



# **Model Validation: The Modeler's Perspective**

**Amber Popovitch, FCAS**

**CAS RPM Seminar**

**March 2012**

## Disclaimer

*The views expressed in this presentation are those of the author and do not necessarily reflect the views of The Travelers Companies, Inc. or any of its subsidiaries. This presentation is for general informational purposes only.*

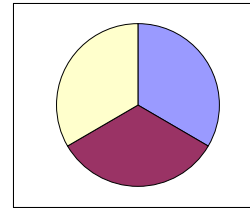
## What Is Model Validation?

From a modeler's perspective, there are two parts:

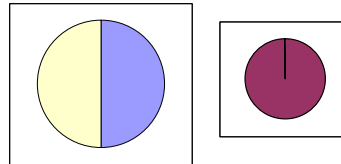
- Model Building
  - Have I chosen the right model? (e.g. are assumptions valid?)
  - Have I selected the right variables?
  - Have I adhered to the principle of parsimony?
  - Have I selected the right factors?
- Model Testing
  - Have I achieved the modeling objectives?
  - Have I avoided over-fitting my data?
  - Have I created a model that will predict future behavior?

# Data Partitioning

- Training / Validation / Holdout Approach



- Out of Time Validation



- Bootstrapping Approach

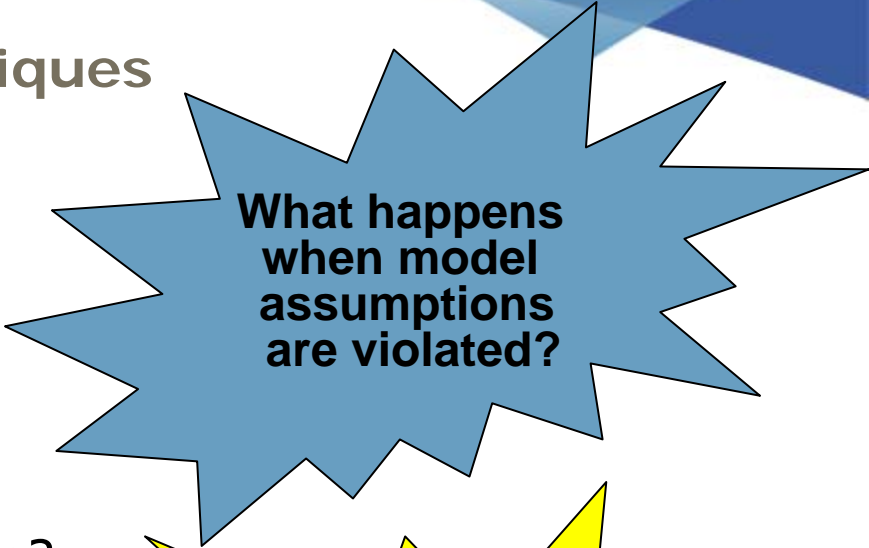
Original	Bootstrap 1	Bootstrap 2	Bootstrap 3
1	1	3	2
2	1	4	2
3	2	5	3
4	3	5	3
5	3	5	4

- Cross Validation Approach


Original	CrossValid1	CrossValid2	CrossValid3	CrossValid4	CrossValid5
1	2	1	1	1	1
2	3	3	2	2	2
3	4	4	4	3	3
4	5	5	5	5	4
5	1	2	3	4	5

## Model Building Tools and Techniques

- Type III statistics
- p-values for variable levels
- Factor assessment
  - Does it make business sense?
  - Does the relationship make sense?  
(e.g. monotonic)
- Comparison with other techniques
  - Univariate analysis
  - Decision trees
- Residual analysis
- AIC / BIC / log-likelihood / deviance measures



**What happens  
when model  
assumptions  
are violated?**

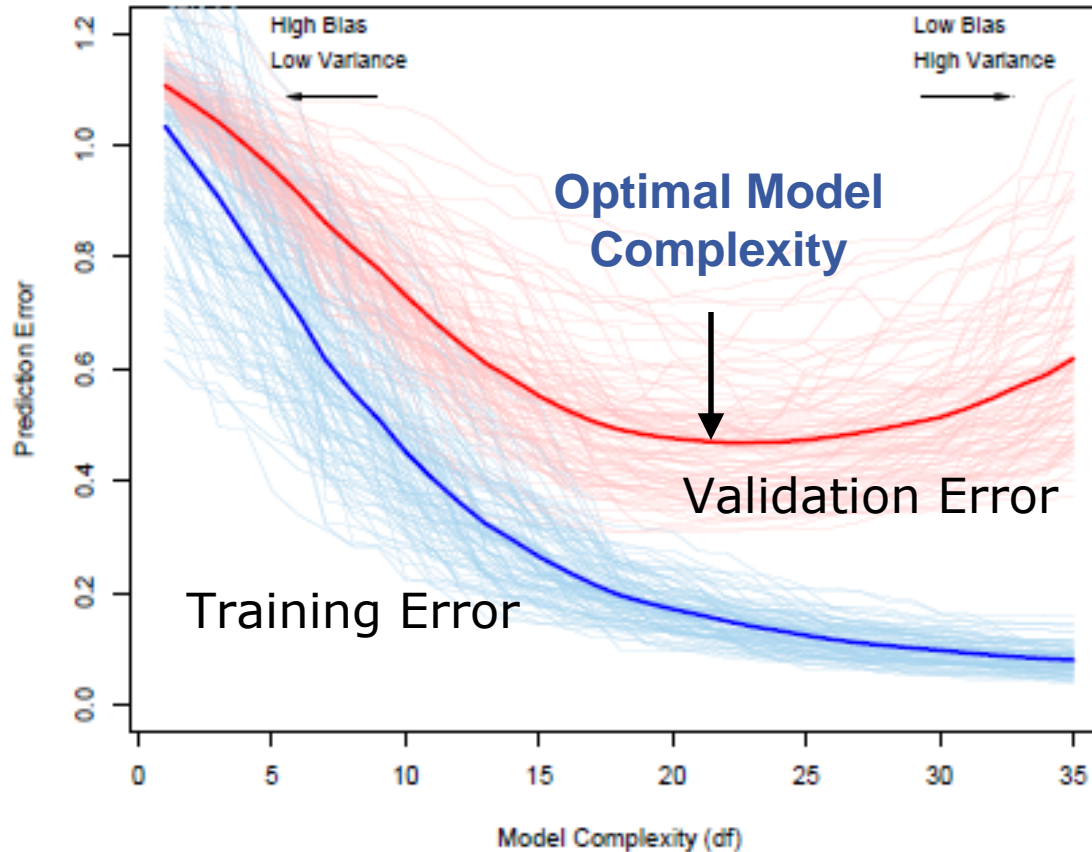


**The easy part  
is coming up with  
the story. . .**



**Beware of  
correlations!**

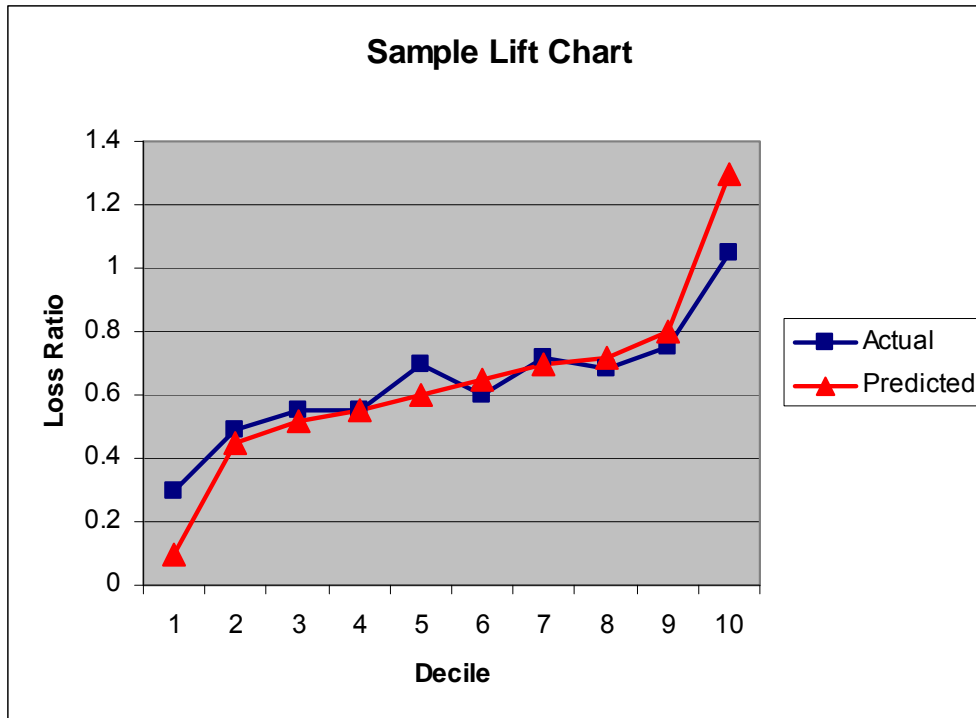
# Connecting Model Building and Model Testing



\* From [Elements of Statistical Learning](#) by Hastie, Tibshirani, and Friedman

# Model Testing Tools and Techniques

## The Lift Chart



### Questions:

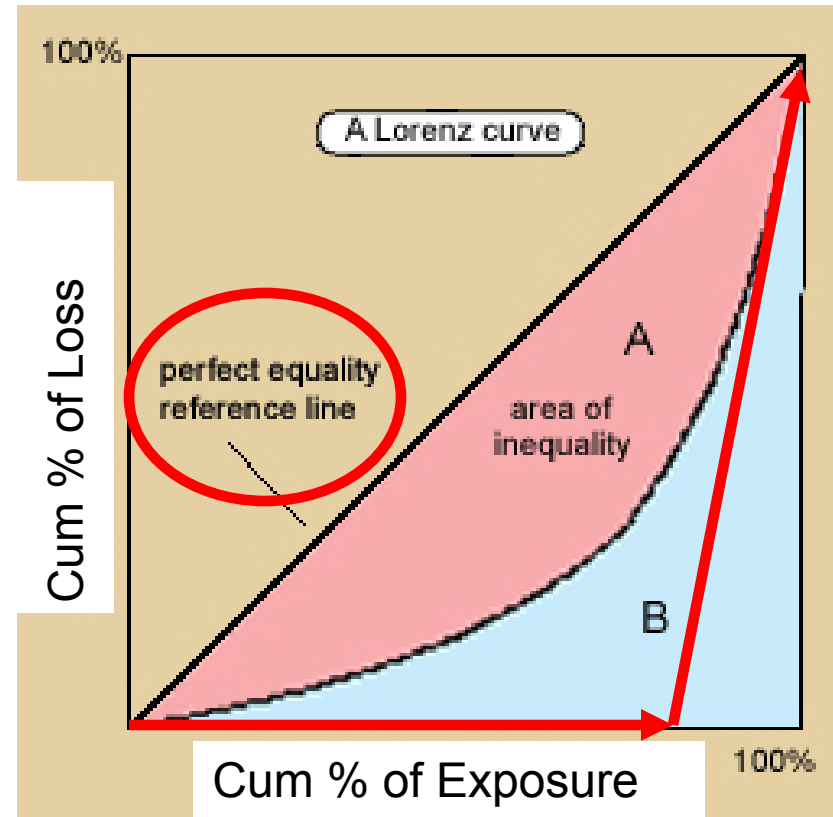
- How should lift be measured?
- How many buckets?
- How should reversals be interpreted?
- Are there variable biases affecting the ordering? (e.g. size, policy year)
- Is there over-fitting?
- Fit vs. Lift?

# Model Testing Tools and Techniques

## The GINI Index

$$Gini = \frac{A}{A + B}$$

- Commonly used to assess income inequality across countries
- More granular assessment of model fit
- Gives information on model segmentation
- $-1 \leq Gini \leq 1$  (1 = more segmentation, better fit)



Sort Predictions Low -> High

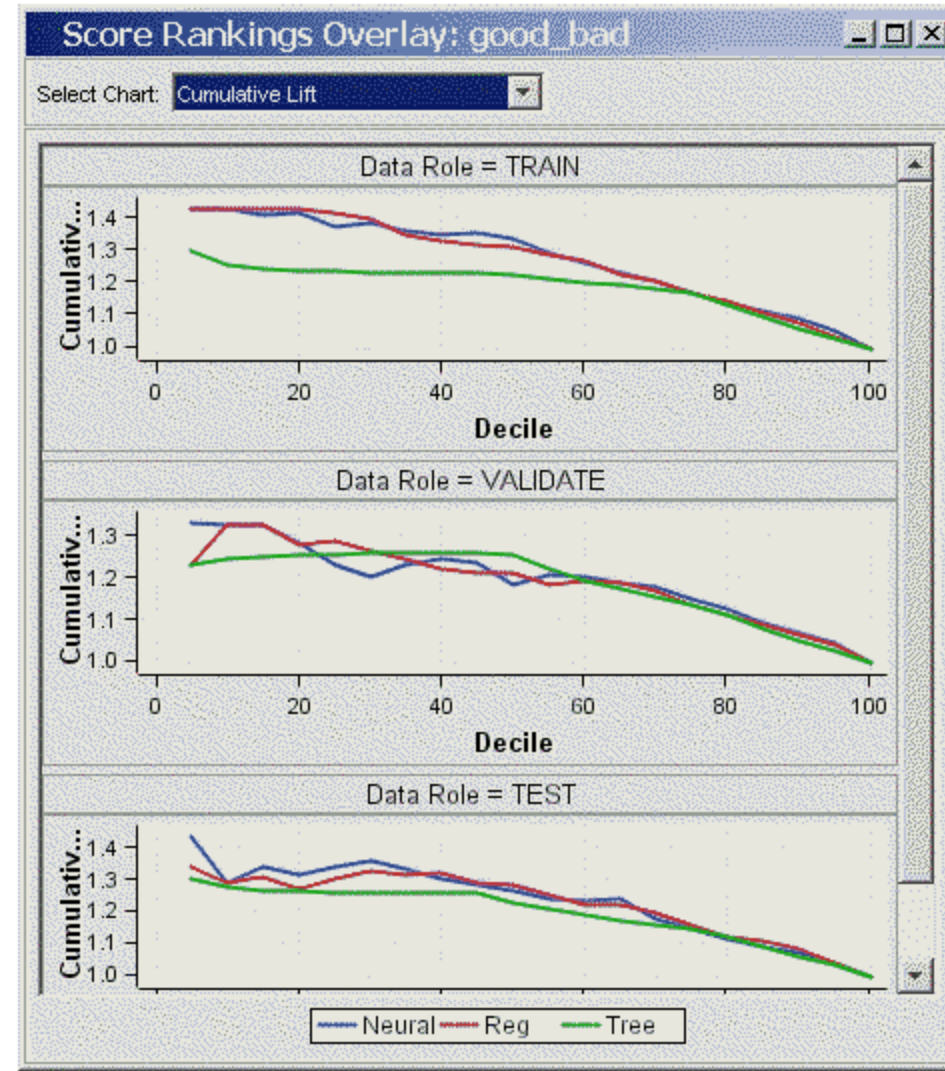
Reference: [http://en.wikipedia.org/wiki/Gini\\_index](http://en.wikipedia.org/wiki/Gini_index)



# Model Testing Tools and Techniques

## Comparing Across Models

- Which modeling technique is best?
- How much better is this version vs. the last one?
- Can use any measure you'd like – lift, GINI index, etc.
- Some software packages have this capability built in (e.g. Enterprise Miner)
- Be careful of over-fitting
- Don't use this on the holdout data as a model building technique!



## Food For Thought. . .

### **Should there be an actuarial standard of practice addressing predictive modeling?**

- Topics such a standard might address
  - When is out-of-time validation rather than just out-of-sample validation critical?
  - What steps should be taken to ensure knowledge of the holdout data has not crept into the model-building process?
    - For instance, split off the holdout data before or after EDA?
    - Splitting it too early makes balancing to control-totals difficult
  - Auditing
    - “Lock up” holdout data?
    - Peer review standards
  - What should be done when holdout data “disagrees?”