

Predictive Modeling for Small Companies

**Prepared For: Casualty Actuarial Society
2009 Ratemaking & Product Management Seminar**

**Prepared By: Paul D. Anderson, FCAS, MAAA
paul.anderson@milliman.com
(262) 641-3531**

March 10, 2009

Antitrust Notice

The Casualty Actuarial Society is committed to adhering strictly to the letter and spirit of the antitrust laws. Seminars conducted under the auspices of the CAS are designed solely to provide a forum for the expression of various points of view on topics described in the programs or agendas for such meetings.

Under no circumstances shall CAS seminars be used as a means for competing companies or firms to reach any understanding – expressed or implied – that restricts competition or in any way impairs the ability of members to exercise independent business judgment regarding matters affecting competition.

It is the responsibility of all seminar participants to be aware of antitrust regulations, to prevent any written or verbal discussions that appear to violate these laws, and to adhere in every respect to the CAS antitrust compliance policy.



Overview of Issues

- How do you start?
 - Ø Literature/Resources, Software, etc.
- Approval/Buy-in from Management
- Preparing the Data
- Analyzing the Data
- Implementation Issues
- Filing/Regulatory Issues
- Communication Issues

Data Prep: Common Challenges

- Combining datasets for the first time
 - ∅ Duplicate records may result in one or more dataset
 - ∅ Records may be lost in the process of merging
 - ∅ Inconsistent formats of similar data fields
 - ∅ Must locate linking variables on each dataset
- Investigating some data fields for the first time
 - ∅ Data field exists, but not used
 - All missing values or all the same value (try to clean up for future use)
 - Not useful -- data field can be removed from analysis
 - ∅ Data field exists, but coding unknown
 - Can be cross-referenced with other fields to determine coding
 - May need to research individual policies to decode

Data Analysis: Common Challenges

Most are generally due to not enough data (of course!):

1. Non-intuitive, unexpected, or exaggerated results
 - ∅ Discount ends up as indicated surcharge
 - ∅ Surcharge ends up indicating a 600% surcharge
 - ∅ Variable with multiple segments indicates factors that bounce up & down instead of a gradual progression from low to high
2. Results that vary significantly from one year to the next
3. Skewed distributions within rating variables
 - ∅ Can result in segments with no exposures
4. High degree of correlation between some variables

Possible Solutions to Challenges

1. Non-intuitive, unexpected, or exaggerated results
 - ∅ Cap individual losses or look at Frequency-only model
 - ∅ Group small, similar segments together (extrapolate later)
 - ∅ Run model iteratively/sequentially
 - Evaluate only a few variables at a time
 - Make selections, adjust data, and rerun model with new variables
 - ∅ Rely on current factors, competitors, or judgment

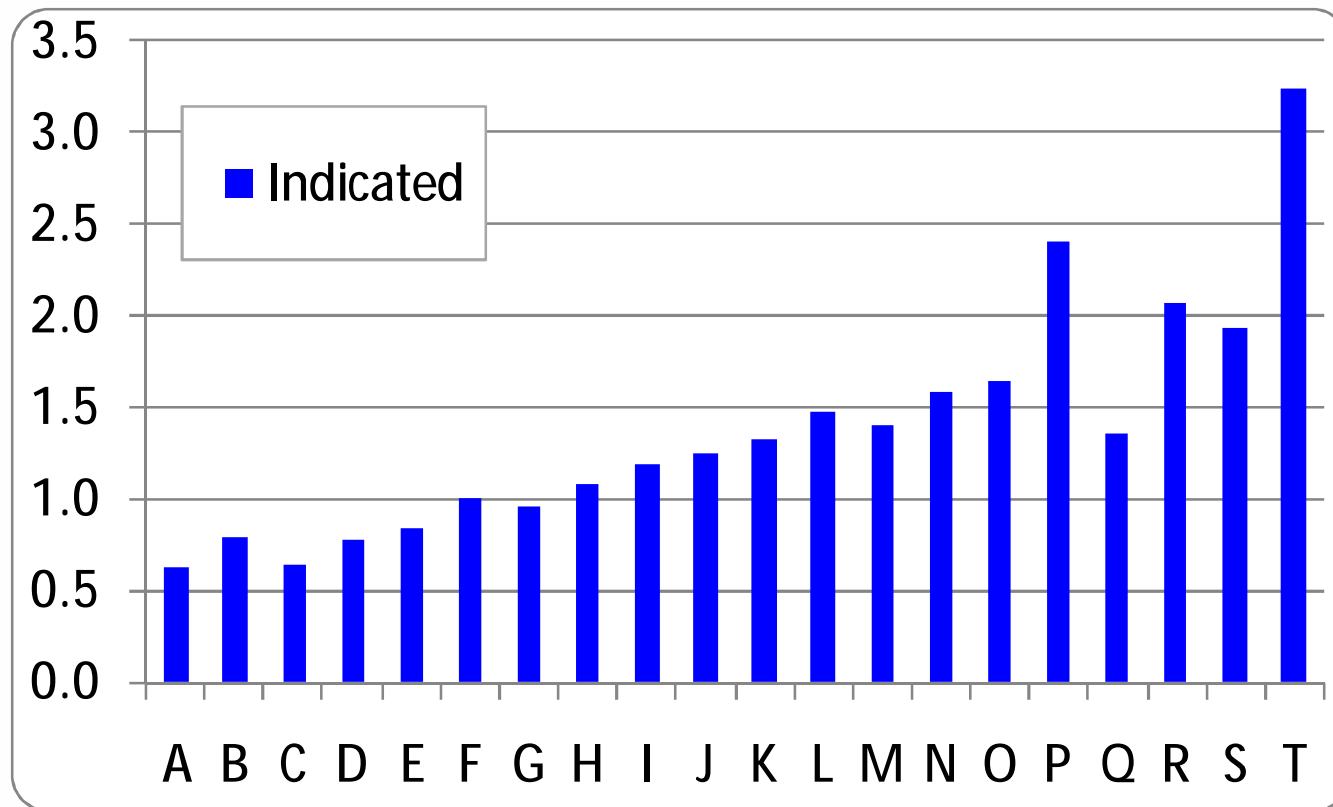
2. Results that vary significantly from one year to the next
 - ∅ Consider & adjust for unique events within individual years
 - ∅ Use as many years as possible to build model
 - ∅ When validating model, be sure to include multiple years

Possible Solutions to Challenges

3. Skewed distributions within rating variables
 - ∅ Rely on larger segments to make selections
 - ∅ For smaller segments, maintain current relationship or use judgment to establish relationship between segments
4. High degree of correlation between some variables
 - ∅ If on-going correlation is expected, remove one or more of the variables or combine into a compound variable
 - ∅ If correlation is due to small dataset, make selections for one or more variables, adjust data, and run model again

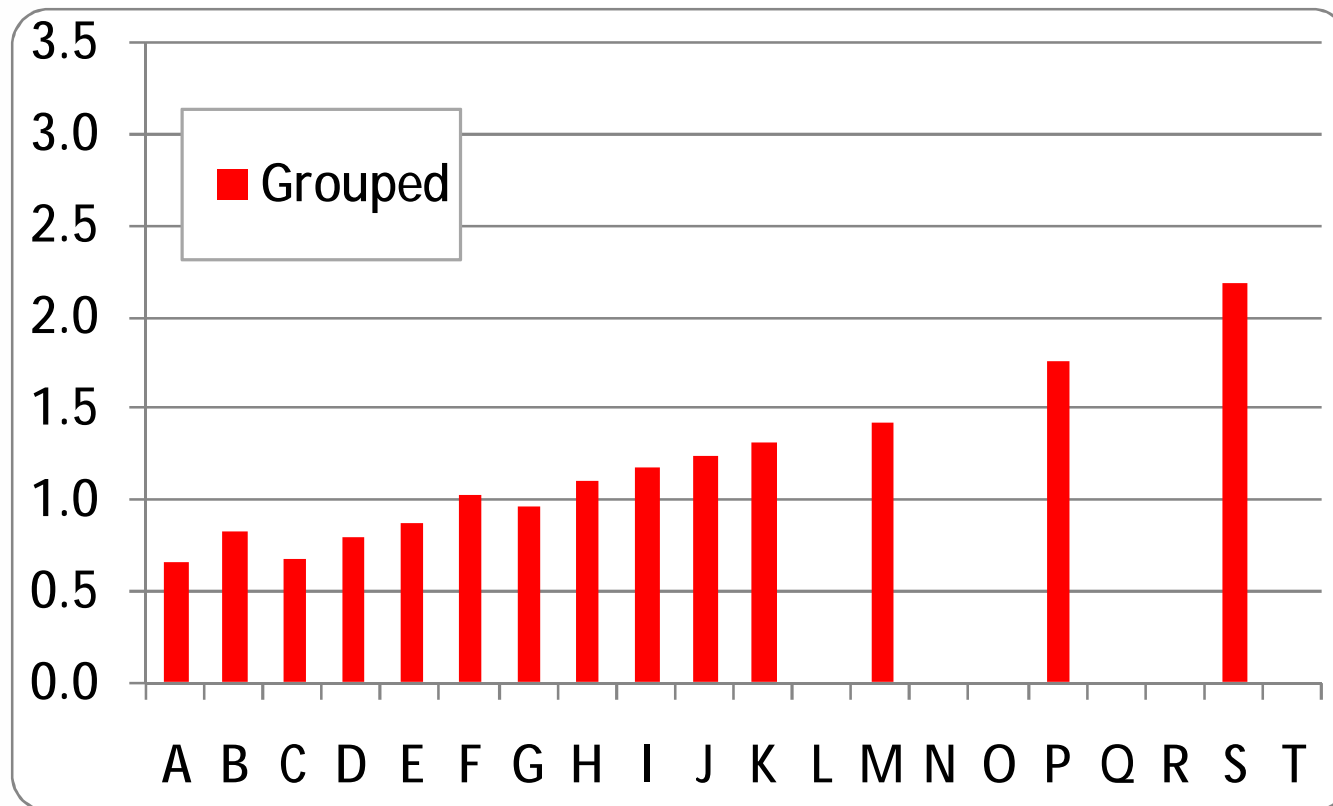
Example of Unexpected Results

Preliminary indicated credit factors show a definite pattern, but they include some “noise” as well



Example of Unexpected Results

Indicated factors after combining several of the smaller segments show the same pattern with less “noise”



Example of Unexpected Results

After grouping some segments, the selection process is simple -- a little smoothing, interpolation, & extrapolation

