And The Winner Is…? How to Pick a Better Model

Model Lift – 2015 CAS RPM Seminar

Verisk Analytics

SERVE | ADD VALUE | INNOVATE

Motivation

- Models that appear to be strong may have weaknesses
  - Fit may not be good
  - Model may be overfit
  - Wrong distribution may have been chosen
  - Results may not be stable across data subsets or over time
  - Results may be highly influenced by several records
  - Model may underperform the status quo

Some Models Used by Actuaries

- Linear regression
- Exponential regression
- Logistic regression
- Minimum bias procedures
- Generalized linear models
- Classification and regression trees
- Clustering procedures

SERVE | ADD VALUE | INNOVATE

Understanding & Validating a Model

- Model Lift
  - How well does the model differentiate between best and worst risks?
  - Does the model help prevent adverse selection?
  - Is the model better than the current rating plan?
- Goodness of Fit
  - What kind of model statistics are available, and how do you interpret them?
  - What kind of residual plots should you consider, and how do you interpret them?
  - What are some considerations regarding actual versus predicted plots?
- Internal Stability
  - How well does the model perform on other data?
  - How will the model perform over time?
  - How reliable are the model's parameter estimates?

4

SERVE | ADD VALUE | INNOVATE

Model Lift

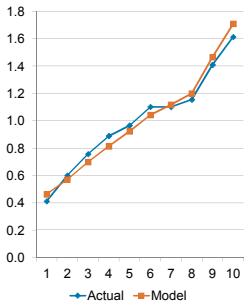
- Ability to differentiate between low and high cost policyholders
  - Sometimes called the "economic value" of the model
- Some tools for measuring and illustrating model lift
  - Simple Quantile plots
  - Double lift charts
  - Gini index
  - Loss ratio charts

5

SERVE | ADD VALUE | INNOVATE

Model Lift – Simple Quantile Plots

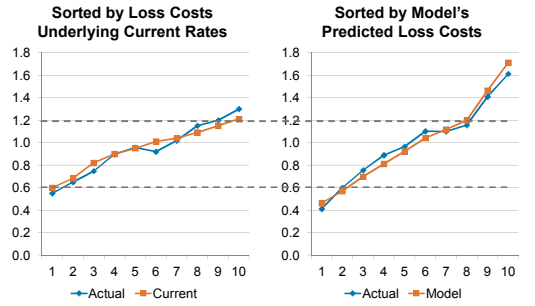
- Creating a quantile plot
  - Use holdout sample.
  - Sort data based on predicted value (frequency, severity, loss cost).
  - Subdivide sorted data into quantiles (quartiles, quintiles, deciles) with equal weight (exposure, claim count).
  - Calculate average actual value and predicted value for each quantile and index to overall average.



6

SERVE | ADD VALUE | INNOVATE

Model Lift – Simple Quantile Plots

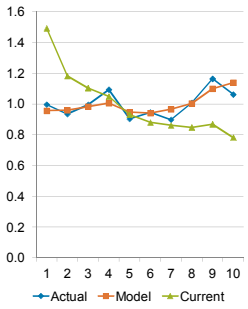


7

SERVE | ADD VALUE | INNOVATE

Model Lift – Double Lift Charts

- Creating a double lift chart
  - Sort data by ratio of model prediction to current premium.
  - Subdivide sorted data into quantiles with equal exposure.
  - For each quantile calculate average actual loss cost, average model predicted loss cost and the average loss cost underlying the current manual premium .
  - Index the quantile averages to the overall averages.

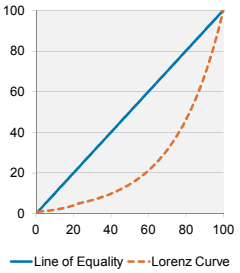


8

SERVE | ADD VALUE | INNOVATE

Economics – The Gini Index

- Gini coefficient or Gini ratio
  - Named after Corrado Gini
- Measure of income inequality
  - Horizontal axis = percentage of country's population
  - Vertical axis = percentage of country's income
  - A = Area between line of equality and Lorenz Curve
  - B = Area beneath Lorenz Curve
  - Gini index = A / (A + B)

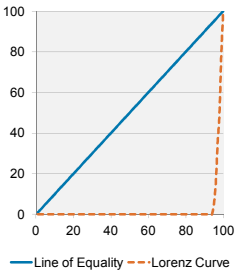


9

SERVE | ADD VALUE | INNOVATE

Model Lift – Simple Gini Index

- Adapting to car insurance
  - Assume claim frequency = 5%
- "The perfect model"
  - Prediction = actual loss, which is $0 for 95% of vehicles insured.
  - Sort holdout data set by model prediction.
  - Horizontal axis = percentage of total earned car years.
  - Vertical axis = Percentage of total incurred loss.
  - Gini Index = A / (A+ B) is very high.



10

SERVE | ADD VALUE | INNOVATE

Model Lift – Simple Gini Index

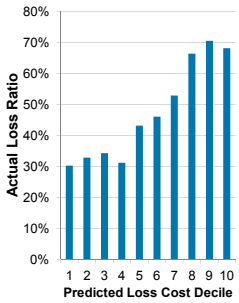
- Exercise:
  - Model X prediction = expected loss cost
  - Model Y prediction = 0.5 (expected loss cost)
  - Model Z prediction = 2.0 (expected loss cost)
  - Which model has the highest Gini index?
- Model A has a Gini index of 15.9 and B has a Gini index of 15.4
  - Is that difference significant, or is it just a quirk of the holdout data?

11

SERVE | ADD VALUE | INNOVATE

Model Lift – Loss Ratio Charts

- Lift charts and Gini index
  - May be unfamiliar to some stakeholders
- Loss ratios
  - Widely used and understood in the industry
- Ranking by predicted loss cost
  - Rank data into quantiles by predicted model loss cost
  - Calculate loss ratio for each quantile



12

SERVE | ADD VALUE | INNOVATE

Model Lift – Summary

- Simple Quantile plots
  - Illustrate how well the model helps prevent adverse selection
- Double lift charts
  - Compare competing models
  - Compare new model against current rating plan
- Simple Gini Index
  - Summarizes model lift into one number
- Loss ratio charts
  - Puts lift in context most people in insurance industry can understand
  - Can be distorted by redundancy or inadequacy of current rating plan

13

SERVE | ADD VALUE | INNOVATE

References

- Dickey, D. A., "Finding the Gold in Your Data: An Overview of Data Mining", SAS Global Forum 2013
- Gini, C. "On the Measure of Concentration with Special Reference to Income and Wealth", Cowles Commission Research Conference on Economics and Statistics, Colorado College Publication, General Series No. 208, 73–79, 1936.
- Jaffery, T. and Liu, S. X., "Measuring Campaign Performance by Using Cumulative Gain and Lift Chart", SAS Global Forum, 2009
- May, E., *Handbook of Credit Scoring*, Global Professional Publishing, 2001
- Parr Rud, O., *Data Mining Cookbook*, John Wiley & Sons, 2001

14

SERVE | ADD VALUE | INNOVATE

Contact Information

Hernan L. Medina, CPCU
Director, Analytical Data Management
ISO Insurance Programs and Analytic Services

hmedina@iso.com
www.verisk.com/insurance