaries. Some of the groups were selected to produce class levels of equal width or exposure content. The base class was generally selected as the largest or most central class.

Section 5. Binomial GLM Model Results

5.1   Explanation of the Graphs

The graphs which follow are relativity plots from the binomial model fit. The relativities have

had one subtracted from them. Therefore, positive relativities denote increased risk while

negative relativities imply decreased risk relative to the base class. The right-pointing triangle

indicates the relativity with the value displayed immediately to its right. The vertical bar to

the left of the triangle indicates the uncertainty of the relativity estimate as measured by its

standard deviation. In these plots, we have shown an 80% confidence interval, based on the

asymptotic normality of the maximum likelihood estimates. In some cases, however, the

extent of the confidence interval has been limited by placing an upper limit on the range

displayed. The base class for each rating factor has a relativity of one, which appears in the

graphs as zero with no error bar. The bars under each relativity indicate the level of exposure

for each category of the rating variable.

To calculate the overall relativity for a given cross-classification, the relativities are multiplied

together.   For example for a borrower with a credit score in the band 698-714, a loan

between $50,000 and $75,000, a financial commitment ratio between 26% and 28%, a loan

term less than 5 years, refinancing an existing loan, has a risk relativity of 43% relative to the

base class (0.43 = 1.42 x 0.77 x 1.71 x 0.23 x 1.00).

69

5.2   Discussion of Results

In this section, we present the raw results of the binomial GLM analysis   In deference to the proprietary nature of these underwriting and rating models, and for ease of presentation, we have treated the data in the following manner for this paper:

- We have transformed the underlying data so the numeric relationships shown in this paper are only illustrative;

- We have fitted a limited model of only five variables to the data, although there are additional explanatory variables; and

- We have treated the continuous variables as categorical, although it is statistically sub-optimal

Given these treatments, no quantitative conclusions should be drawn from the examples shown herein.

5.2.1 Credit Score


Risk Factor Relativities for Log-Odds
Binomial GLM
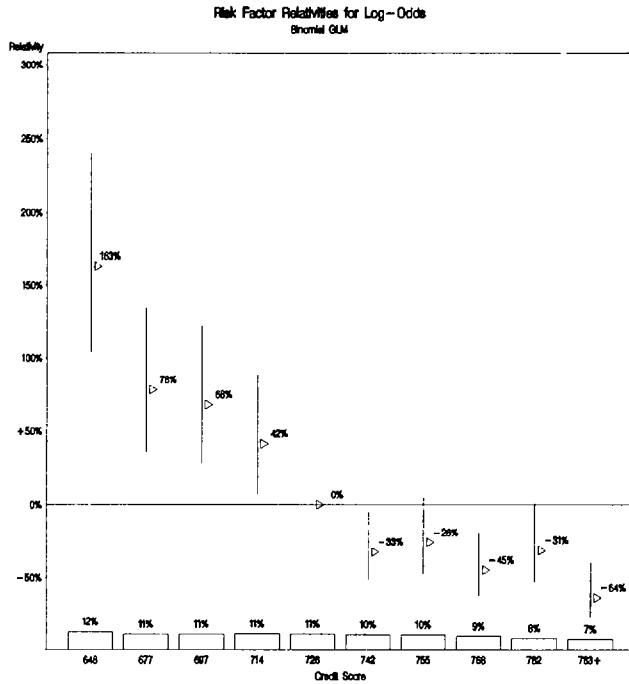
The shape of relativities for credit score is as expected, an almost monotonically decreasing function of credit score. In practice, it may be preferable to use credit score as a continuous variable (albeit transformed) and fit only one parameter instead of nine. A sensible transformation might be of the exponential or logistic form.

71

5.2.2 Loan Amount

Risk Factor Relativities for Log-Odds
Binomial GLM
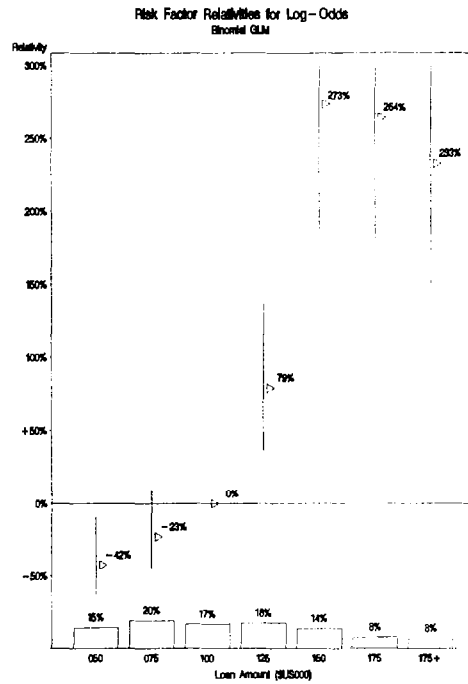
The results for loan amount behave generally as expected. We had thought that there would be a more gradual and monotonic pattern past $150,000. However, the error bars for the larger loan classes are particularly wide. It may well be that the apparent reduction in risk is a result of management action, such as increased underwriting for large loan values. If the indications between $150,000-200,000 were lower, loan amount could be fit as a continuous quantity, using a suitable transformation such as the hyperbolic tangent.

72

5.2.3 Commitment Ratio

Risk Factor Relativities for Log-Odds
Binomial GLM
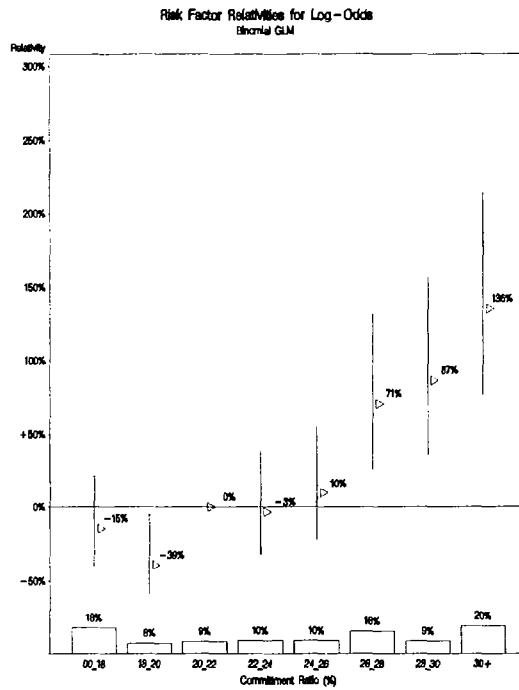
Again, the results for commitment ratio are largely in line with expectations, except perhaps for the apparent upturn for the lowest band. This hook may indicate that there is a base level of relative risk reduction. A hyperbolic tangent transformation may be appropriate for modeling this as a continuous variable. The transformation would also imply that at the upper end of the scale, there would be a limiting level of risk deterioration.

5.2.4   Loan Term

Risk Factor Relativities for Log-Odds Binomial GLM

Relativity axis with values: 300%, 250%, 200%, 150%, 100%, +50%, 0%, -50%

Bars: 000_60 = 8% (-77%), 61+ = 0% (92%)

Loan Term (Months)
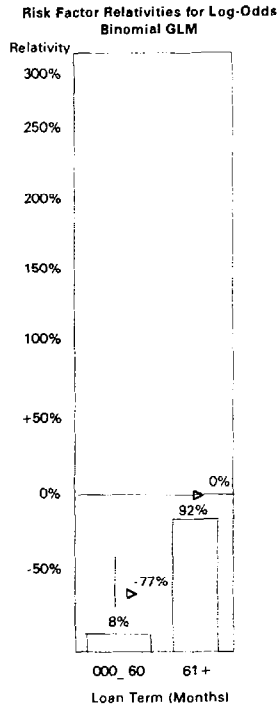
Loan terms of less than five years appear to be substantially less risky than loan terms greater than five years. This may be due to the quicker build-up of equity in the loan or more careful underwriting of shorter duration loans, which have lower profit potential.

74

5.2.5 Loan Purpose

Relativity / Loan Purpose chart showing: Other 14%, Refinance 86%, with markers at 239% and 0%.
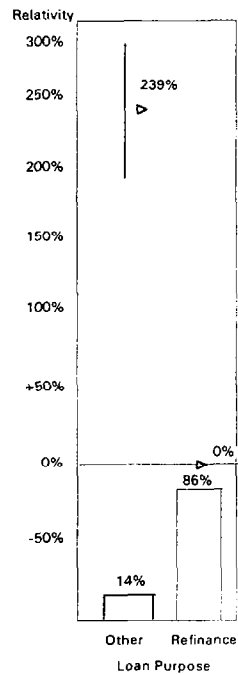
Refinanced loans appear to be less risky than new ventures. This is likely due to the stable history required for a bank refinance, while new ventures may be more uncertain.

5.3 Smoothing Results

The graphs presented in Section 5.2 display the raw relativities that come from the model fitting process. In practice, adjustments may be required before implementing the model. The error bars, as well as the exposure measures give an indication of the reliability of the particular estimates and the potential for these adjustments. In the case of the continuous variables, the shape of the relativities gives an indication for possible functional forms to be used for refitting. In addition, a practical model must take account of the fact that manual management intervention may not be in place in the future (such as for loan-to-value) and hence the shape of the relativities may need to be altered to reflect this. Finally, expense and profit allocation issues, as well as marketing focus, must be considered.

Appendix - Additional Models and Examples

As we've mentioned in the paper, Bailey [2], Bailey and Simon [3], and Brown [4] have introduced a number of models in their studies. The Poisson Maximum Likelihood and Least Squares Multiplicative models of earlier sections were from Bailey and Brown, respectively. However, Bailey did not develop his model as a MLE for the Poisson distribution. He had developed this model by assuming "the balance principle," or that the average error for any given class should be zero.

Expressed mathematically: For all i,

$$\frac{\sum_j n_{ij}\left(r_{ij} - x_i y_j\right)}{\sum_j n_{ij} r_{ij}} = 0$$

$$\sum_j n_{ij} r_{ij} - \sum_j n_{ij} x_i y_j = 0$$

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

which happens to be the Poisson model. The second line in the derivation above provides an additional interpretation of the balance principle. When viewing a fixed level of one of the row rating or column rating variables, we see that the total of the actual row (column), $\sum n_{ij} r_{ij}$, must equal the total estimated by the rating factors, $\sum_j n_{ij} x_i y_j$ for a row, or $\sum_i n_{ij} x_i y_j$ for a column.

Bailey also developed an additive model using the same constraint, which can be shown as:

$$\frac{\sum_j n_{ij}\left(r_{ij} - x_i - y_j - 1\right)}{\sum_j n_{ij} r_{ij}} = 0$$

$$\sum_j n_{ij}\left(r_{ij} - y_j - 1\right) - x_i \sum_j n_{ij} = 0$$

$$x_i = \frac{\sum_j n_{ij}\left(r_{ij} - y_j - 1\right)}{\sum_j n_{ij}}$$

which is <u>not</u> the MLE for an additive Poisson model.

We have chosen to present the additive model in a slightly different format than Bailey and Brown. Brown presents the "base rate" as $BR_a + x_i + y_j$ and $BR_m x_i y_j$ for the additive and multiplicative model, respectively. We have chosen to present these forms as, $BR\left(1 + x_i + y_j\right)$ and $BR x_i y_j$, respectively. The change in the additive form makes the discussion easier to follow because:

1) The loss cost relativity, r_{ij}, is on the same scale for either model. For example, a class that is 25% worse than the base class will have a relativity of 1.25, regardless of the model format. The scale in Bailey and Brown's interpretation is not so clear for the additive model. For example $r_{ij} = x_i + y_j$ could equal $75.00.

2) The same scale certainly makes the loss cost -vs- loss ratio discussion in Section 2 more easily understood.

In Bailey and Simon, a second multiplicative model was derived which minimizes the Chi-Squared value, rather than adhering to the balance principle. The Chi-Squared statistic is equal to:

$$Q = c \sum_i \sum_j \frac{n_{ij}\left(r_{ij} - x_i y_j\right)^2}{x_i y_j}$$

78

To minimize this, we set the partial derivatives equal to zero: $\dfrac{\partial Q}{\partial x_k} = 0$.

$$\sum_j \frac{-2y_j^2 x_k n_{kj}\left(r_{kj} - x_k y_j\right) - n_{kj} y_j \left(r_{kj} - x_k y_j\right)^2}{x_k^2 y_j^2} = 0$$

$$\sum_j 2x_k n_{kj}\left(r_{kj} - x_k y_j\right) + \frac{n_{kj}}{y_j}\left(r_{kj}^2 - 2x_k y_j r_{kj} + x_k^2 y_j^2\right) = 0$$

$$\sum_j 2x_k n_{kj} r_{kj} - 2x_k^2 n_{kj} y_j + \frac{n_{kj} r_{kj}^2}{y_j} - 2x_k n_{kj} r_{kj} + n_{kj} x_k^2 y_j = 0$$

$$\sum_j \frac{n_{kj} r_{kj}^2}{y_j} - x_k^2 n_{kj} y_j = 0$$

$$\Rightarrow \quad x_k^2 \sum_j n_{kj} y_j = \sum_j \frac{n_{kj} r_{kj}^2}{y_j}$$

$$\Rightarrow \quad x_k = \sqrt{\sum_j \frac{n_{kj} r_{kj}^2}{y_j} \Big/ \sum_j n_{kj} y_j}$$

Brown chose to approach the classification problem from the statistical standpoint. If the losses for ij^{th} cell (class) L_{ij} equals $n_{ij} r_{ij} P_B$, where P_B is the pure premium for the base class, then $E\left(L_{ij}\right) = n_{ij} p_B E\left(r_{ij}\right)$

$$= n_{ij} p_B x_i y_j, \quad \text{(or } n_{ij} P_B \text{ } (x_i + y_j + 1) \text{ for an additive model).}$$

Suppose we assume the losses in each cell to be distributed exponentially with parameter θ_{ij}, then $E\left(L_{ij}\right) = \theta_{ij}$, $f\left(L_{ij}\right) = \dfrac{1}{\theta_{ii}} e^{-L_{ij}/\theta_{ij}}$ and $n_{ij} p_B x_i y_j = \theta_{ij}$, so

$$f\left(L_{ij}\right) = \frac{1}{n_{ij}\,p_B\,x_i\,y_j}\,e^{-\left(n_{ij}r_{ij}p_B/n_{ij}p_B x_i y_j\right)}$$

$$= \frac{1}{n_{ij}\,p_B\,x_i\,y_j}\,e^{-\left(r_{ij}/x_i y_j\right)}$$

The likelihood function

$$L = \prod_{i,j} f\left(L_{ij}\right)$$

$$= \prod_{i,j} \frac{1}{n_{ij}\,p_B\,x_i\,y_j}\,e^{-\left(r_{ij}/x_i y_j\right)}$$

$$\ell n L = \sum_i \sum_j \ell n\left(\frac{1}{n_{ij}\,p_B\,x_i\,y_j}\right) + \ell n\left(e^{-r_{ij}/x_i y_j}\right)$$

$$= -\sum_i \sum_j \ell n\left(n_{ij}\right) + \ell n\left(p_B\right) + \ell n\left(x_i\right) + \ell n\left(y_j\right) + \left(r_{ij}/x_i\,y_j\right)$$

$$\frac{\partial\,\ell n L}{\partial\,x_k} = -\sum_j\left(\frac{1}{x_k} + \frac{r_{kj}}{x_k^2 y_j}\right) = 0$$

$$\Rightarrow \qquad \frac{1}{x_k}\sum_j 1 = \frac{1}{x_k^2}\sum_j r_{ij}/y_j$$

$$x_k = \frac{\sum_j r_{ij}/y_j}{\sum_j 1}$$

This approach can be used for additive models and with different distributions. For example, assume

$$L_{ij} \sim N\left(\mu_{ij}, \sigma_{ij}^2\right)$$

$$\mu_{ij} = n_{ij} p_B \left(x_i + y_j + 1\right)$$

$$\sigma_{ij}^2 = n_{ij}\sigma^2 \quad (see\ footnote\ 7)$$

$$f\left(L_{ij}\right) = \frac{1}{\sigma_{ij}\sqrt{2\pi}} e^{-\left[\frac{1}{2\sigma_{ij}^2}(L_{ij} - \mu_{ij})^2\right]}$$

$$= \frac{1}{\sigma\sqrt{2\pi\, n_{ij}}} e^{-\left[\frac{1}{2n_{ij}\sigma^2}\left(r_{ij} n_{ij} p_B - n_{ij} p_B (x_i + y_j + 1)\right)^2\right]}$$

$$= \frac{1}{\sigma\sqrt{2\pi\, n_{ij}}} e^{-\frac{p_B^2 n_{ij}}{2\sigma^2}\left(r_{ij} - x_i - y_j - 1\right)^2}$$

$$\Rightarrow \quad \ell n L = \sum_i \sum_j - \ell n\left(\sigma\sqrt{2\pi/n_{ij}}\right) - p_B^2 \sum_i \sum_j n_{ij}\left(r_{ij} - x_i - y_j - 1\right)^2 / 2\sigma^2$$

$$\frac{\partial\, \ell n L}{\partial\, x_k} = \frac{p_B^2}{2\sigma^2}\sum_j 2n_{kj}\left(r_{kj} - x_i - y_j - 1\right) = 0$$

$$\sum_j n_{kj}\left(r_{kj} - y_j - 1\right) = x_k \sum_j n_{kj}$$

$$x_k = \sum_j n_{kj}\left(r_{kj} - y_j - 1\right) \Big/ \sum_j n_{kj}$$

which is the same as the Bailey [2] additive model. This solution can also be used for a multiplicative lognormal model by taking the logarithm of the data.

These are obviously just samples of many possible models involving different distributional assumptions. For a three (or more) variable model, one could use a mixed additive-multiplicative model, where $r_{ijk} = x_i y_j + z_k$. This would be solved using the same process.

---

7 This form essentially assumes that each exposure is independent and distributed $N(P_B(x_i + y_j + 1), \sigma^2)$. As the sum of normal random variables is normal, the distribution of the cell losses, L_{ij}, follows.

In concluding this appendix, we thought it might assist the reader in working with the models shown if we gave the values of the class plan, for the two variable example in the paper, solved using each of the models discussed.

First Iteration

	Bailey Additive	Bailey-Simon Multiplicative	Exponential Multiplicative
X_1	0.500	1.155	1.083
X_2	4.588	3.448	3.367
X_3	13.556	6.867	7.356
Y_1	-3.408	1.054	.994
Y_2	8.111	3.911	4.026
Initial Y_1, Y_2	(0, 3)	(1, 4)	(1, 4)

Second Iteration

	Bailey Additive	Bailey-Simon Multiplicative	Exponential Multiplicative
X_1	-2.482	1.175	1.083
X_2	5.490	3.439	3.365
X_3	13.177	6.870	7.363
Y_1	-3.738	1.054	.994
Y_2	8.607	3.910	4.026

Converged Solution

	Bailey Additive	Bailey-Simon Multiplicative	Exponential Multiplicative
X_1	-2.802	1.175	1.083
X_2	5.587	3.439	3.365
X_3	13.136	6.870	7.363
Y_1	-3.774	1.054	.994
Y_2	8.660	3.910	4.026

The Bailey additive model appears to be much more sensitive than the other two models to the choice of initial values. In fact, the implied rate for the base class is negative. This result occurs in part because L1 is the smallest class in terms of exposures and has the lowest frequency. The failure of the simple additive model to reflect interactions contributes to the dilemma as well. These observations, coupled with the balance principle, result in a nonsensical model. Using another base class or Bailey's original model, as previously noted, continues to produce the unreasonable result. If there were more levels for each class, the model could also be constrained to have x_l and y_l equal zero, but the iterative formulas would change. This entire problem is one argument in favor of multiplicative models rather than additive models.

These relativities can be multiplied (or added) together, and compared to the actual relativities using validation techniques discussed in the paper. In the example, for the largest class, M1, the various model relativities are displayed in the table below.

M1 Class Relativities

Method	Relativity
Actual Data (r_{ij})	3.083
One-Way Method	1.725
Bailey - Simon Multiplicative	2.926
Exponential Multiplicative	3.108
Poisson MLE (Bailey Multiplicative)	2.920

Bibliography

1. SAS Institute Inc., *SAS/STAT Software, Changes and Enhancements through Release 6.11 (55356)*, (Cary: SAS Institute Inc., 1996).

   SAS Institute Inc., *SAS/INSIGHT Users Guide, Version 6, Third Edition (55382)*, (Cary: SAS Institute Inc., 1996).

2. R. A. Bailey, *Insurance Rates with Minimum Bias* (PCAS L, 1963), p. 4.

3. R. A. Bailey and L. Simon, *Two Studies in Automobile Insurance Ratemaking* (Astin Bulletin, 1, 1960) p. 192.

4. R.L. Brown, *Minimum Bias with Generalized Linear Models* (PCAS LXXV, 1988) p. 187.

5. M. Aitkin, D.Anderson, B. Francis, and J. Hinde, *Statistical Modelling in GLIM*, (Oxford: Clarendon Press, 1989).

6. P. McCullagh and J. A. Nelder, *Generalized Linear Models*, (London: Chapman and Hall, 1989).

7. S. R. Searle, *Linear Models*, (New York: John Wiley and Sons, 1971).

8. D. A. Belsley, E. Kuh, and R. E. Welsch, *Regression Diagnostics*, (New York: John Wiley and Sons, 1980).

9. M. J. Brockman and T. S. Wright, *Statistical Motor Rating*, (JIA No. 474, with discussion).

10. R. V. Hogg and S. A. Klugman, *Loss Distributions*, (New York: John Wiley and Sons, 1984).

11. D. G. Hart, R. A. Buchanan, and B. A. Howe, *Actuarial Practice of General Insurance, 5th Edition*, (Sydney: Institute of Actuaries of Australia,1996).