Marrying Underwriter Intuition & Predictive Modeling - A Workers' Compensation Perspective

CAS RPM Seminar

Las Vegas Mar 10, 2008

Presented by:
Gaétan Veilleux, FCAS
Matt Frazier, Client Lead

VALEN technologies

A Short Vignette

After Several Weeks of Analysis...

the Predictive Modeler and
the Underwriter

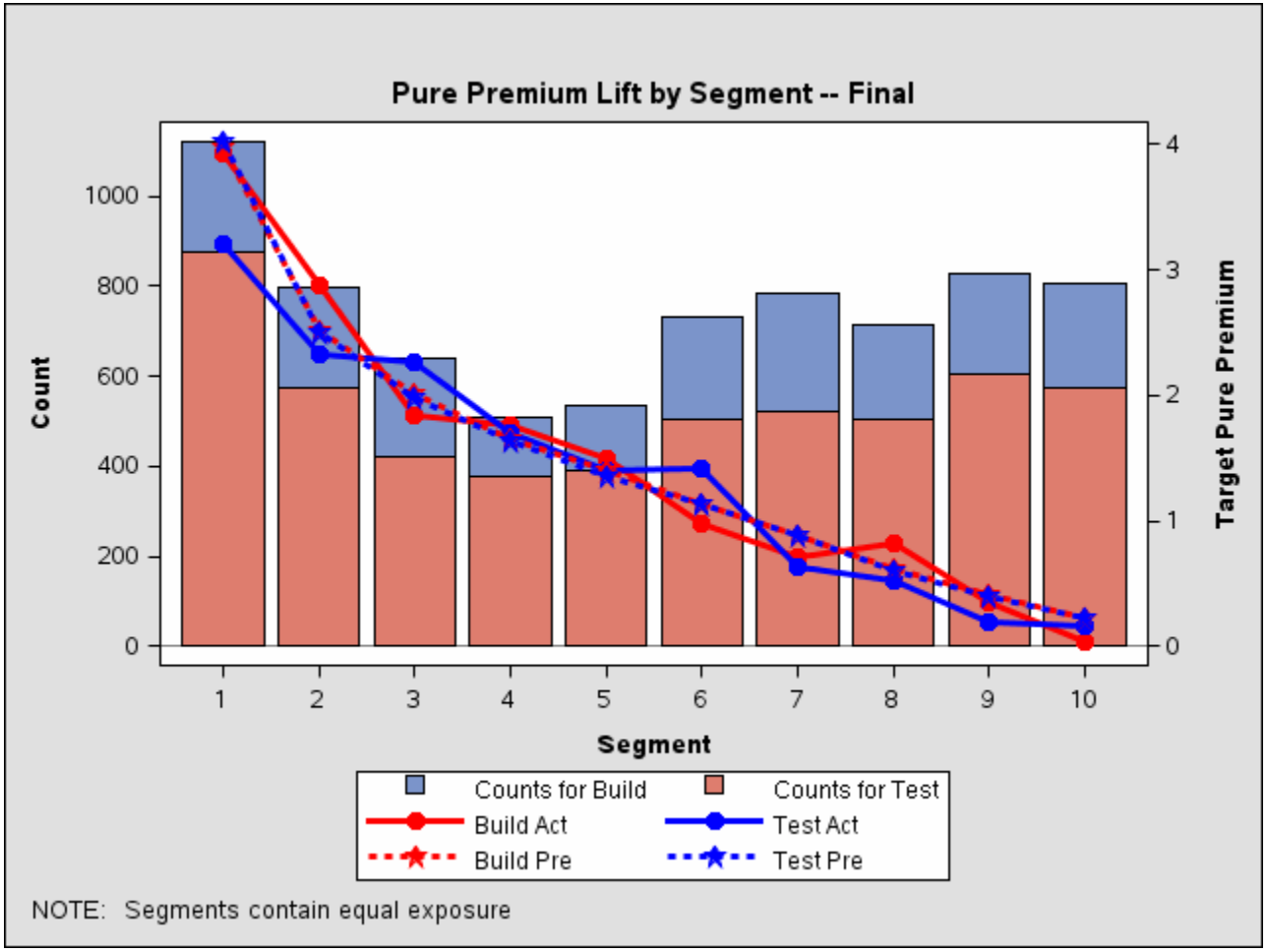
Meet...

VALEN

Initial Candidate Model Risk Factors

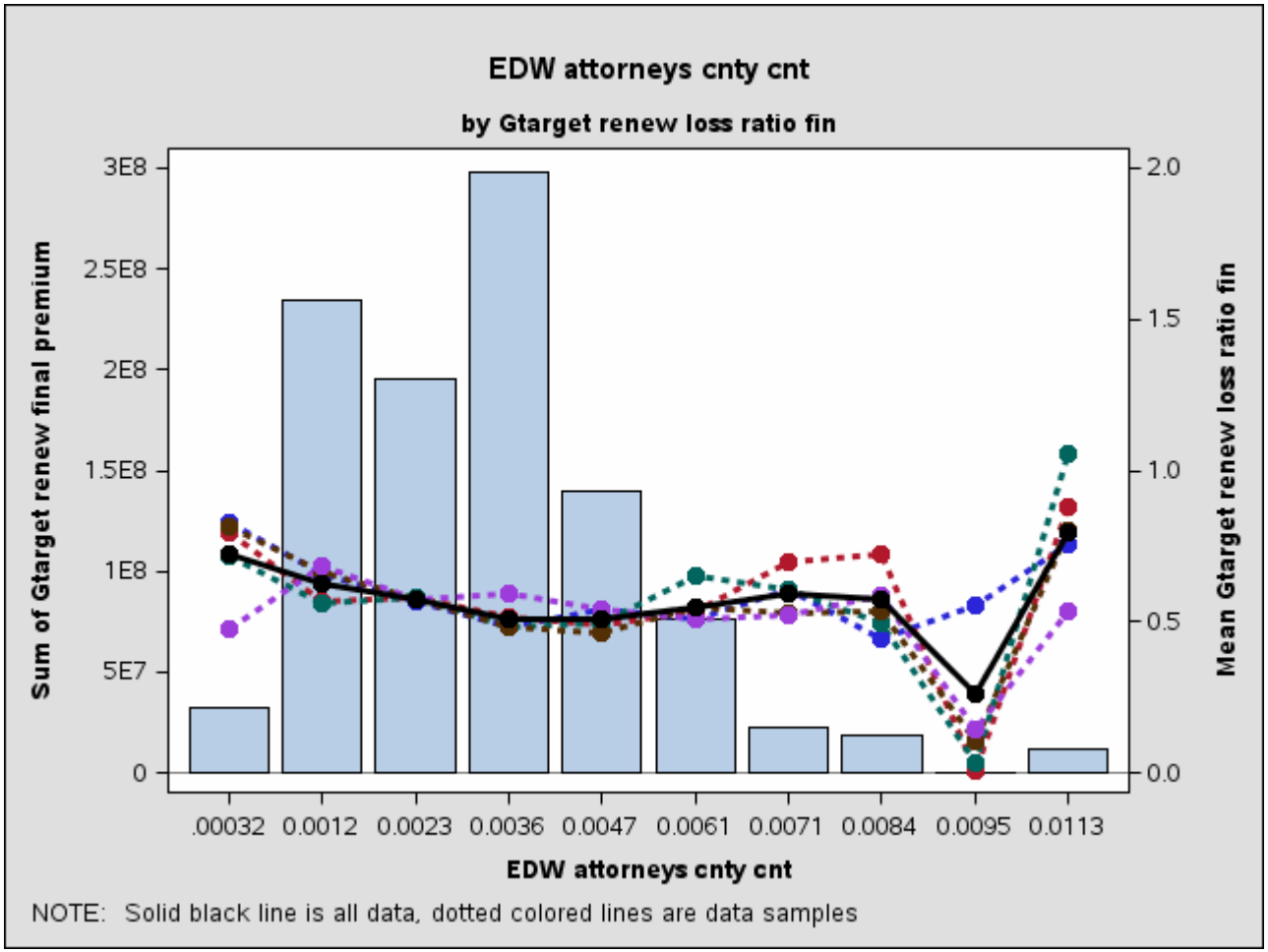
- Attorney count by county
- Hail size by county
- Late reporting >30 days
- Frequency –1 year back
- Principal factor – up to 3 term LR, up to 3 term PP
- Term sequence
- Binned Manual Premium
- Natural log (Class code/176 + 11,237)…(?!)

Initial Candidate Model



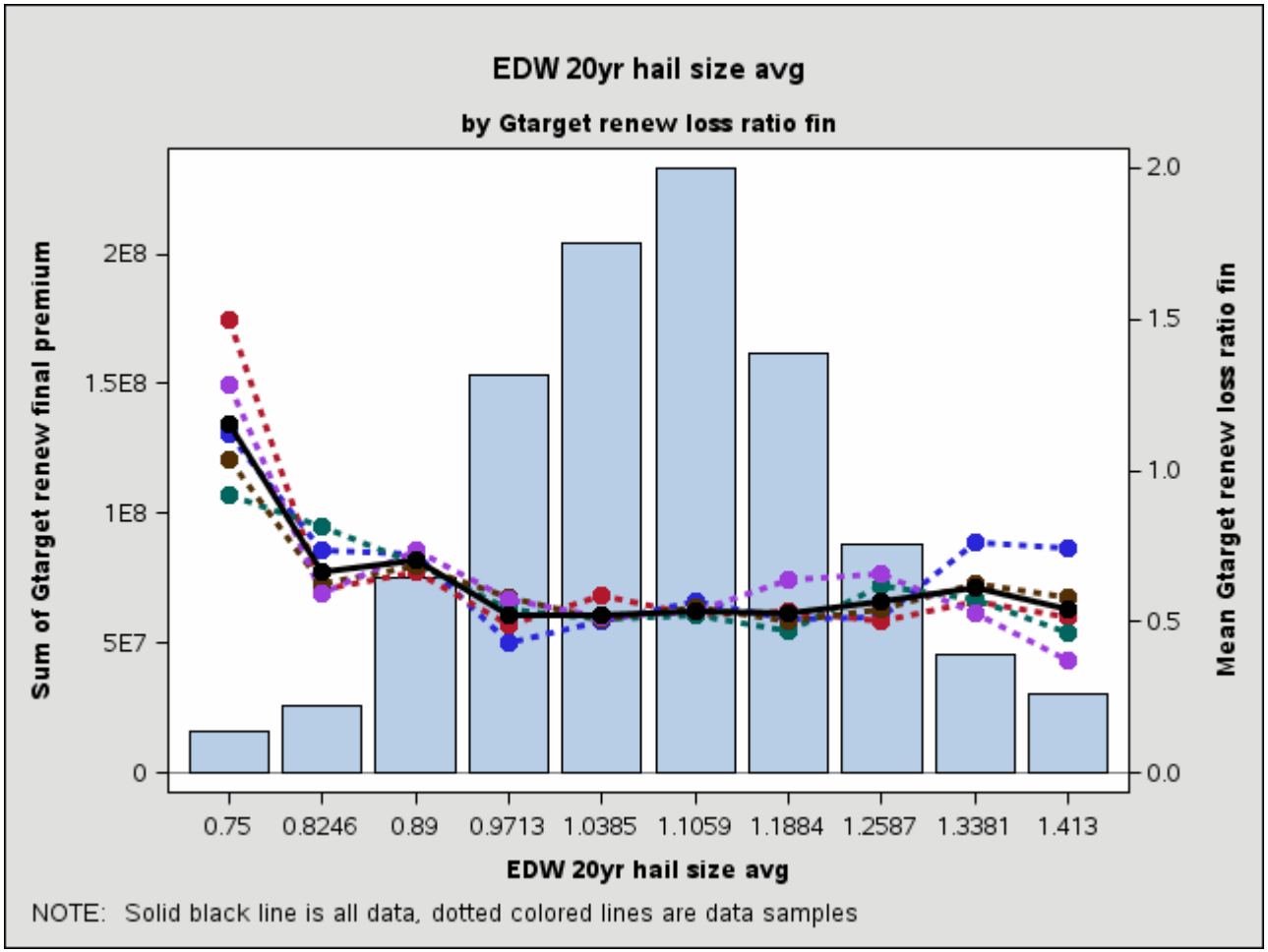
VALEN

Number of Attorney's per County



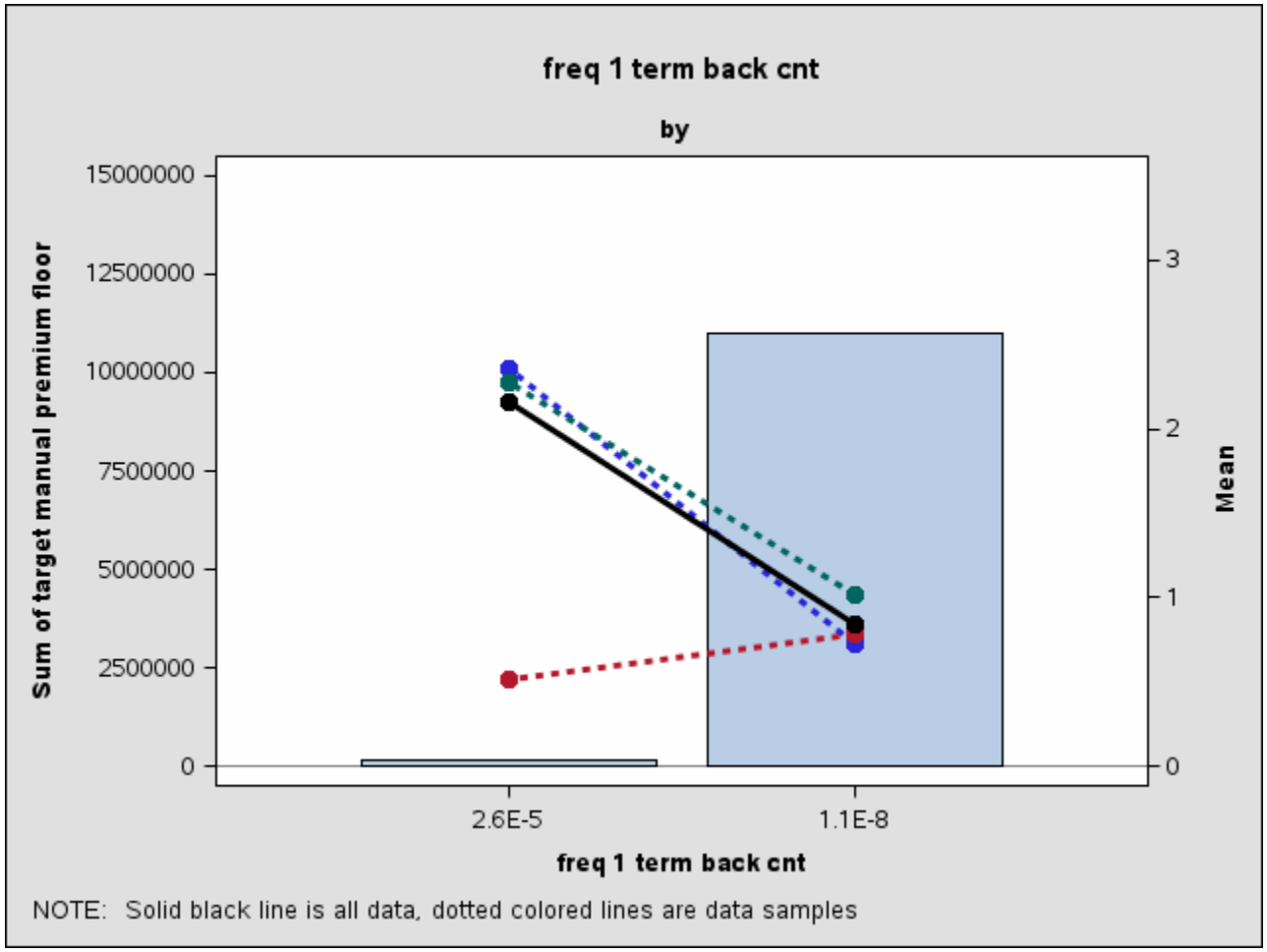
VALEN

20-Year Average Hail Size



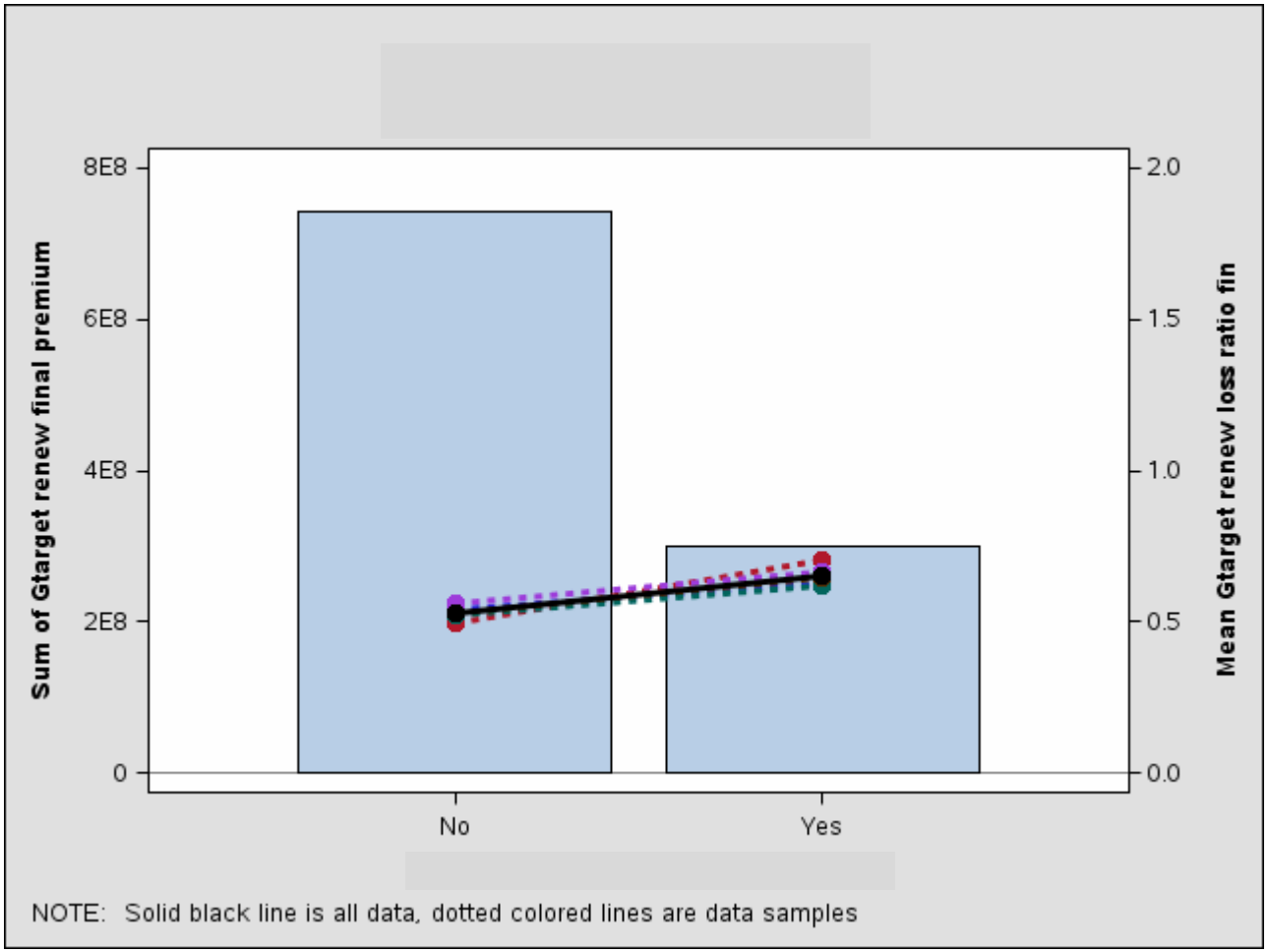
VALEN

Frequency 1 Term Back

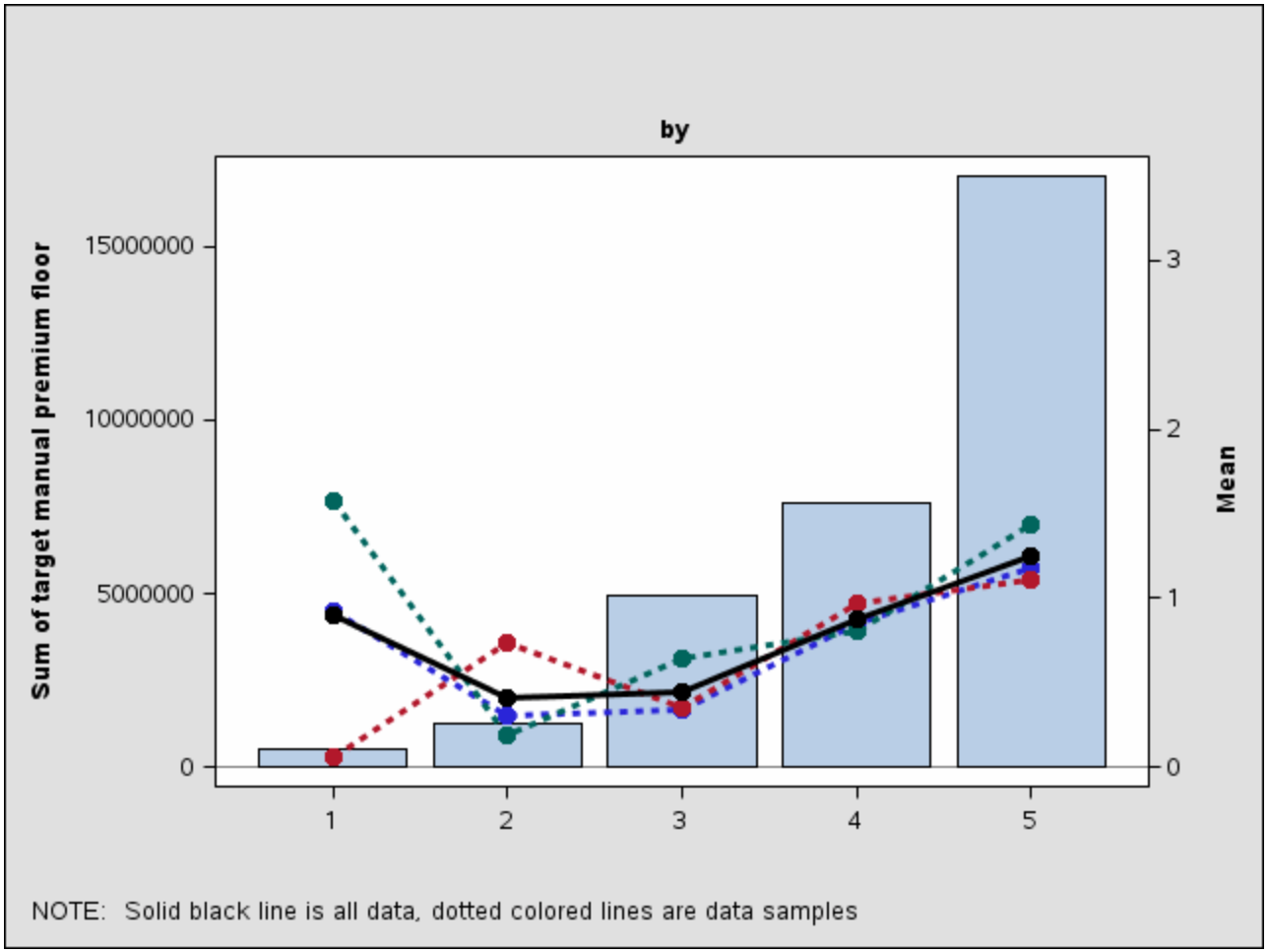


VALEN

Late Reporting 30 days

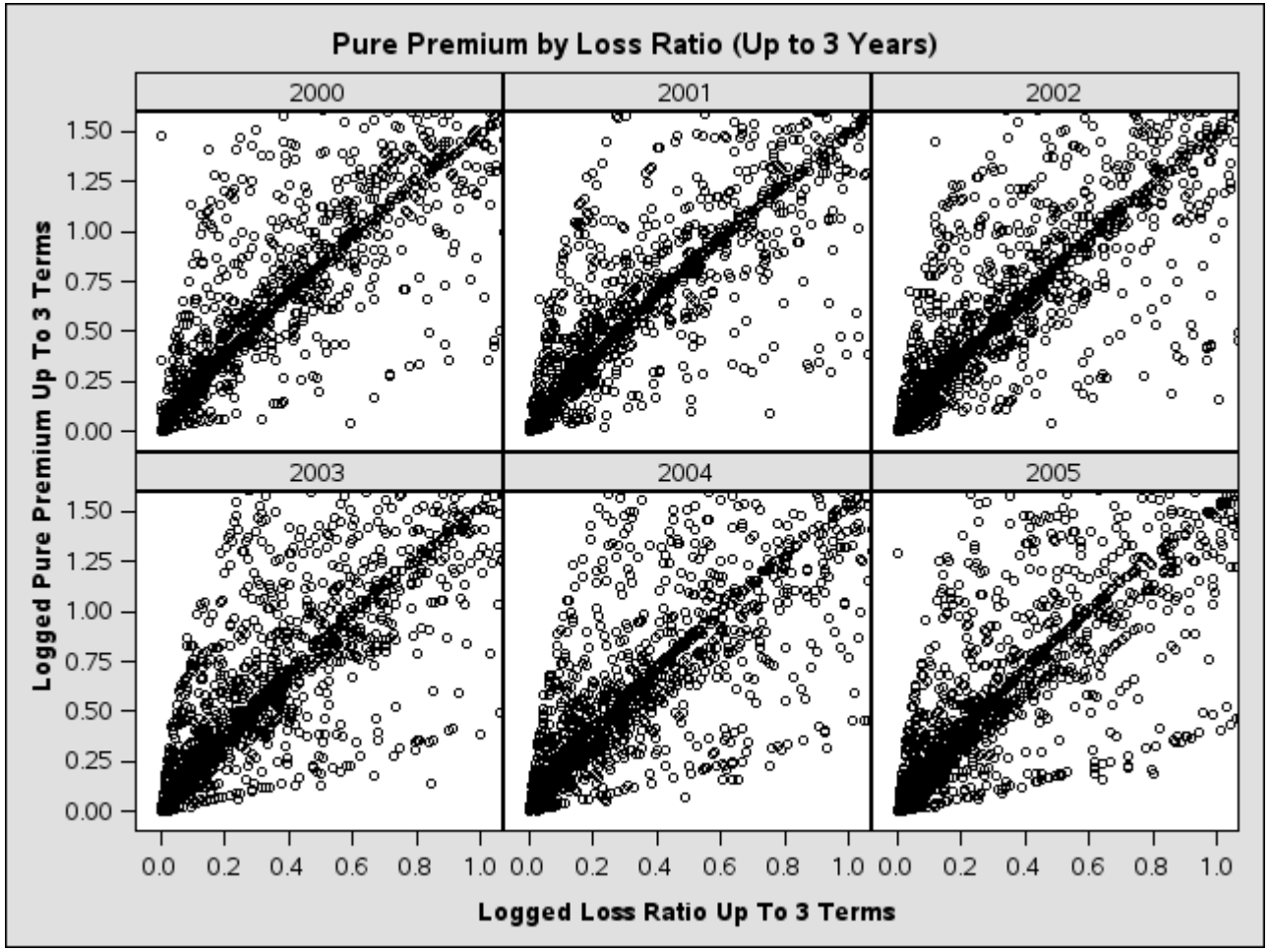


Binned Manual Premium



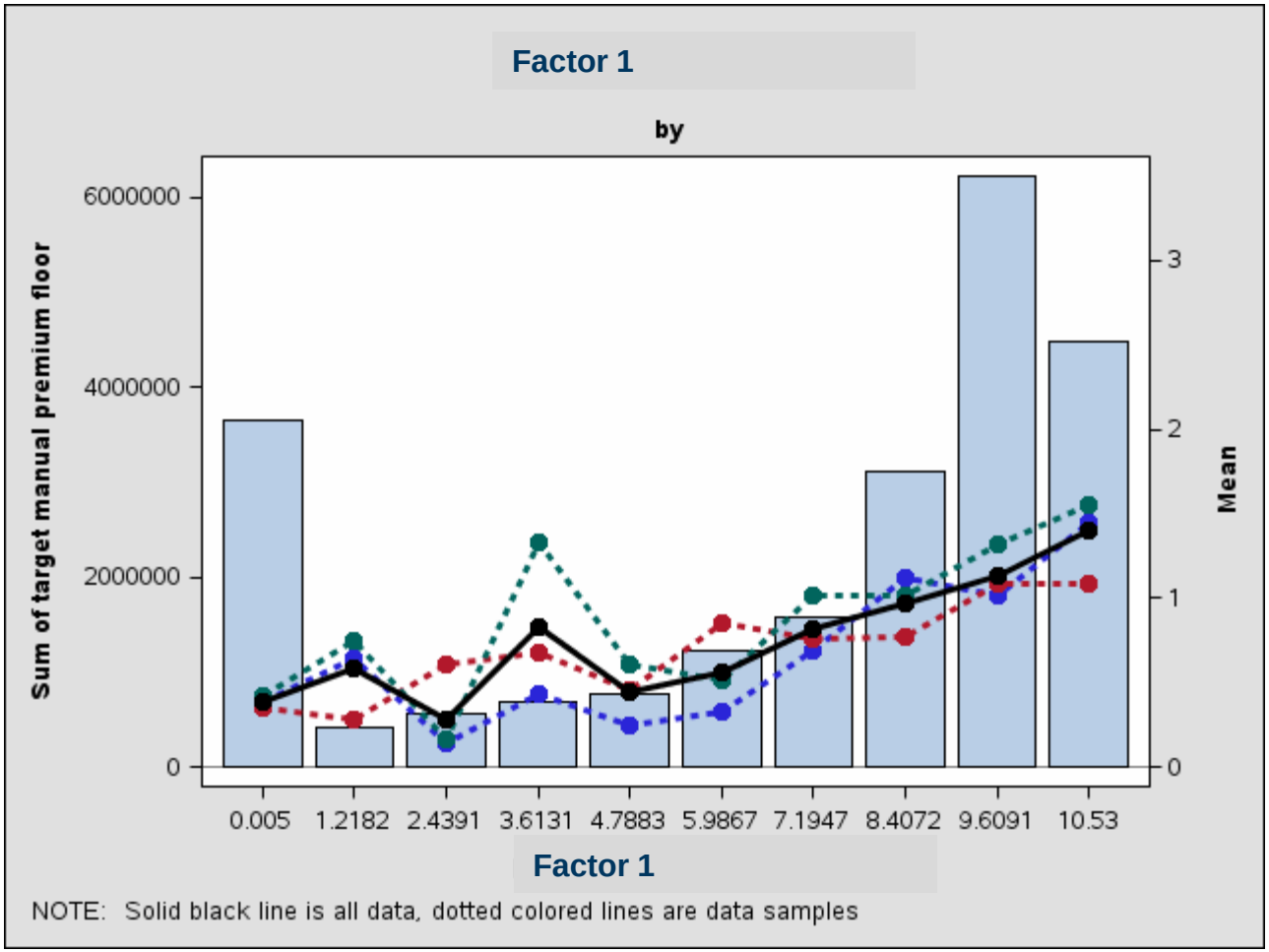
VALEN

Principal Component



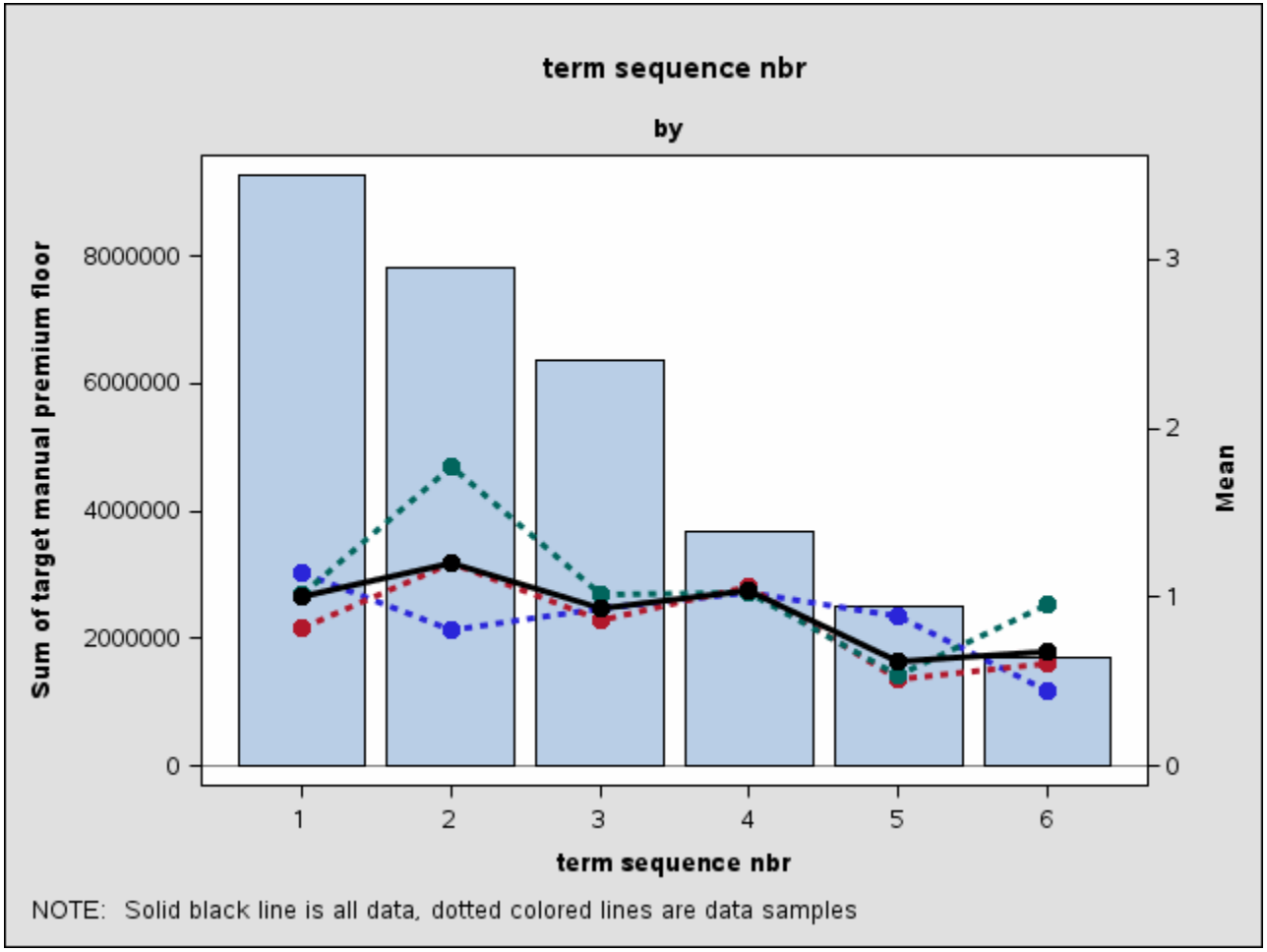
VALEN

Principal Component



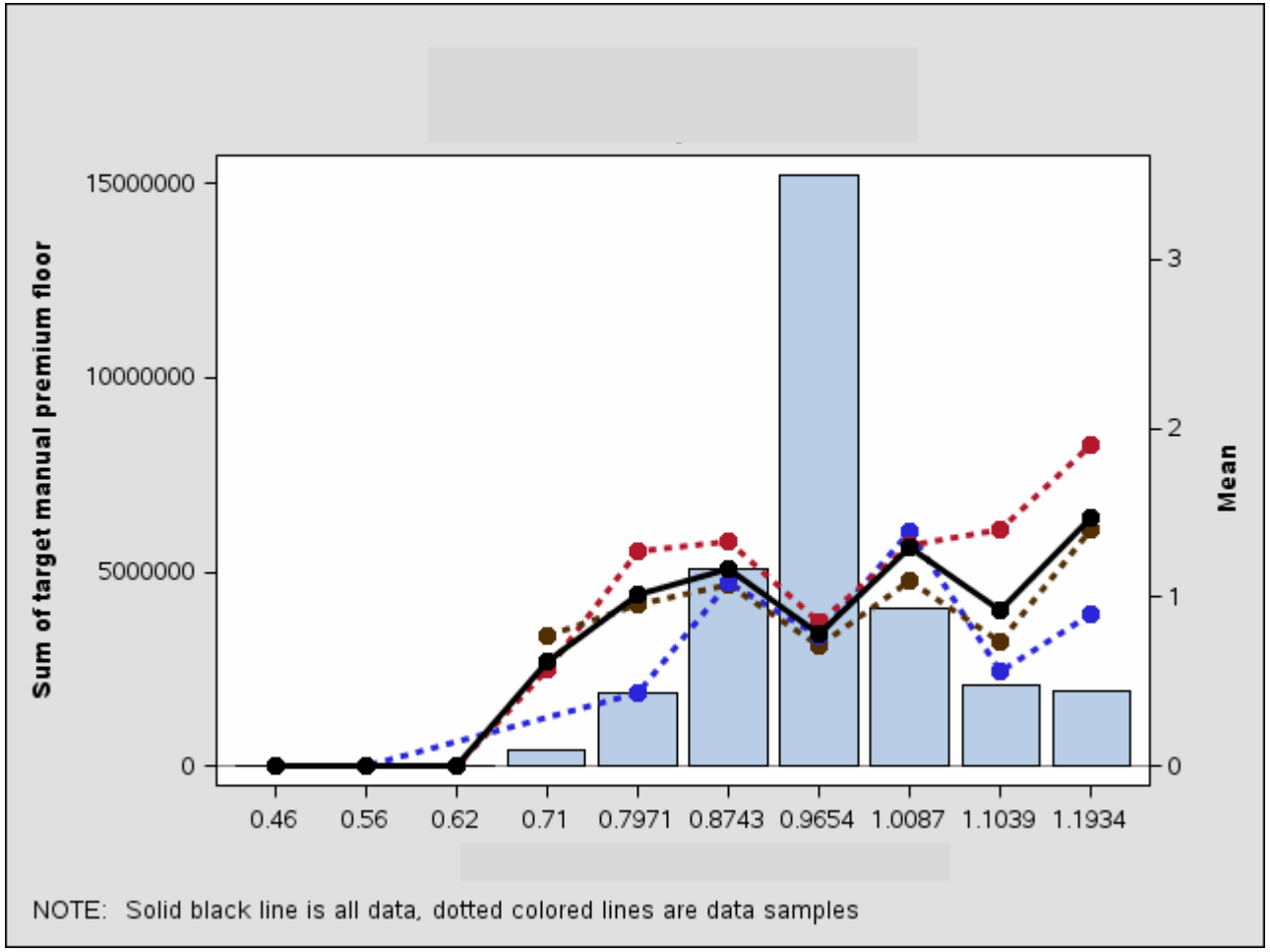
VALEN

Term Sequence



VALEN

Natural Log of Class Code



VALEN

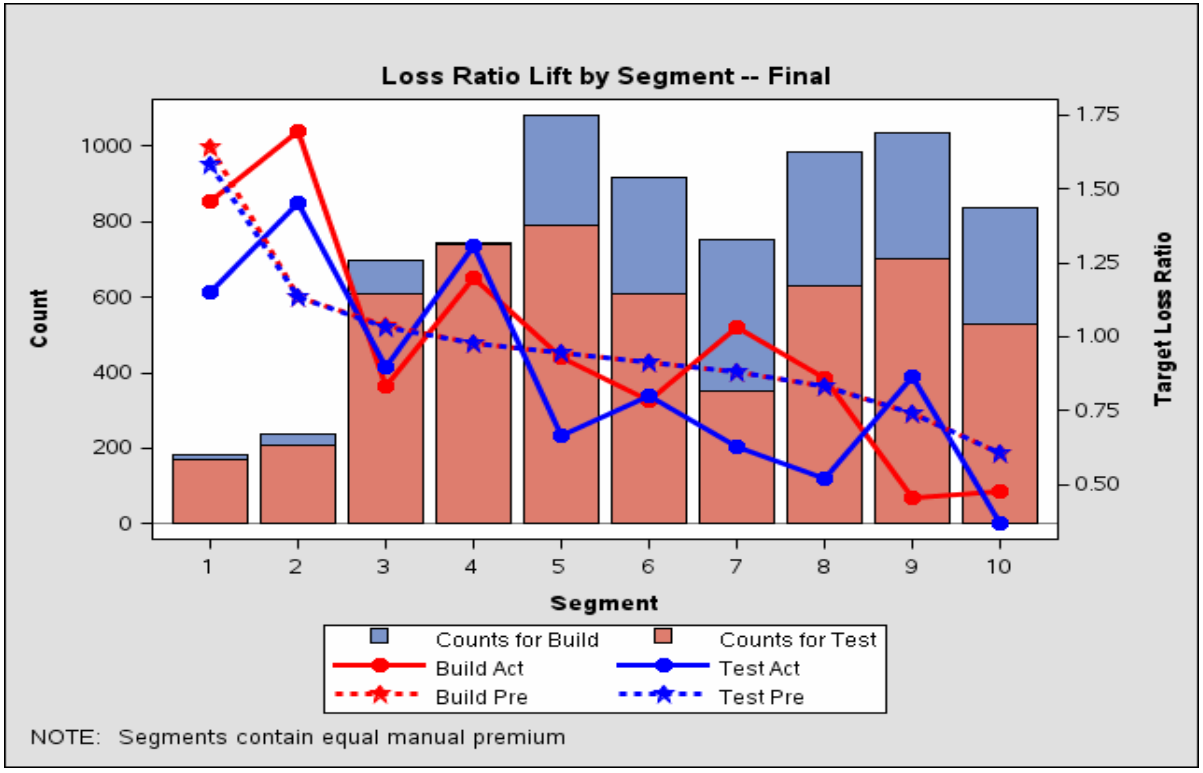
The Modeler Modifies Constraints and Assumptions

- Changed target to loss ratio
- Minimum premium floor
- Implied exposure for minimum premium
- Syncs risk variables to time of score (60 days prior to inception)
- Creates surrogates for risk factors not available in production
- Tries several other transformations or constructions of variables based on underwriter inquiries
- Changes target to score prior to schedule rating and after experience modification
- Changes independent variables to exclude ALAE

What happened???

Upon converting to Loss ratio...

the Predictive Modeler finds an interesting quirk…

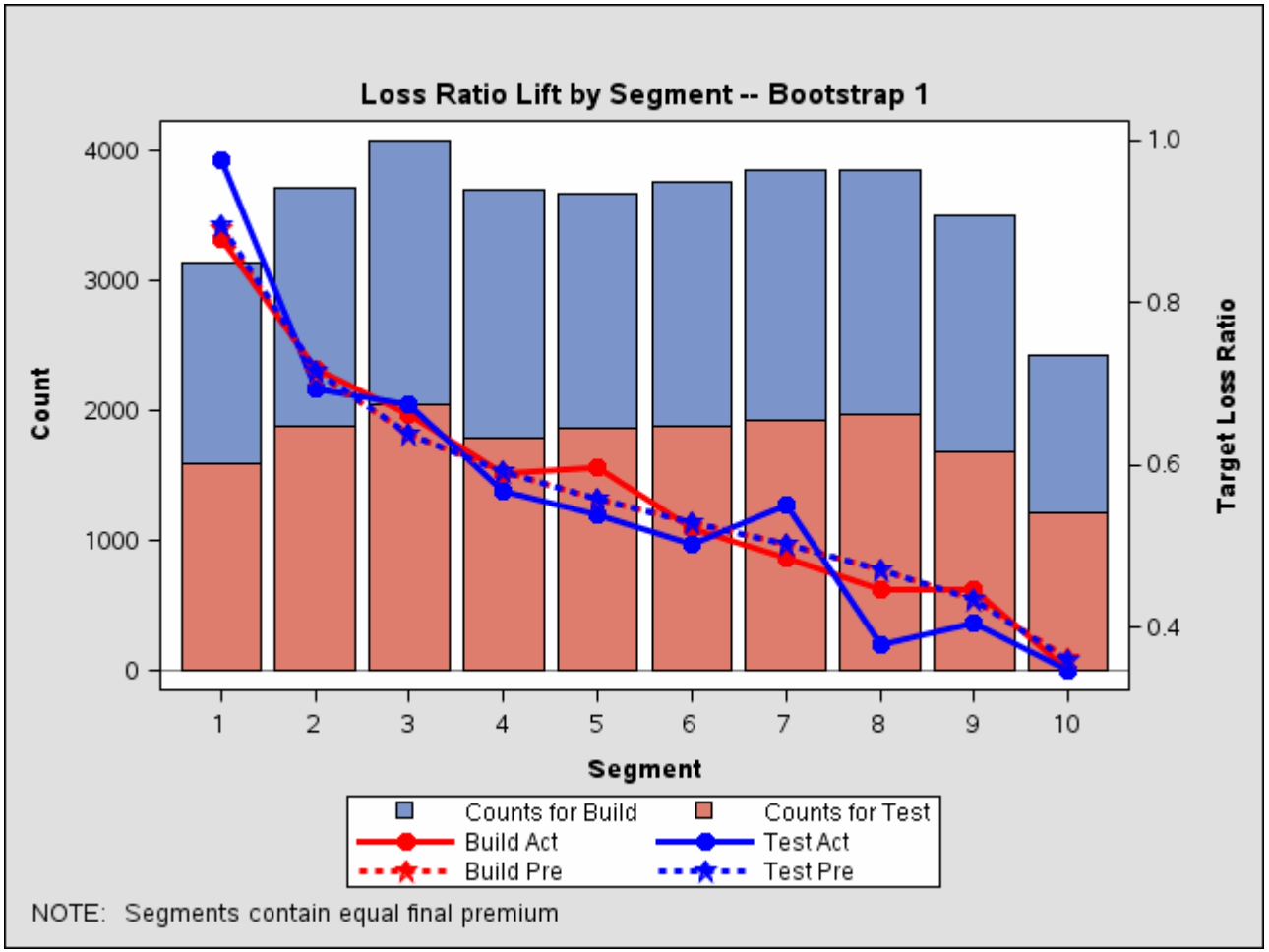


VALEN

After Several Months

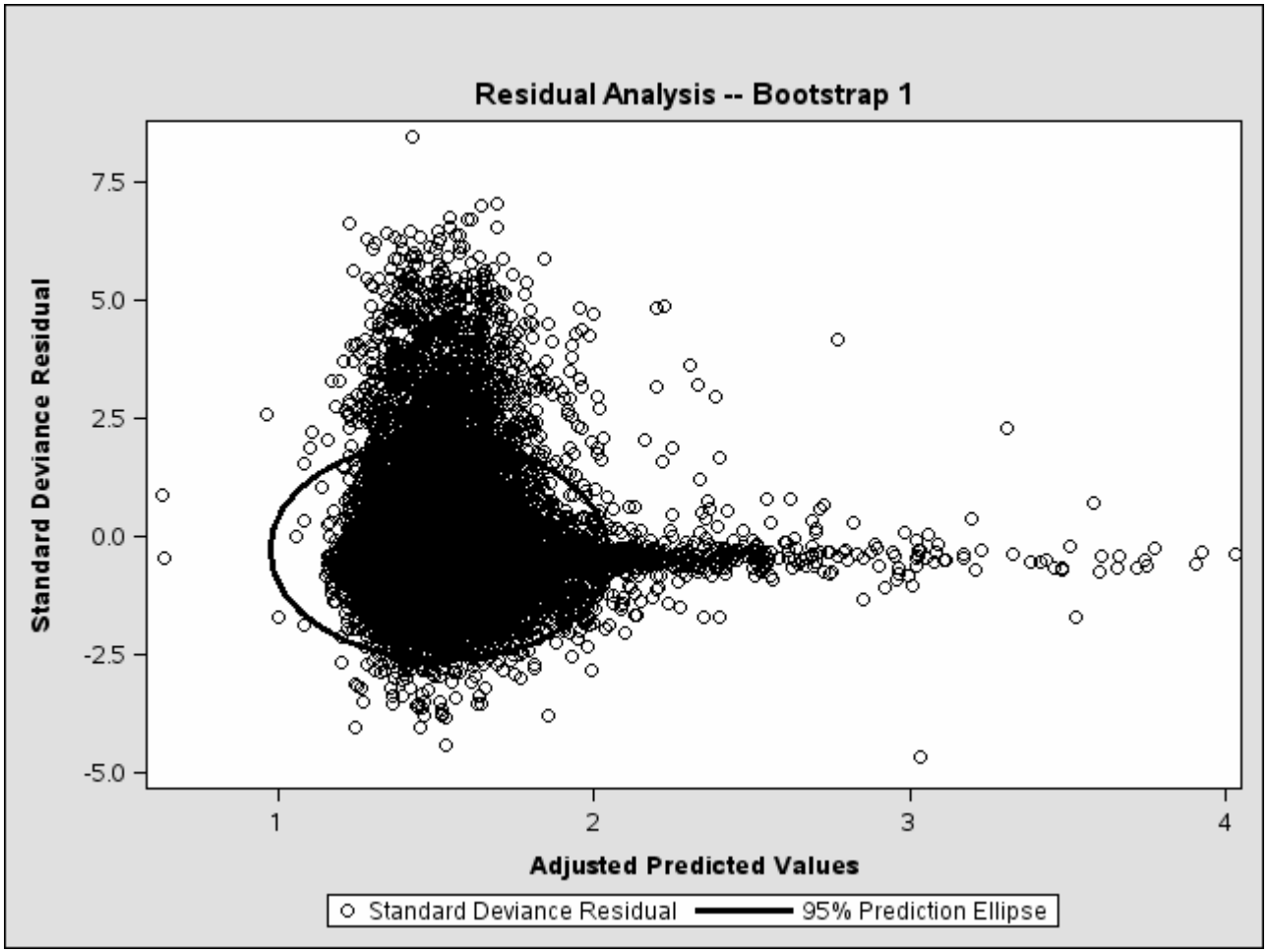
The underwriting department and the actuarial department get on the same page

Candidate Model Lift Curve



VALEN

Standard Deviance by Adjusted Predicted



VALEN

A Short Vignette

The end...

Or

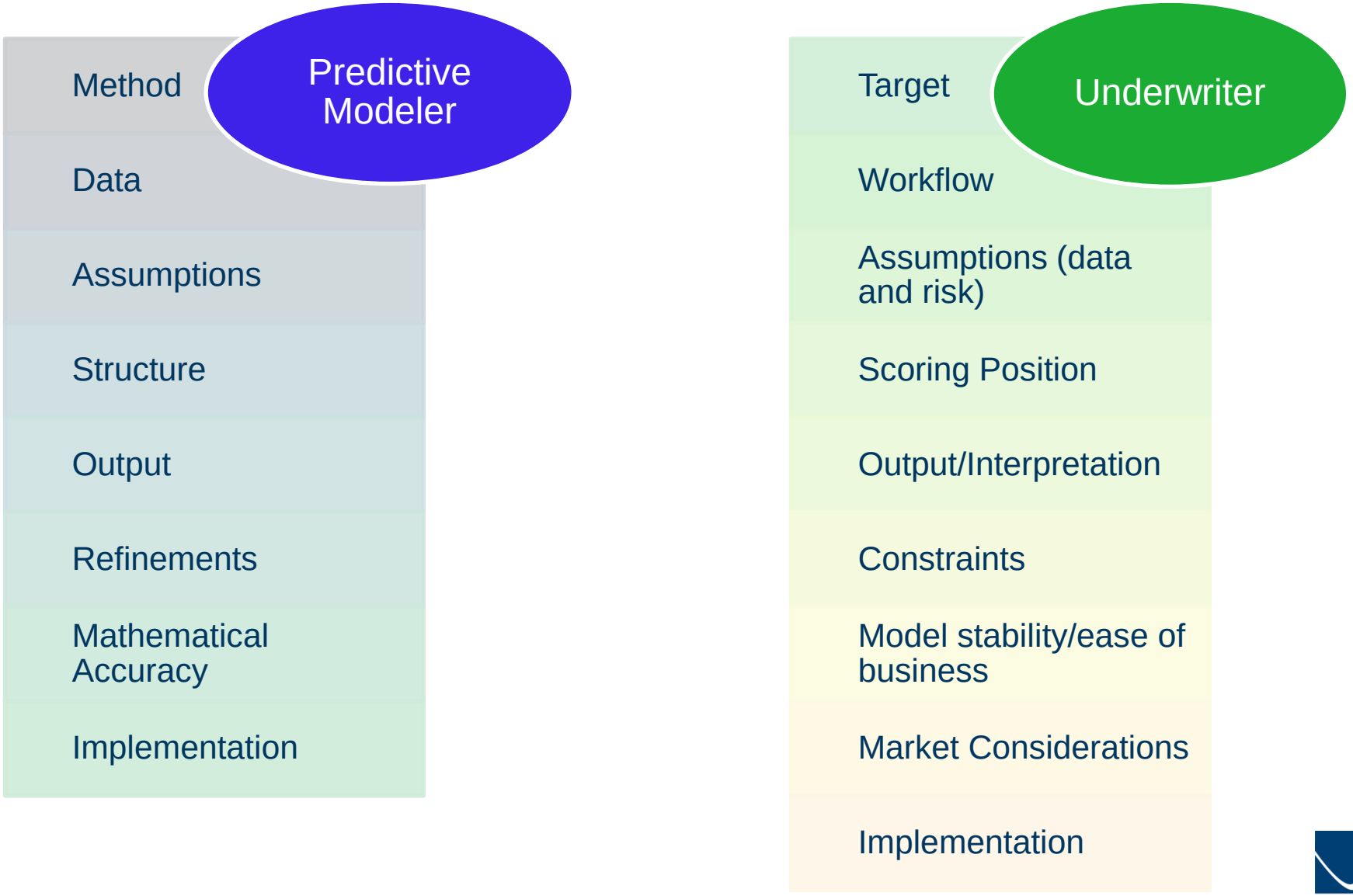
Is it?...

VALEN

The Stage

- Intuition versus Empiricism
- Art versus Math
- Underwriting versus Predictive Modeling/Actuarial
- Collaborative Effort

Perspectives



VALEN

Method

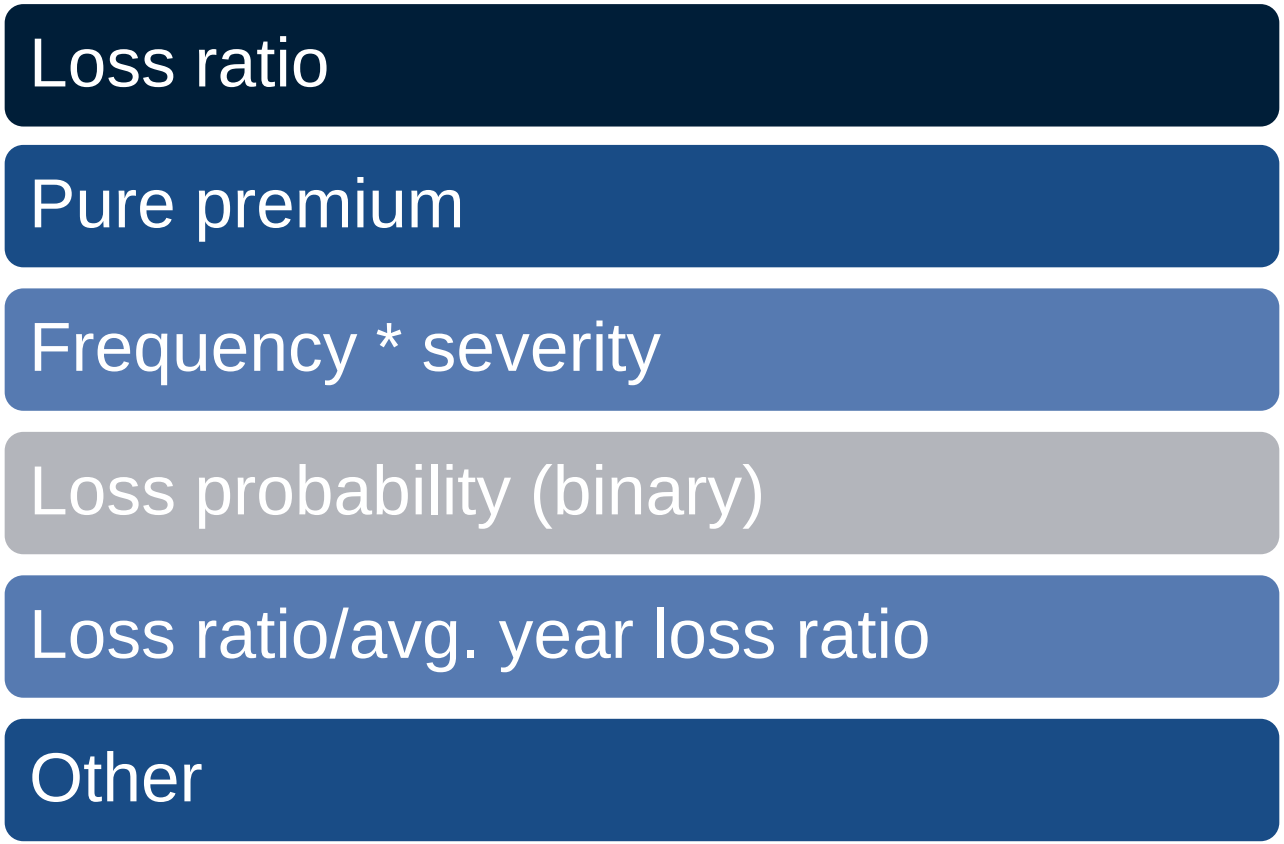
- If you can't explain it, don't expect others to understand it
- For example: generalized linear model (GLM) is …
  - Layman's explanation…
  - Statistician's explanation

$$E[Y_i] = \mu_i = g^{-1}\left(\sum X_{ij}\,\beta_j + \xi_i\right)$$

$$Var[Y_i] = \phi V(\mu_i)/\omega_i$$

…Huh?

Target



- What is the intent of the model?
- What other business rules are applied in the underwriting process
- How can the model build synergy with current best practices without being “constrained to fail”

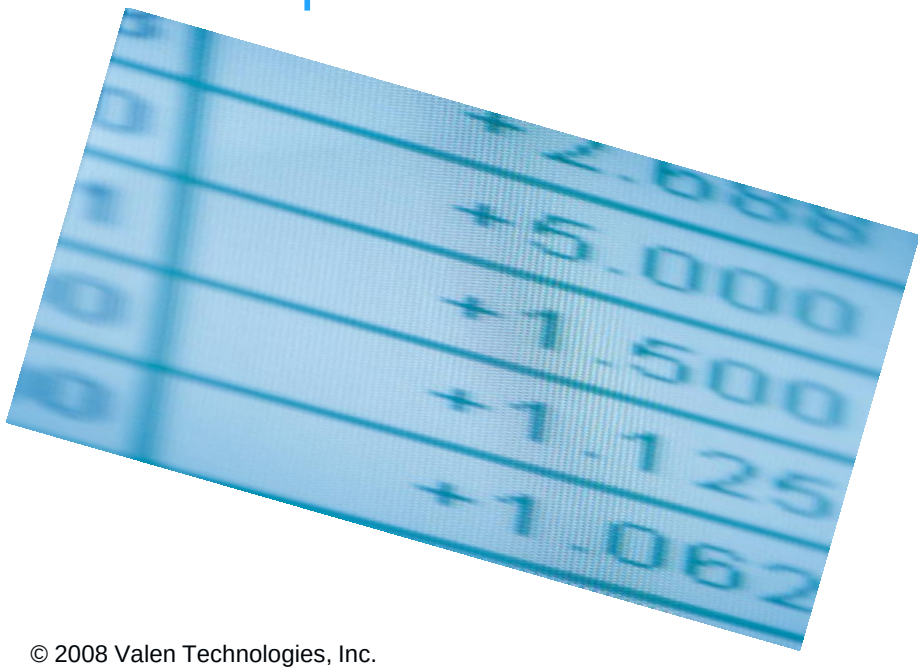
VALEN

Workflow

- Where will the scoring engine live?
- At what point during the logical process will the model be used?
  - When will the output values be rendered to make a risk judgment?
- What data is available at time of scoring?
- What are the underwriters going to do before and after they render the outputs?
  - What judgments have they already made?
  - What judgments will they make post scoring?
- How does the workflow constrain the data or the scoring parameters?

Data

- How much (e.g. exposure, claim counts)
- States, years, class codes
- Explanatory variables
  - Huge opportunity to build rapport
- Data adjustments/transformations
- Data splits (e.g. hold-out samples for model validation)



VALEN

Assumptions

- Model assumptions
  - Link function
    - log ==> multiplicative
    - Identity ==> additive
    - Logit ==> probability
  - Distribution assumptions
    - Frequency (Poisson)
    - Severity (gamma)
    - Combine Frequency/Severity model results
    - Loss Costs (Tweedie)
    - Probability of loss (Binomial)
- Model assumptions
  - Loss ratio goal
  - Market penetration goal
  - Tier cuts based on production and market assumptions
  - Sync market forces with model tier cuts and constraints
  - Model uses production available data
  - Independent variables must pass test of reasonableness

VALEN

Model structure

Clearly define the variables in your models and the model structure.

- Equation:

$$E[\text{Loss Ratio}] = g^{-1}(\alpha + \beta.\text{freq 2 term avg} + \gamma.\text{emod amount} + \ldots)$$

- Table:

Variable	Intuitive	Production Available	Data Support
Freq 2T	Yes	Yes	Yes
Emod Amt	Yes or No	Maybe	Yes
Class bin	Yes	Yes	Maybe
…			

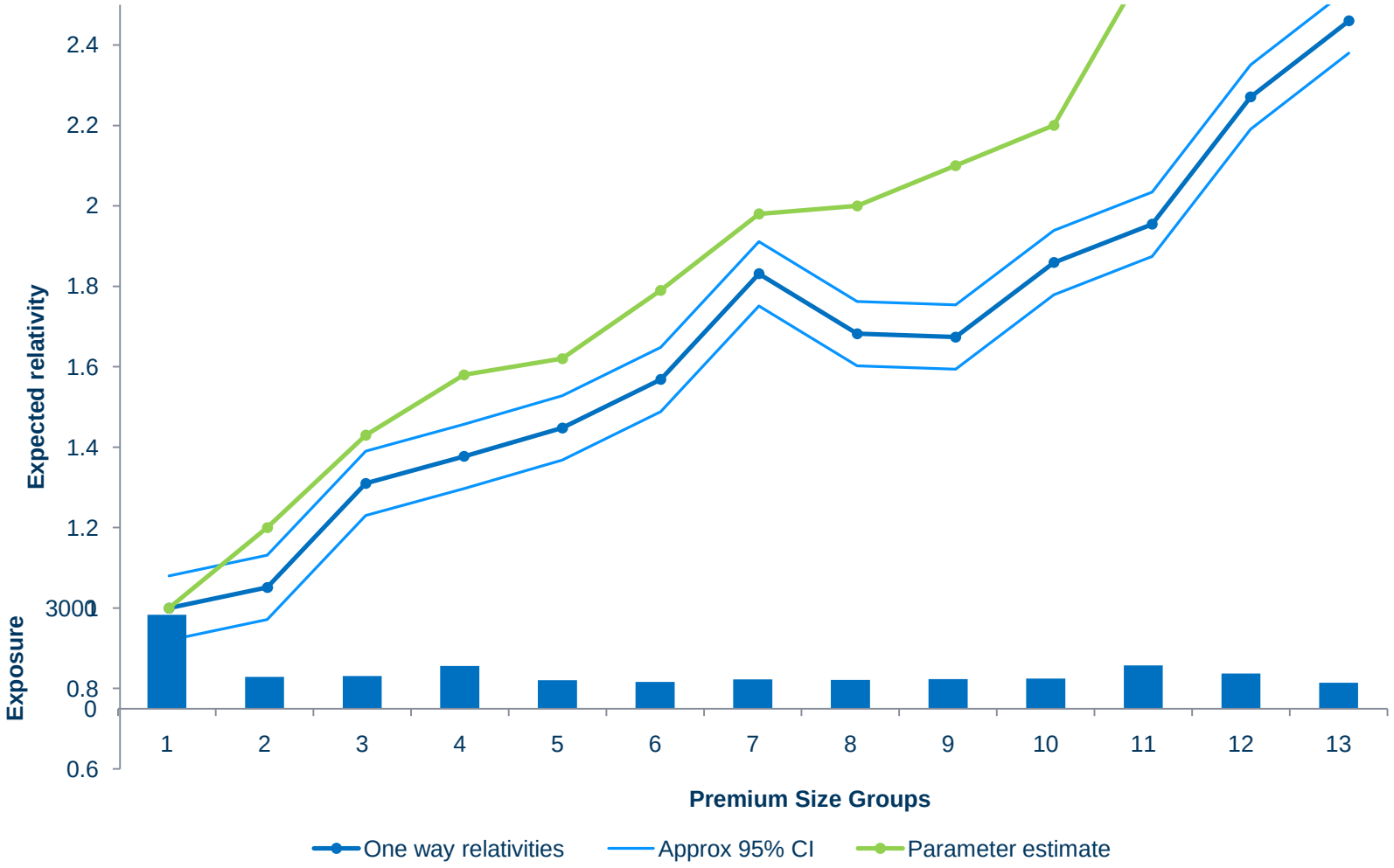
Data/Variable structure

Explain how specific variables are defined/structured.

- Continuous
- Categorical
- Interactions
- Restrictions
- Constructed
- Transformations

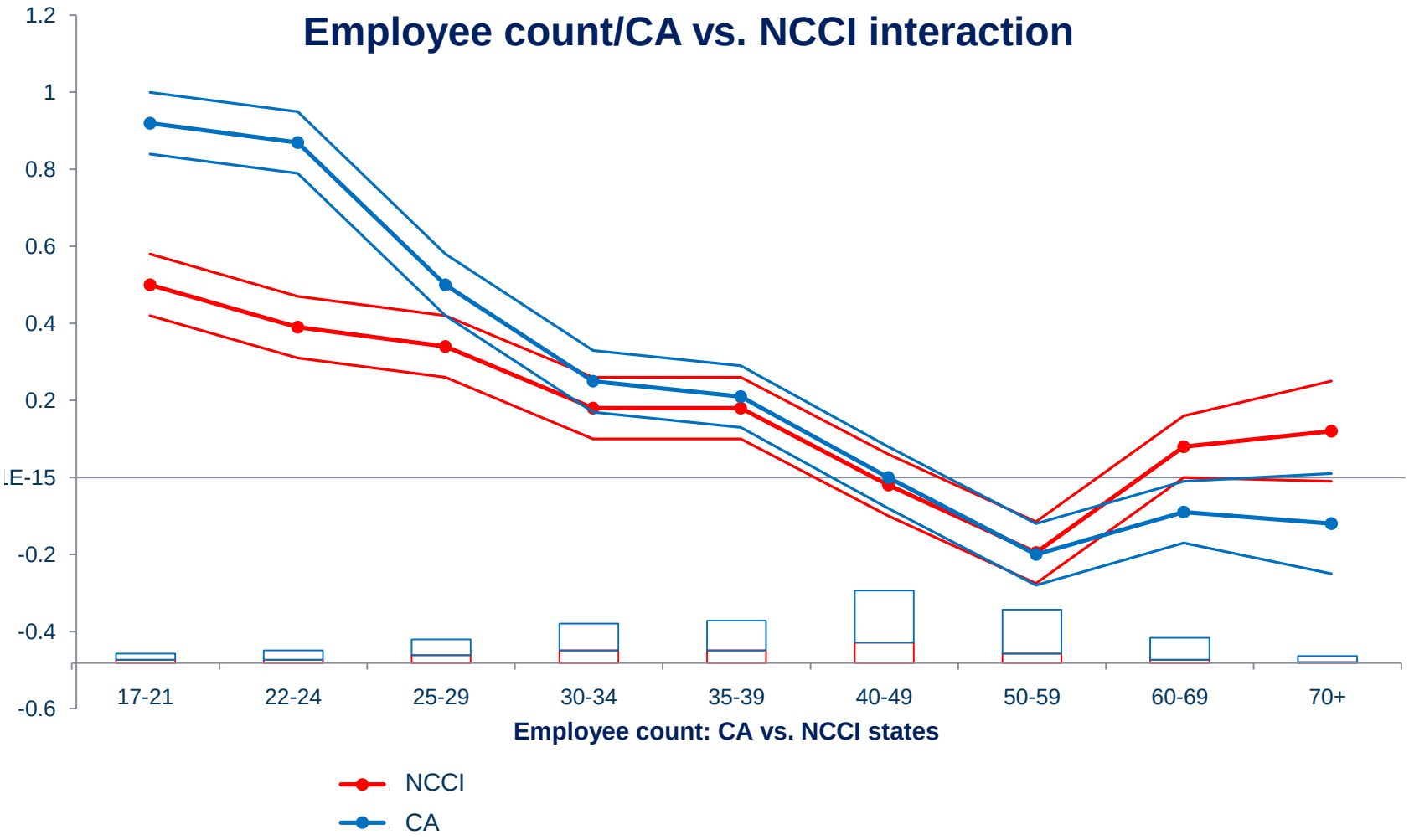
Reviewing univariate analysis with underwriting is key!

Model output

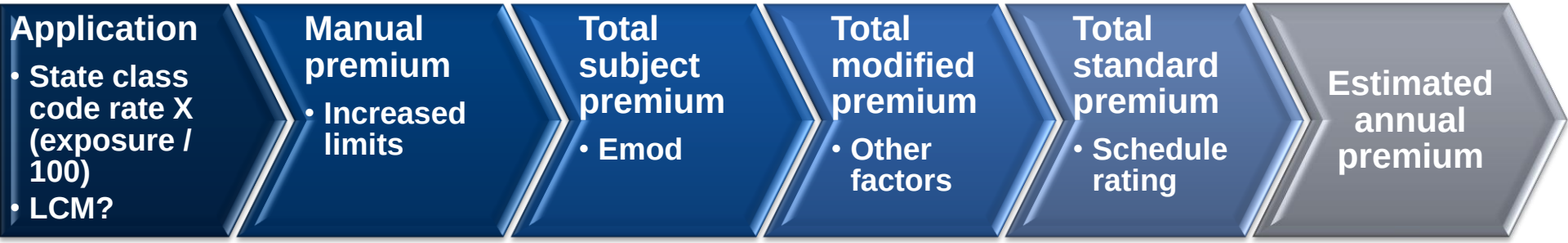


VALEN

GLM output



Scoring Position



VALEN

Production Outputs and Interpretation

- Loss ratio
- Pure premium
- Expected Loss
- Avg. bin loss ratio
- Score?
- Risk grade

How will underwriters interpret results. What pitfalls can be avoided upfront?

Refinements

- Based on discussions with underwriting and business partners
- Likely changes arising from:
  - Regulatory
  - Acceptability
  - Explainability
  - Business knowledge
  - Business rules/philosophy
  - Availability

Constraints

- Rating plan limitations
  - Minimum premiums
  - Pool business
  - Per capita rated policies
  - Schedule rating caps
  - Discontinued classes/risk types
  - Business rules
    - Are they available in the data?

Model Stability vs. Ease of Business

- 3 year historical loss and pure premium variables
- Frequency variables
- Severity variables
- Exposure
- Historical premium
- Employee count

How much information is too much? This question must be asked at the outset.

Statistical Accuracy

- Parameter point estimate
- Standard errors
- Policy level prediction
- Credibility of policy level results?
- Bin Framework
- New paradigm shift for underwriting
  - Policy predictions by virtue of its inclusion to a bin with policies of like charateristics

A (Very) Short Vignette (Revisited)

Collaboration is critical...

the Predictive Modeler and
the Underwriter

Learn from each other...

VALEN

Questions???

Marrying Underwriter Intuition & Predictive Modeling - A Workers' Compensation Perspective

CAS RPM Seminar

Las Vegas Mar 10, 2008

Presented by:
Gaétan Veilleux, FCAS
Matt Frazier, Client Lead

VALEN technologies