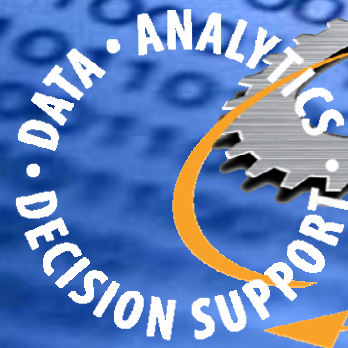


2008 PREDICTIVE MODELING SEMINAR

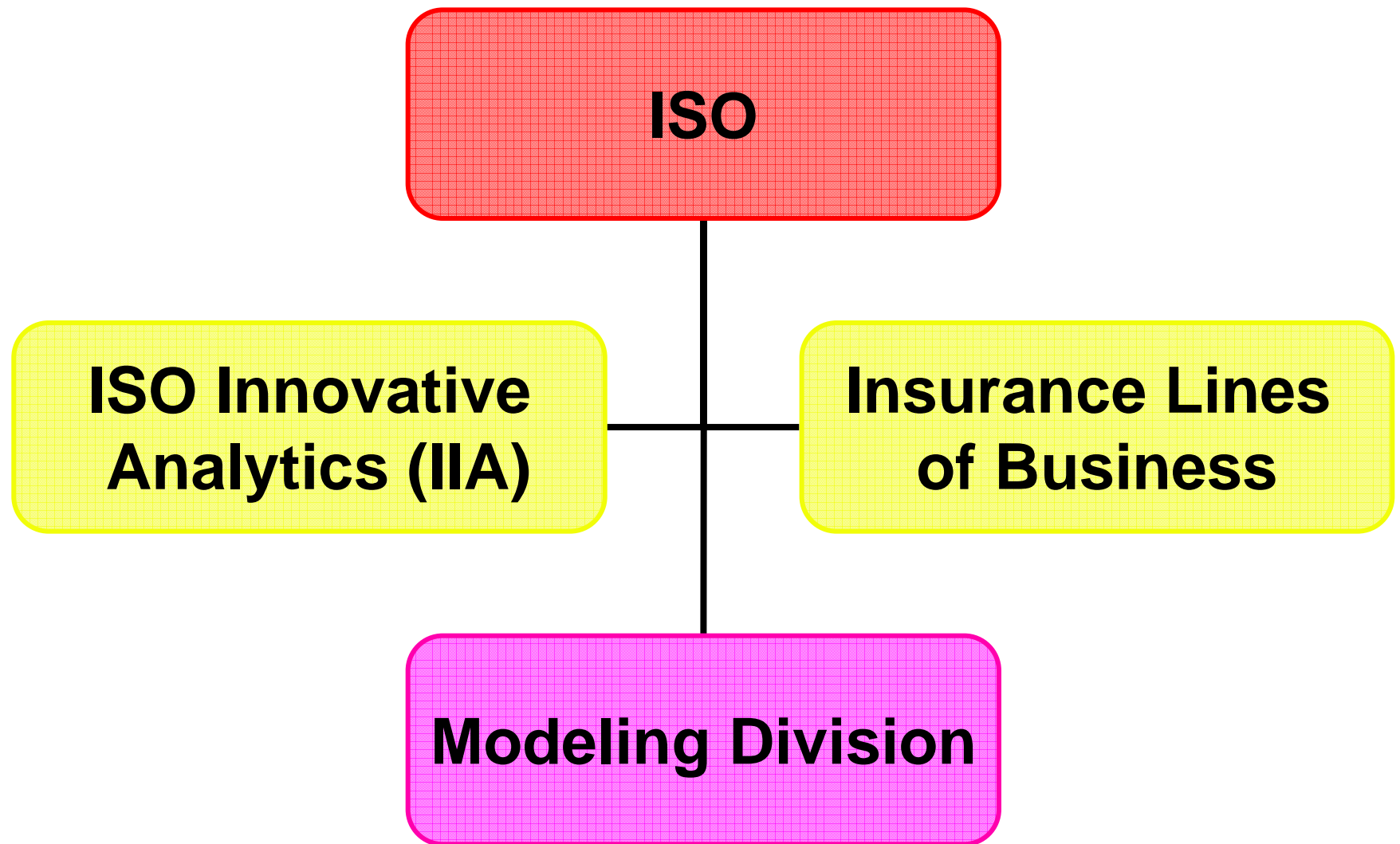
PROJECT MANAGEMENT FOR PREDICTIVE MODELS



ANALYTICS
DATA • SUPPORT

JOHN BALDAN, ISO

Where Does Modeling Fit In?



Modeling Division – Client-Facing Goals

- **Work with IIA to develop predictive models for the major lines of insurance, using insurer data.**
 - Personal Auto
 - Homeowners
 - Commercial Lines

Modeling Division – Internal Company Goals

- **Make greater use of predictive modeling techniques within areas of traditional ISO ratemaking:**
 - Loss costs
 - Classifications
- **Disseminate modeling knowledge and expertise to pricing actuaries via training, modeling projects.**

Developing Resources

- **Staff Knowledge**

- Good sources re: GLMs:

- “A Practitioner’s Guide to Generalized Linear Models” – Anderson, Feldblum, et. al.
 - “Generalized Linear Models” – McCullagh and Nelder
 - Generalized Linear Models for Insurance Data – de Jong and Heller
 - “A Systematic Relationship Between Minimum Bias and Generalized Linear Models” – Mildenhall

Developing Resources

- **Software**

- PC SAS
 - Interactive aspect
 - (Major) Advantage: more efficient, flexibility coding
 - (Minor) Disadvantage: documentation
 - Graphics
 - Avoid divisional chargebacks!
- R (used for MARS, for example)
 - Incredibly flexible, since object oriented
 - Special purpose modules available on Web
 - Widely adopted in statistical, academic world
 - Cost
- Other software packages, as needed

Be vigilant !



Building Predictive Models

- **Know your data**

- Get data in common format, at level of individual risk.
- Interface with insurer to:
 - Understand their database structure
 - Examine univariate distributions to clarify the meanings of data elements, unclear codes and missing values.

Deploying the Model Output

- **Not all model variables should be used to calculate the deployed model output. Decide which ones should.**
- **Considerations:**
 - Keep user input simple (e.g., deductible)
 - Keep some rating variables unchanged (specify these as offsets)
 - Consider variable retrieval time.
 - Consider model type (marketing vs. rating models).

Specifying the form of the Model Output

- **Frequency/severity models vs. pure premium models**
 - Risk-level data vs. aggregate data
- **Pure premium models vs. loss ratio models**
- **Relativize results to a specified base risk?**

Measuring Model Effectiveness

- **Measures of Lift**

- Decile plot
- Gini index

What these share is an ordering of risks, against which experience is evaluated.

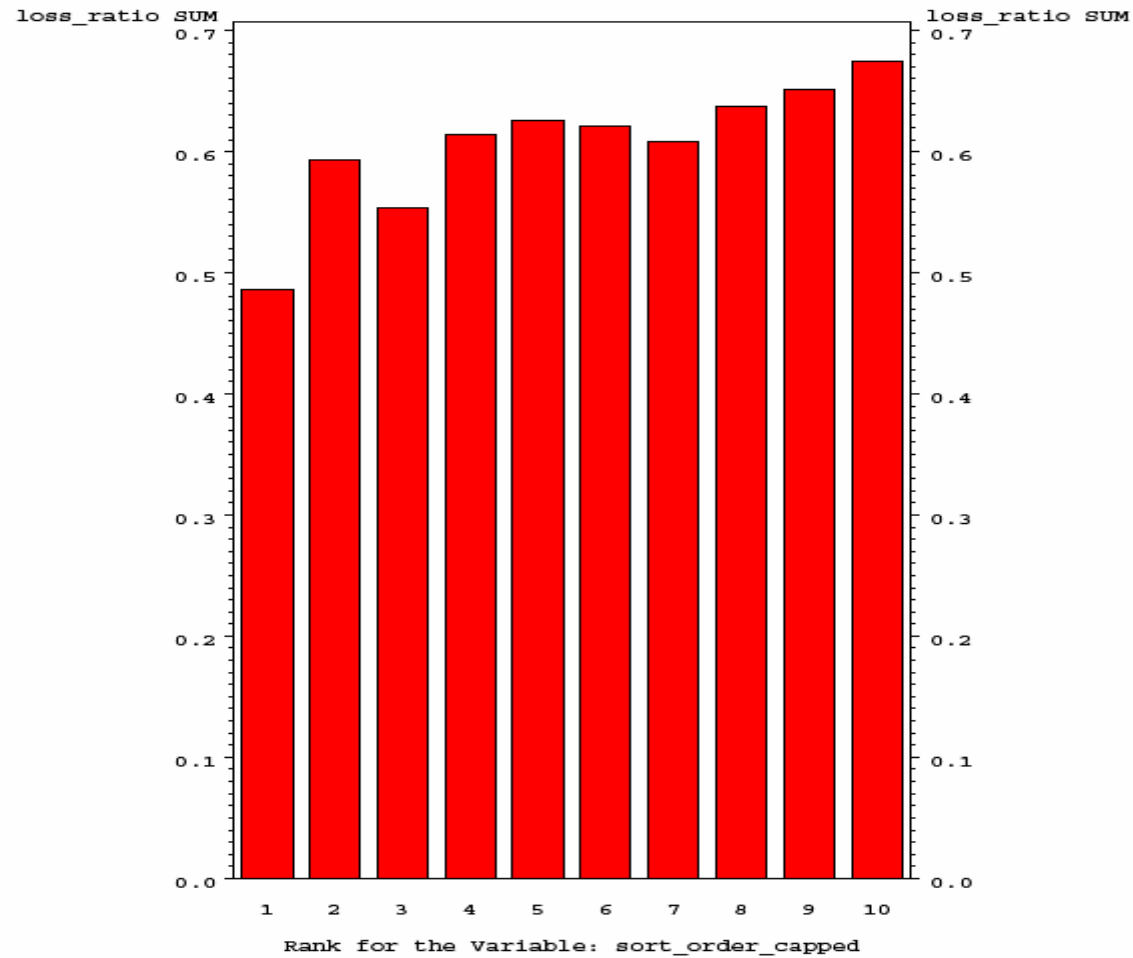
- **Lift is measured against the rating system currently in place. Define sort order by:**

- Modeled loss cost to current loss cost; or
- Modeled loss cost to average modeled loss cost for risks currently rated identically (for example, in a location model, risks in same territory)

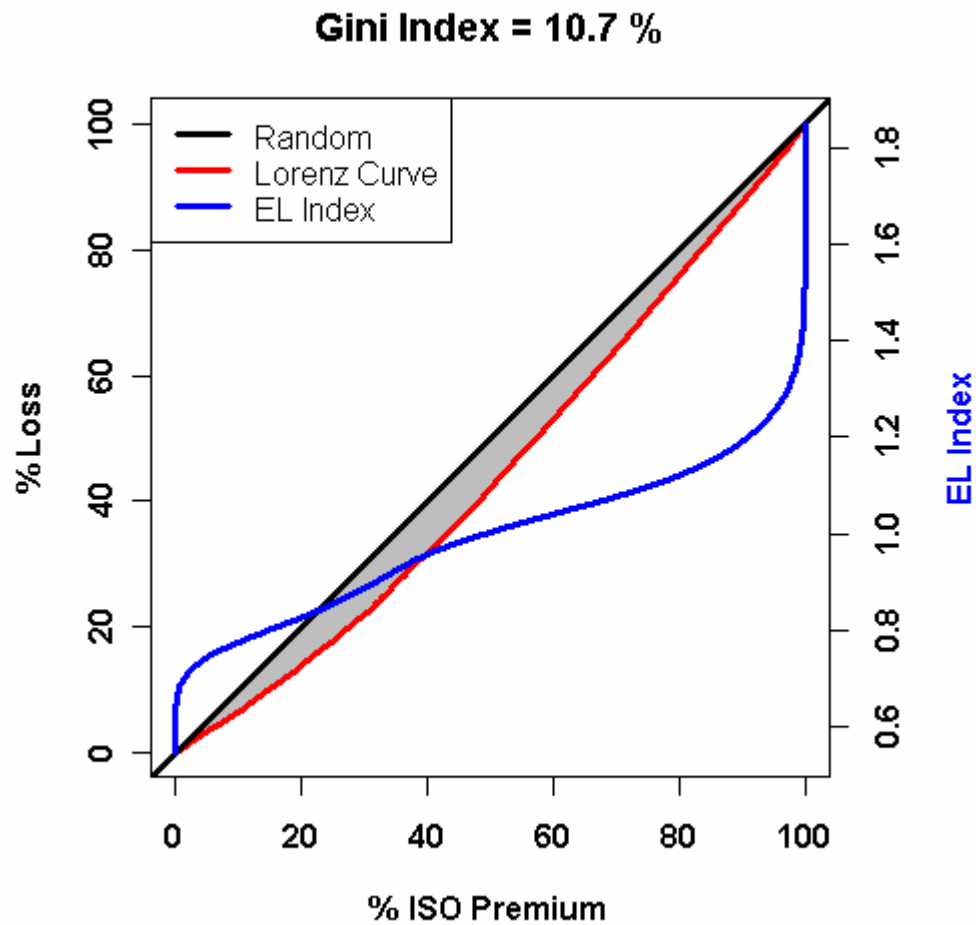
Measuring Lift – Decile Plot

Gini Index for Score `sort_order_capped`

Base = ISO_IC_with_symbol
SAS Dataset = johnb.PD_lift_liabsymbol_data

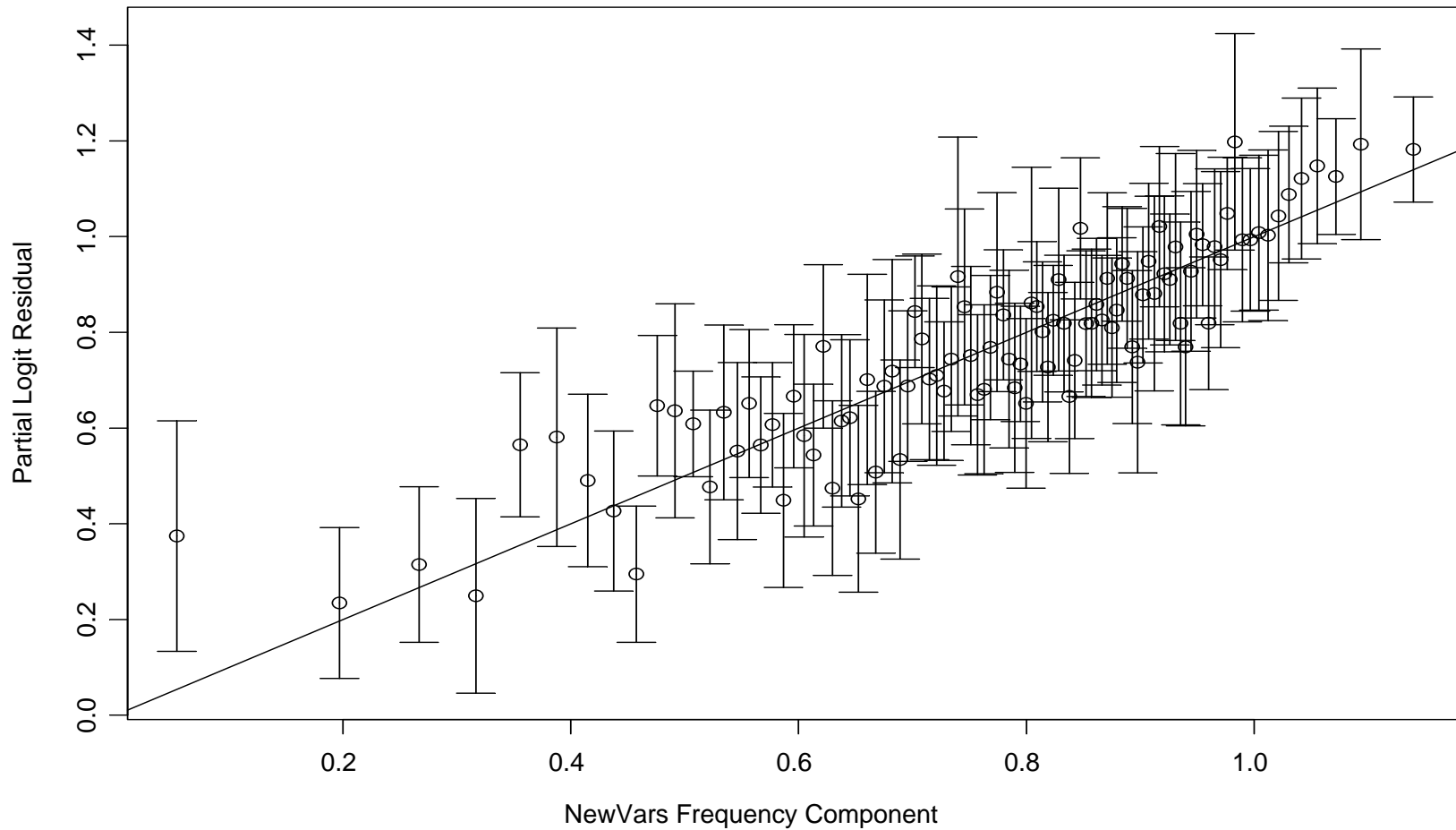


Measuring Lift – Gini Index



Model Diagnostics

BI Model Diagnostics : Partial Residuals



Models and Regulation

- **Establish dialogue between modelers and legal/regulatory experts**
- **Modeling ground rules**
 - Restricted modeling variables
 - Model variable creation techniques
 - Interpretability of final model form
- **Model Smoothing**
- **Reason codes**
- **Diagnostics**

Updating the Model

- **Know your product life cycle**
- **Input updates**
 - Insurance data
 - Third-party data
- **Output/Model updates**
 - Recalculation
 - Re-estimation
 - Rebuilding