Introduction to Ratemaking
Multivariate Methods

Part 2 of Ratemaking Relativites - *Introduction to Multivariate Methods*

March 9, 2009
Mirage Hotel
Las Vegas, NV

Chris Cooksey, FCAS, MAAA
Nationwide Insurance Company

Ain Milner, FCAS, MAAA
Safeco Insurance Company, a Member of Liberty Mutual

Chris Stoll, FCAS, MAAA
Allstate Insurance Company

Introduction to Ratemaking

Multivariate Methods

Insurance is inherently a stochastic (random) process. Any set of data you examine will contain random results in addition to true relationships.

$$\text{Dependent Variable} = \text{Signal} + \text{Noise}$$

$$\text{Dependent Variable} = \text{Systematic Component} + \text{Random Component}$$

Note: the presence of noise along with our signal is the basic reason credibility was conceived. Due to the presence of noise, we don't fully believe our point estimate.

Introduction to Ratemaking
Multivariate Methods

One-way pure premiums, loss per exposure, are a straight-forward method for determining relativities.

Why don't we just look at pure premiums by relativity in order to set relative rates?

The problem with them is they are blind to the rest of the class plan.

Introduction to Ratemaking
Multivariate Methods

One-way pure premiums, loss per exposure, are a straight-forward method for determining relativities.

For example, you look at pure premiums for youthfuls and find they deserve to be charged 2.00 times the rate of adults.

Then you look at points and pure premiums say that pointed drivers should be charged 1.50 times that of clean drivers.

Should a youthful with points get charged 3.00 times the rate of clean adults (1.50 * 2.00)?

Maybe, maybe not. It depends on whether there is a distributional correlation between age and points. Are young drivers more likely also to have points? If so, you've overcharged.

Introduction to Ratemaking
Multivariate Methods

One-way loss ratios are the most convenient alternative to pure premiums. They are inherently multivariate because the premium "takes into account" the rest of the class plan.

For example, if you look at the relative loss ratios between youthful and adult drivers, the premium within that loss ratio will reflect the current factors for points.

Because youthfuls have a higher percentage of points, their average premium will be higher due to the higher pointed factors. This will lower the loss ratio. In this way we don't "double count" the effect of points and age.

Side note…what if points didn't exist? Appropriate age factors would change.

Introduction to Ratemaking
Multivariate Methods

Why aren't one-way loss ratios sufficient?

One-way studies using loss ratios assume that the rest of the class plan is good. This is a big assumption when there are multiple changes which need to be made.

Suppose you want to examine the adequacy of both your age and points curves. When you look at loss ratios by age, you are assuming your current points factors are good. Vice versa for when you look at loss ratios by points.

Introduction to Ratemaking
Multivariate Methods

Another point of confusion: correlations versus interactions.

Correlations between two variables' exposure distributions cause the results to be linked (remember points and age). This is NOT an interaction. It is an important effect and using multivariate techniques solves this problem.

Interactions are correlations between two variables' indicated factors. When you don't know what factor to use until both variables are specified, you have an interaction.

It is perfectly possible for two variables to be correlated but have no interaction. It is also possible for two variables to have an interaction but not be correlated!

Correlation but no interaction

Exposures	*Clean*	*Pointed*	
Younger	50	100	150
Older	500	500	1000
	550	600	1150

Loss	*Clean*	*Pointed*	
Younger	1,500	4,500	6,000
Older	5,000	7,500	12,500
	6,500	12,000	18,500

Pure Premium	*Clean*	*Pointed*
Younger	30	45
Older	10	15

Introduction to Ratemaking
Multivariate Methods

Interaction but no correlation

Exposures	*Clean*	*Pointed*	
Younger	50	100	150
Older	450	900	1350
	500	1000	1500

Loss	*Clean*	*Pointed*	
Younger	1,500	6,000	7,500
Older	6,750	40,500	47,250
	8,250	46,500	54,750

Pure Premium	*Clean*	*Pointed*
Younger	30	60
Older	15	45

Introduction to Ratemaking
Multivariate Methods

<u>Correlation but no interaction</u> – Prove that a one-way pure premium approach doesn't work.

Exposures	*Clean*	*Pointed*	
Younger	50	100	150
Older	500	500	1000
	550	600	1150

Loss	*Clean*	*Pointed*	
Younger	1,500	4,500	6,000
Older	5,000	7,500	12,500
	6,500	12,000	18,500

Pure Premium	*Clean*	*Pointed*
Younger	30	45
Older	10	15

Introduction to Ratemaking
Multivariate Methods

<u>Correlation but no interaction</u> – Prove that a one-way pure premium approach doesn't work.

Age	*Exp*	*Loss*	*PP*	*PP Rel*
Younger	150	6,000	40.00	3.20
Older	1000	12,500	12.50	1.00

Points	*Exp*	*Loss*	*PP*	*PP Rel*
Clean	550	6,500	11.82	1.00
Pointed	600	12,000	20.00	1.69

Pure Premium Relativity	*Clean*	*Pointed*
Younger	3.20	5.41
Older	1.00	1.69

20% too high!

Multivariate Approach

Pure Premium	*Clean*	*Pointed*
Younger	30	45
Older	10	15

Pure Premium Relativity	*Clean*	*Pointed*
Younger	3.00	4.50
Older	1.00	1.50

Minimum Bias Techniques

- Multivariate procedure to optimize the relativities for two or more rating variables
- Calculate relativities which are as close to the actual relativities as possible
- "Close" measured by some bias function
- Bias function determines a set of equations relating the observed data & rating variables
- Use iterative technique to solve the equations and converge to the optimal solution

Introduction to Ratemaking
Multivariate Methods

Minimum Bias Techniques

- Two rating variables with relativities X_i and Y_j
- Select initial value for each X_i
- Use model to solve for each Y_j
- Use newly calculated Y_js to solve for each X_i
- Process continues until solutions at each interval converge

Introduction to Ratemaking
Multivariate Methods

Minimum Bias Techniques

- Least Squares
- Bailey's Minimum Bias

Minimum Bias Techniques – Least Squares Method

- Minimize weighted squared error between the indicated and the observed relativities

- i.e., $\text{Min}_{xy} \ \sum_{ij} w_{ij} (r_{ij} - x_i y_j)^2$

where

x_i and y_j = relativities for rating variables i and j

w_{ij} = weights

r_{ij} = observed relativity

Minimum Bias Techniques – Least Squares Method

Formula:

$$x_i = \frac{\sum_j w_{ij} \; r_{ij} \; y_j}{\sum_j w_{ij} \; (y_j)^2}$$

Note: this formula is specific to a multiplicative model

where

x_i and y_j = relativities for rating variables i and j

w_{ij} = weights

r_{ij} = observed relativity

Minimum Bias Techniques – Bailey's Minimum Bias

- Minimize bias along the dimensions of the class system
- "Balance Principle" :

  $\sum$ observed relativity = $\sum$ indicated relativity
- i.e., $\sum_j w_{ij} r_{ij} = \sum_j w_{ij} x_i y_j$

where

x_i and y_j = relativities for rating variables i and j

w_{ij} = weights

r_{ij} = observed relativity

Minimum Bias Techniques – Bailey's Minimum Bias

Formula:

$$x_i = \frac{\sum_j \; w_{ij} \; r_{ij}}{\sum_j \; w_{ij} \; y_j}$$

Note: this formula is specific to a multiplicative model

where

x_i and y_j = relativities for rating variables i and j

w_{ij} = weights

r_{ij} = observed relativity

Minimum Bias Techniques

- Bailey's method is less sensitive to the experience of single cells than the Least Squares method.
- Can be multiplicative or additive.
- Can be used for many dimensions (conversion can be difficult).
- Possible to code the calculation directly into a spreadsheet.

Minimum Bias Techniques

These techniques give only point estimates, yet we know all data contains both signal and noise. Minimum bias techniques provide no method for quantifying the extent and impact of the noise.

Introduction to Ratemaking
Multivariate Methods

Classical Statistical Techniques

$$\text{Dependent Variable} = \text{Signal} + \text{Noise}$$

Simple linear regression…

$$y = a_1 x + a_0 + \varepsilon$$

Multiple linear regression…

$$y = a_n x_n + \ldots + a_1 x_1 + a_0 + \varepsilon$$

Classical Statistical Techniques

This example has a couple of categorical variables, so we would formulate the model as…

$y = a_1x_1 + a_2x_2 + a_3x_3 + \varepsilon$

where 1 means younger,
2 means older, and
3 means clean)

Loss	*Clean*	*Pointed*
Younger	1,500	4,500
Older	5,000	7,500

Classical Statistical Techniques

With 4 observations of this proposed relationship, we get…

Loss	*Clean*	*Pointed*
Younger	1,500	4,500
Older	5,000	7,500

$$y = a_1x_1 + a_2x_2 + a_3x_3 + \varepsilon$$

$$1500 = a_1 + 0 + a_3 + \varepsilon_1$$

$$4500 = a_1 + 0 + 0 + \varepsilon_2$$

$$5000 = 0 + a_2 + a_3 + \varepsilon_3$$

$$7500 = 0 + a_2 + 0 + \varepsilon_4$$

To find an answer, we need a criterion for what is the “best” answer. A typical approach is to minimize the sum of the squared errors (SSE).

$$SSE = \varepsilon_1^2 + \varepsilon_2^2 + \varepsilon_3^2 + \varepsilon_4^2$$

Classical Statistical Techniques

Minimizing the SSE in this simple example is easily done by taking partial derivatives (with respect to each coefficient) and setting them equal to zero.

This gives you a system of 3 equations with 3 unknowns…easy to solve.

Loss	*Clean*	*Pointed*
Younger	1,500	4,500
Older	5,000	7,500

Try it out. I got… $a_1 = 4,375$
$a_2 = 7,625$
$a_3 = -2,750$

Classical Statistical Techniques

Finding the optimal answer for a multiple linear regression boils down to systems of equations.

However, systems of equations, especially as the number of variables and observations get more numerous, are more conveniently expressed through matrix notation and linear algebra.

Go back to our example…

Classical Statistical Techniques

$\underline{Y} = [Y_1, Y_2, Y_3, Y_4] = [1500, 4500, 5000, 7500]$ ← 4 observations

$\underline{X}_1 = [1, 1, 0, 0]$ ← 4 observations; 1 if younger, 0 if older

$\underline{X}_2 = [0, 0, 1, 1]$ ← 4 observations; 1 if older, 0 if younger

$\underline{X}_3 = [1, 0, 1, 0]$ ← 4 observations; 1 if clean, 0 if pointed

$\underline{A} = [a_1, a_2, a_3]$ ← solving for these; one per x-variable

$\underline{\varepsilon} = [\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4]$ ← 4 observations

Classical Statistical Techniques

Now collapse the separate x-variables into one matrix…

$$\mathbf{X} = \begin{bmatrix} 1, 1, 0, 0 \\ 0, 0, 1, 1 \\ 1, 0, 1, 0 \end{bmatrix}$$

← This is called the "design matrix" because it specifies your model.

With this final simplification, we can express the example situation using matrix notation…

$$\underline{Y} = \mathbf{X}.\underline{A} + \underline{\varepsilon}$$

(Note: this is the "fancy" version of Y = signal + noise)

Introduction to Ratemaking
Multivariate Methods

We've developed MLR into a simplified form where we can see it is one approach for separating the signal and the noise. But what assumptions have we made along the way, and do we like them?

1. (*Random Component*) Observations are independent and come from a normal distribution with a common variance.

2. (*Systematic Component*) $\mathbf{X}.\underline{A}$ is called the linear predictor, or $\underline{\eta}$.

3. (*Link function*) The expected value of $\underline{Y}$, $E(\underline{Y})$, is equal to $\underline{\eta}$.

Introduction to Ratemaking
Multivariate Methods

One by one…how well do these assumptions work for insurance?

1. (*Random Component*) Observations are independent and come from a normal distribution with a common variance.

For each variable in our model, there is an expected mean and some randomness about that mean. The average loss for "younger drivers" may be $725, but why should the distribution of individual observations be normal about this? In fact, normal distributions extend to negative numbers. What's a negative loss?

This assumption doesn't work well for insurance.

Introduction to Ratemaking
Multivariate Methods

One by one…how well do these assumptions work for insurance?

1. (*Random Component*) Observations are independent and come from a normal distribution with a common variance.

Another problem of this assumption is the common variance. Why should the distribution of losses for 15K limits have the same variance as the distribution of losses for 100K limits? Wouldn't you assume that 15K limits, with a low mean, would have less variance than 100K limits?

Again, this assumption doesn't work well for insurance.

Introduction to Ratemaking
Multivariate Methods

One by one…how well do these assumptions work for insurance?

2. (*Systematic Component*) **X**.$\underline{A}$ is called the linear predictor, or $\underline{\eta}$.
3. (*Link function*) The expected value of $\underline{Y}$, E($\underline{Y}$), is equal to $\underline{\eta}$.

This pair assumes that $\underline{Y}$ is predicted by the additive combination of the variables. However, most insurance effects tend to combine multiplicatively.

Again, this assumption doesn't work well for insurance.

Introduction to Ratemaking
Multivariate Methods

If the MLR assumptions don't work well for insurance, then change them! With the same general approach, but the following assumptions, you've transitioned from MLRs to GLMs.

1. (*Random Component*) Observations are independent, but come from one of the family of exponential distributions.
2. (*Systematic Component*) $\mathbf{X}.\underline{A}$ is called the linear predictor, or $\underline{\eta}$.
3. (*Link function*) The expected value of $\underline{Y}$, $E(\underline{Y})$, is equal to $g^{-1}(\underline{\eta})$.

Introduction to Ratemaking
Multivariate Methods

One by one…how well do these new assumptions work for insurance?

1. (*Random Component*) Observations are independent, but come from one of the family of exponential distributions.

Now we can assume that the distribution of severities about the mean follows a Gamma, and frequencies follow a Poisson. These functions happen to match empirical evidence fairly well and they don't allow negative output.

Also, the variances of these functions are functions of the mean, so that variables with low means also have low variances.

Introduction to Ratemaking
Multivariate Methods

One by one…how well do these new assumptions work for insurance?

2. (*Systematic Component*) **X**.$\underline{A}$ is called the linear predictor, or $\underline{\eta}$.
3. (*Link function*) The expected value of $\underline{Y}$, E($\underline{Y}$), is equal to $g^{-1}(\underline{\eta})$.

Instead of having $\underline{Y}$ automatically equal the additive effects of the predictors, we can let the predictors equal some function of the expectation of $\underline{Y}$.

$$g(\, E(\underline{Y})\,) = \underline{\eta} \quad \leftarrow \text{g is called the link function}$$

or

$$E(\underline{Y}) = g^{-1}(\underline{\eta})$$

Introduction to Ratemaking
Multivariate Methods

One by one…how well do these new assumptions work for insurance?

2. (*Systematic Component*) $\mathbf{X}.\underline{A}$ is called the linear predictor, or $\underline{\eta}$.
3. (*Link function*) The expected value of $\underline{Y}$, $E(\underline{Y})$, is equal to $g^{-1}(\underline{\eta})$.

For example, say we pick the log-link, or $g(x) = \ln(x)$

$$g(\, E(\underline{Y})\,) = \underline{\eta} = \ln(\, E(\underline{Y})\,)$$

or

$$E(\underline{Y}) = e^{(\underline{\eta})}$$

If $\underline{\eta}$ is an additive combination of $x_1a_1 + x_2a_2$, we get…

$$E(\underline{Y}) = e^{(\underline{\eta})} = e^{(x_1a_1+x_2a_2)} = e^{(x_1a_1)}\, e^{(x_2a_2)} \qquad \leftarrow \text{multiplicative!}$$

Introduction to Ratemaking
Multivariate Methods

GLMs have the general form… $\underline{Y} = \mathbf{X}.\underline{A} + \underline{\varepsilon}$
and make the preceding assumptions.

A significant first step in modeling is choosing which link and error functions you will use.

After that, you are deciding the final form of your design matrix. In other words, which variables do you want in your model and how will you combine them? This process is best done through an evaluative, trial and error process that combines both statistics and judgment.

Introduction to Ratemaking
Multivariate Methods

In Summary…

As a statistical model, GLMs allow us to have some measure of the noise as well as the signal.

As statistical models go, GLMs and their attending assumptions are flexible enough to reasonably fit real-world insurance situations.

While we discarded other approaches along the way, it turns out that many minimum bias techniques and all one-way and linear regression approaches are just special forms of GLMs.

GLMs are multivariate and automatically solve the "double counting" problem presented by correlated variables. They also allow for many model forms, including interactions.

Introduction to Ratemaking
Multivariate Methods

Other Multivariate Techniques

While GLM's appear to be the current industry standard, there are other multivariate techniques. These include…

- Decision Trees (CART, C5, CHAID, etc.)
- Neural Networks
- Polynomial Networks
- Clustering
- Kernels
- Others…