



Insurance and Actuarial
Advisory Services

 **ERNST & YOUNG**

Quality In Everything We Do

Similarity Modeling

Chuck Boucek

**CAS Seminar on Predictive Modeling – Las Vegas, Nevada
October 11 – 12, 2007**

www.ey.com/us/actuarial

Agenda

- Overview
- Underwriting Model
- Similarity Model
 - Conceptual framework
 - Process
 - Model results
- Market Deployment

Overview

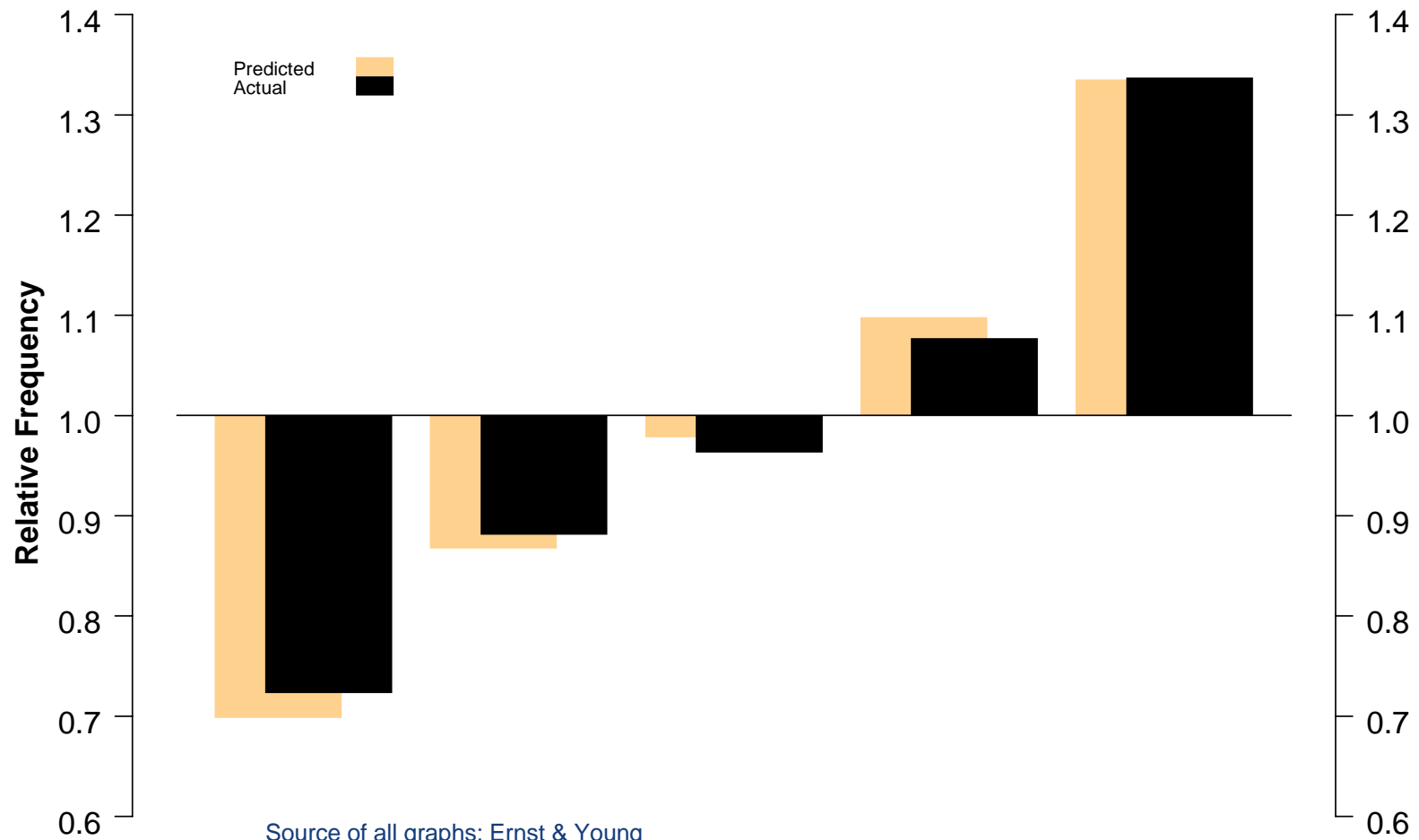
- Company's book of business is a function of its
 - Target market
 - Underwriting practices
 - Location of agencies
- Underwriting model is based on company's own data
 - Model is sometimes employed to score risks outside of target market
- Company approaches to addressing policyholders at extremes of historical data
 - Limits on risks that are scored via model
 - Manual underwriting of selected groups
 - Modeling approach

(Continued)

Overview

- Business risks of predictive modeling
 - Model does not appropriately assess policyholders
 - Model deployment worsens policyholder retention
 - Model is inappropriately extrapolated to policyholders
- Can an underwriting model be used to score all potential policyholders?
 - How to treat policyholders at extremes of historical data?
- Is there a structured way of assessing how similar a policyholder is to the policyholders in the data used to build the model?

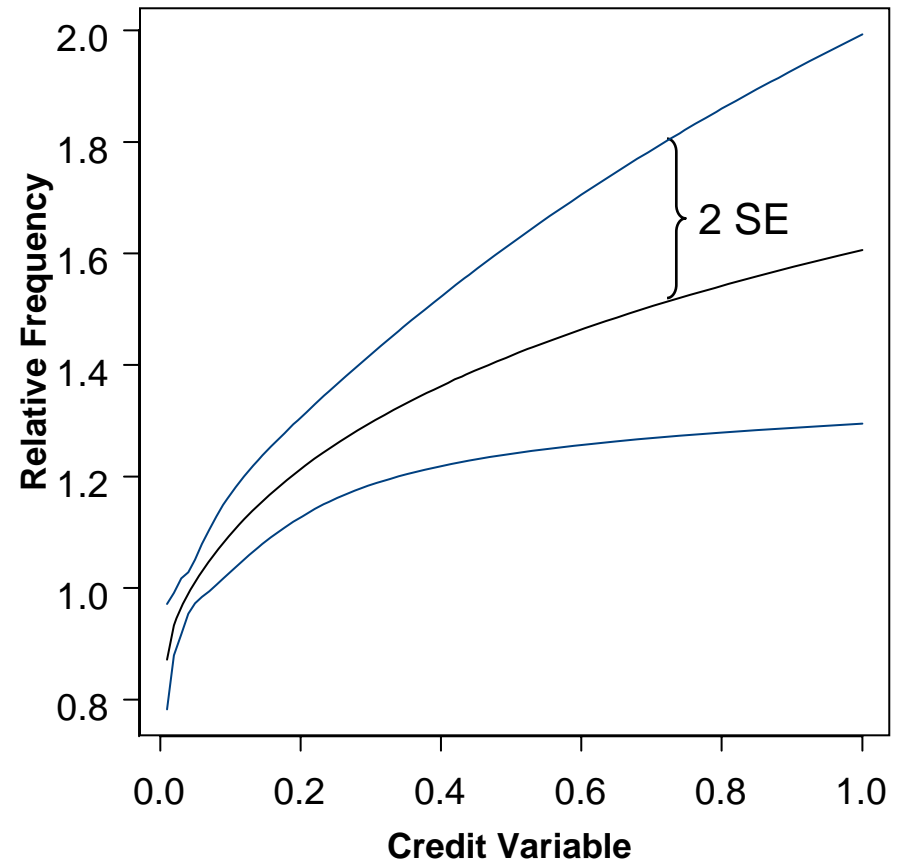
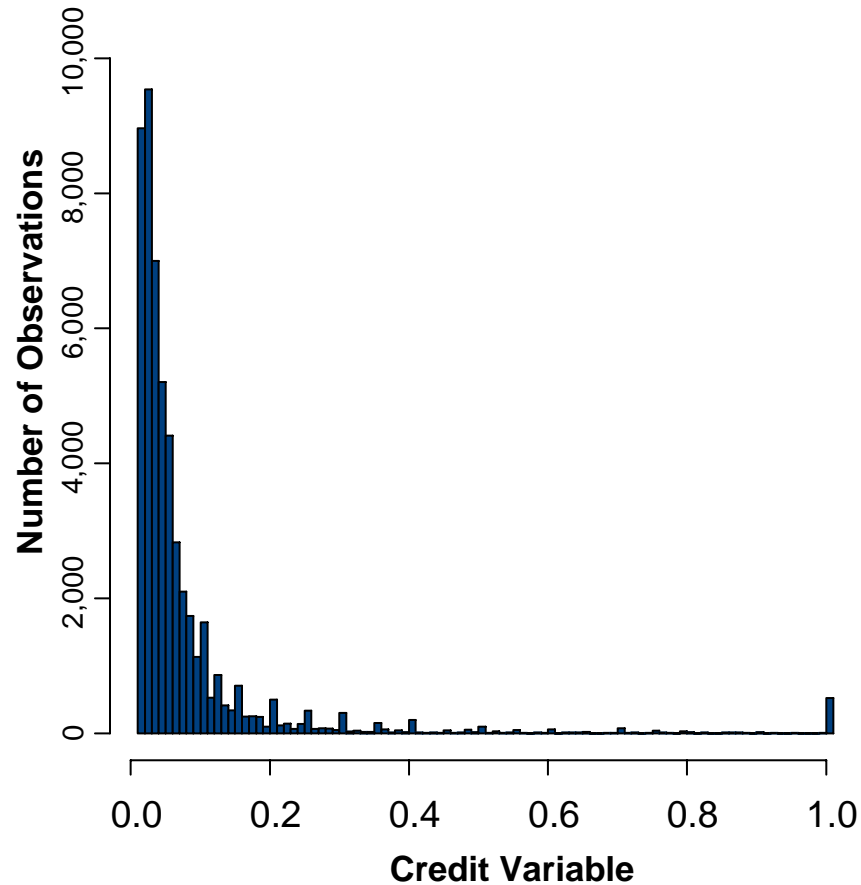
Underwriting Model – Lift Chart



Source of all graphs: Ernst & Young
Insurance and Actuarial Advisory Services

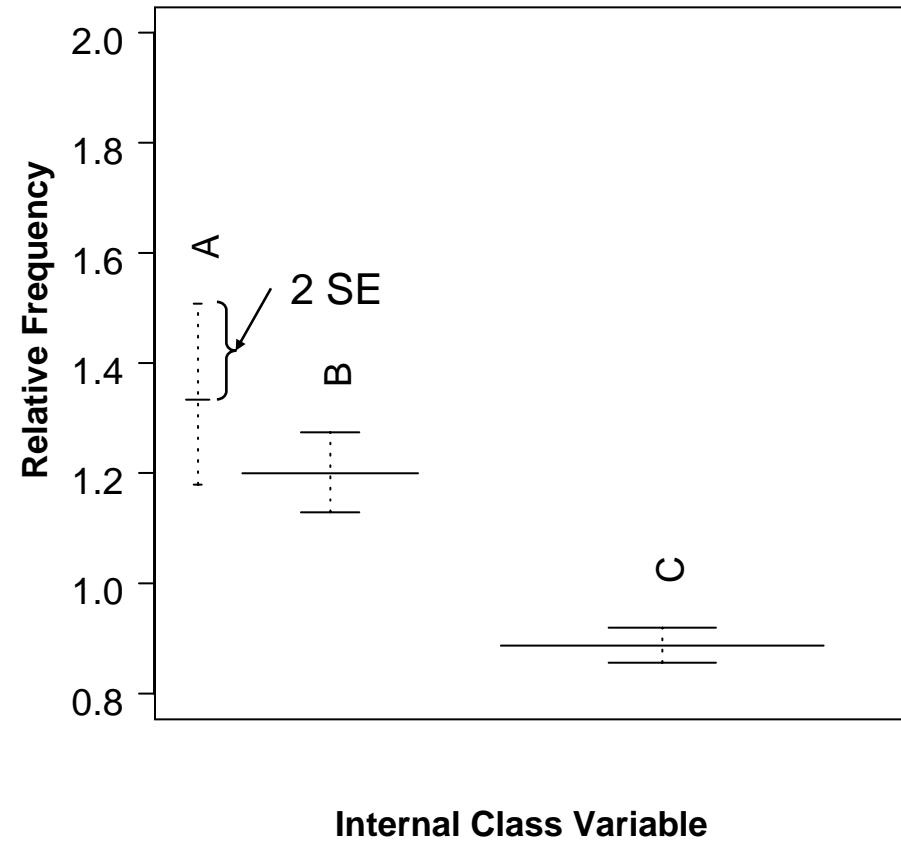
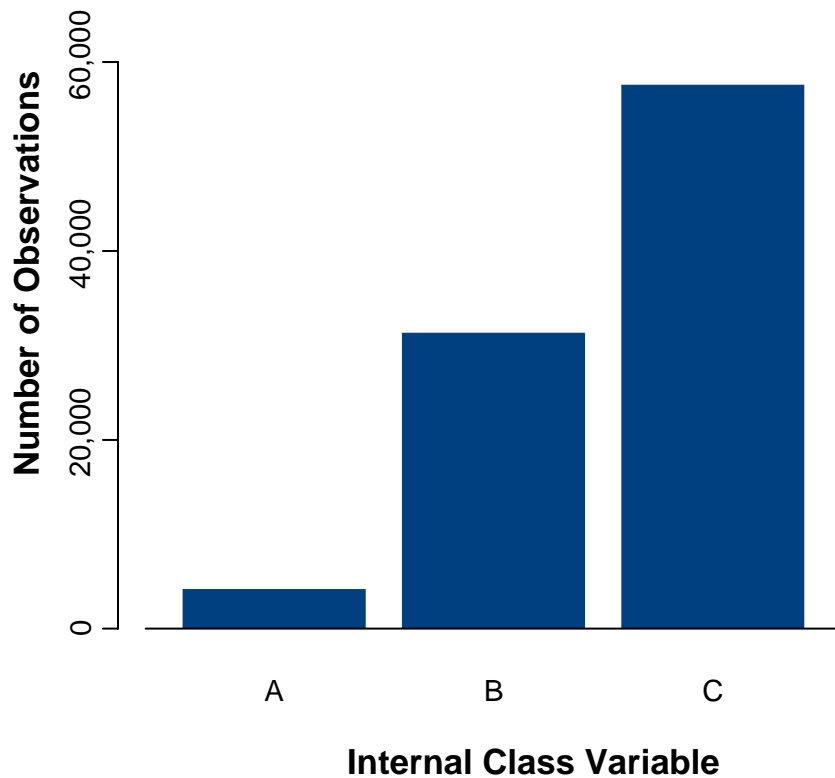
(Continued)

Underwriting Model – Predictor Variable #1



(Continued)

Underwriting Model – Predictor Variable #2



Similarity Model

- Conceptual framework
 - Policyholders with greater certainty in the claim frequency prediction should score higher
 - Generally, there is greater uncertainty in predictions from areas with more sparse data
 - Want to be able to assess similarity in a multivariate framework

(Continued)

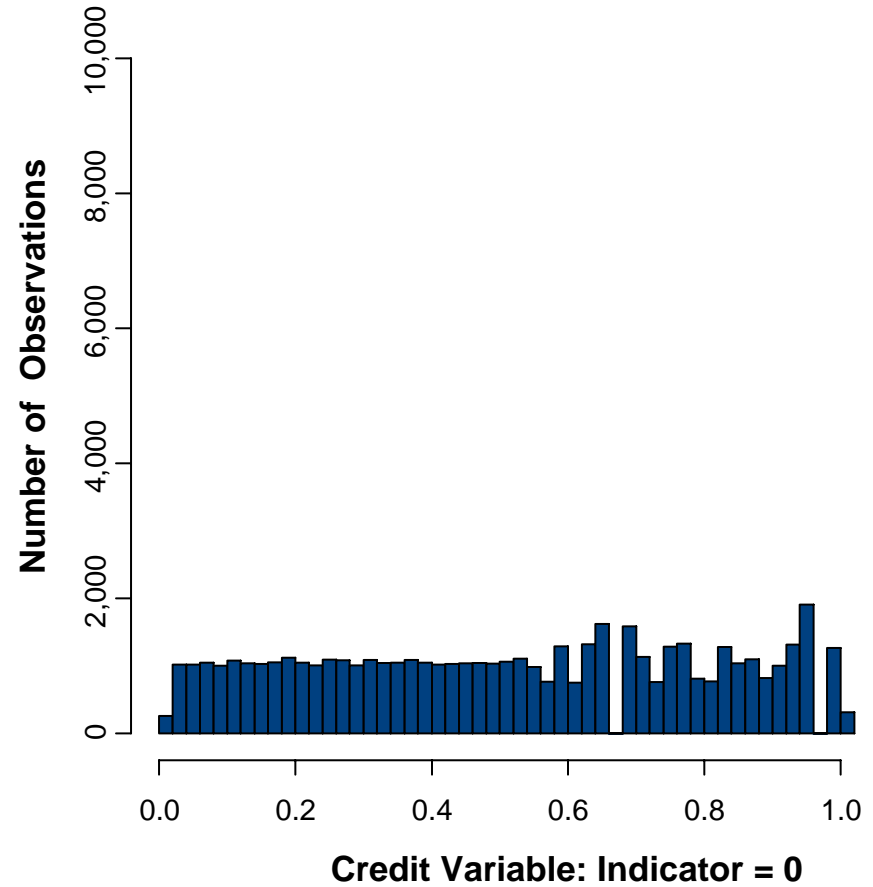
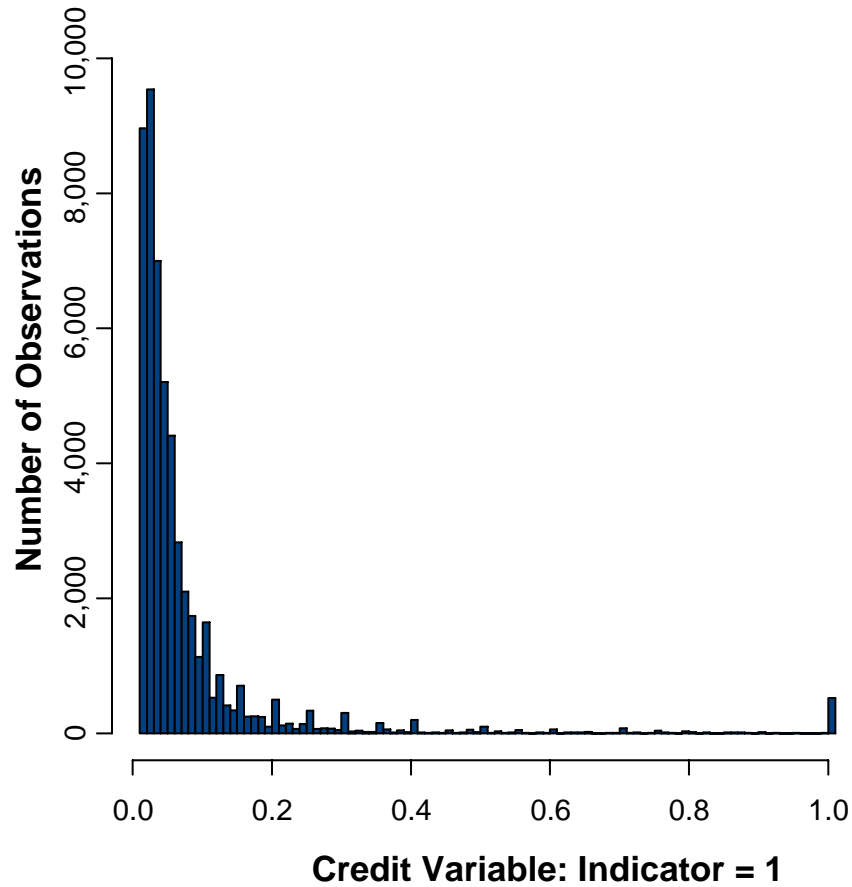
Similarity Model

- Process

- Generate one record for every record in modeling database
 - Numeric variable will be from a uniform distribution over the range of actual values
 - Factor variable will have an equal frequency for each level of the factor
 - Variables are limited to those in underwriting model
- Create an indicator variable
 - 1 = Record is from actual data
 - 0 = Record is from generated data
- Perform a logistic regression with the indicator variable as response variable
 - Response variable assumed to be binomial
 - Logit link function

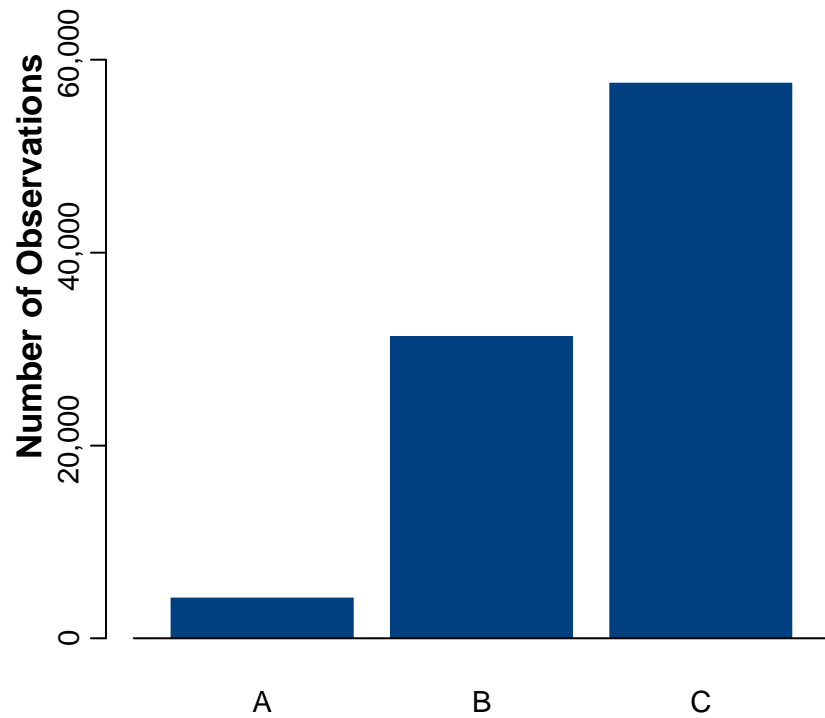
(Continued)

Similarity Model – Predictor Variable #1

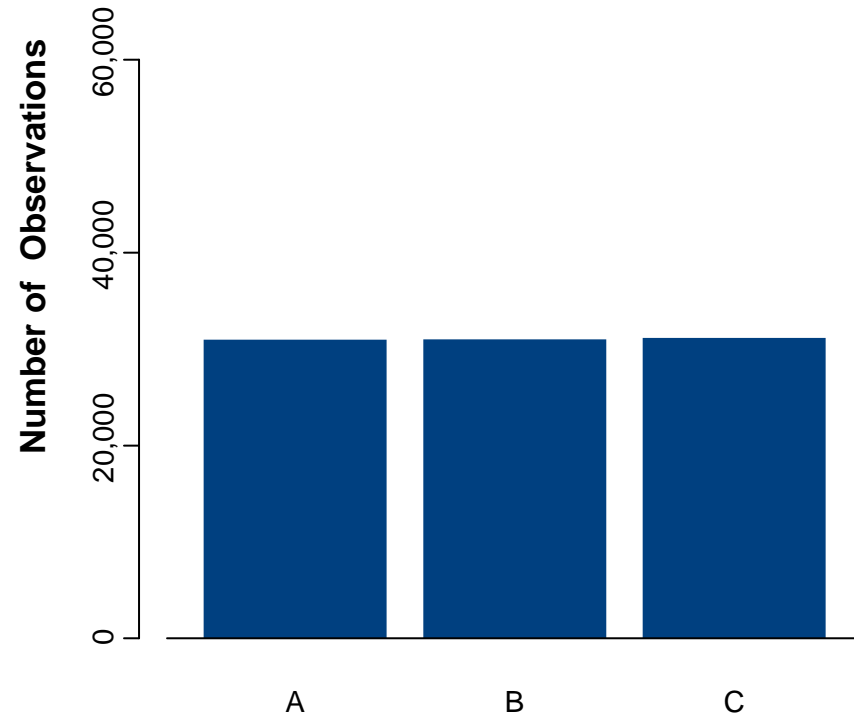


(Continued)

Similarity Model – Predictor Variable #2



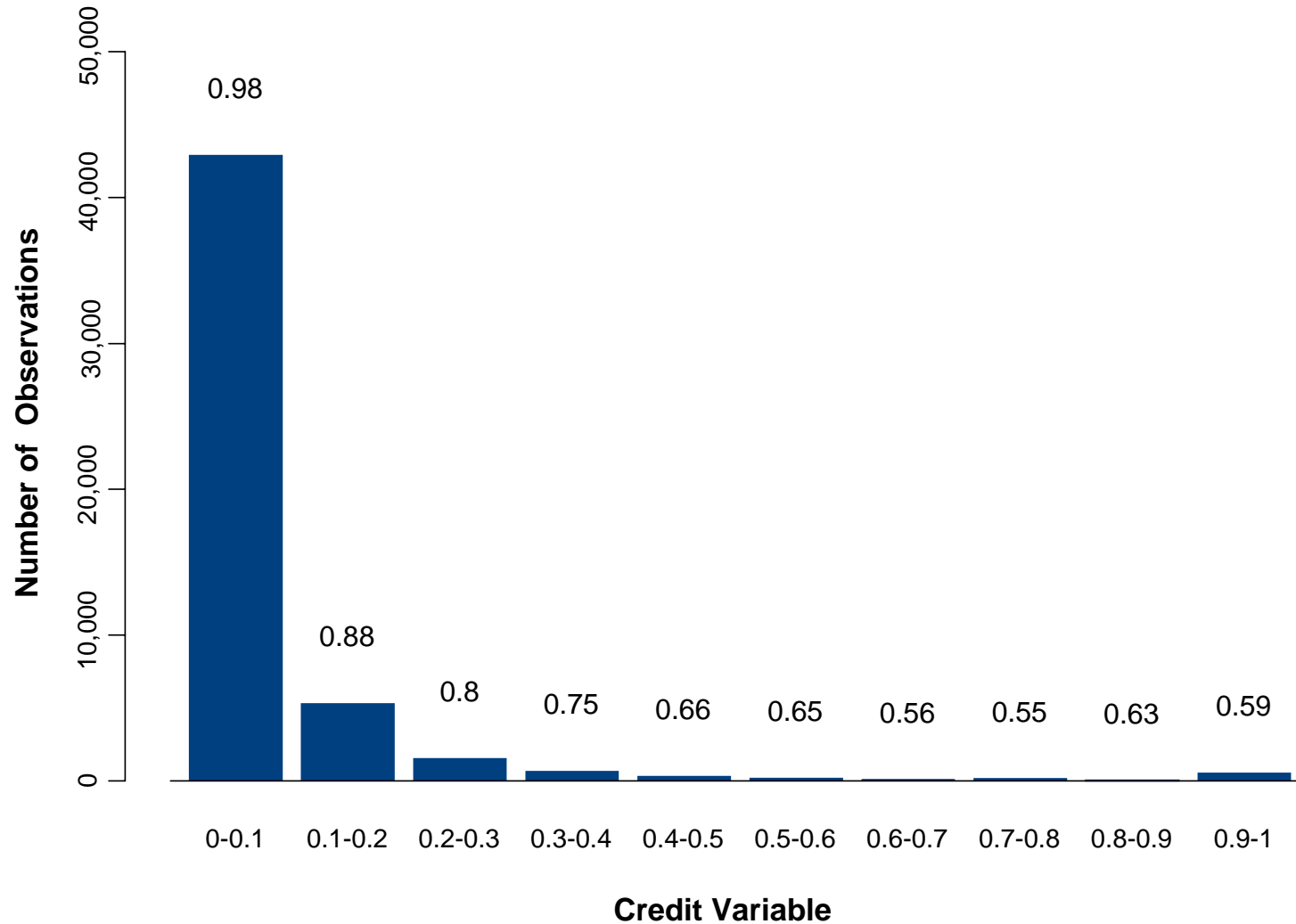
Internal Class Variable: Indicator = 1



Internal Class Variable: Indicator = 0

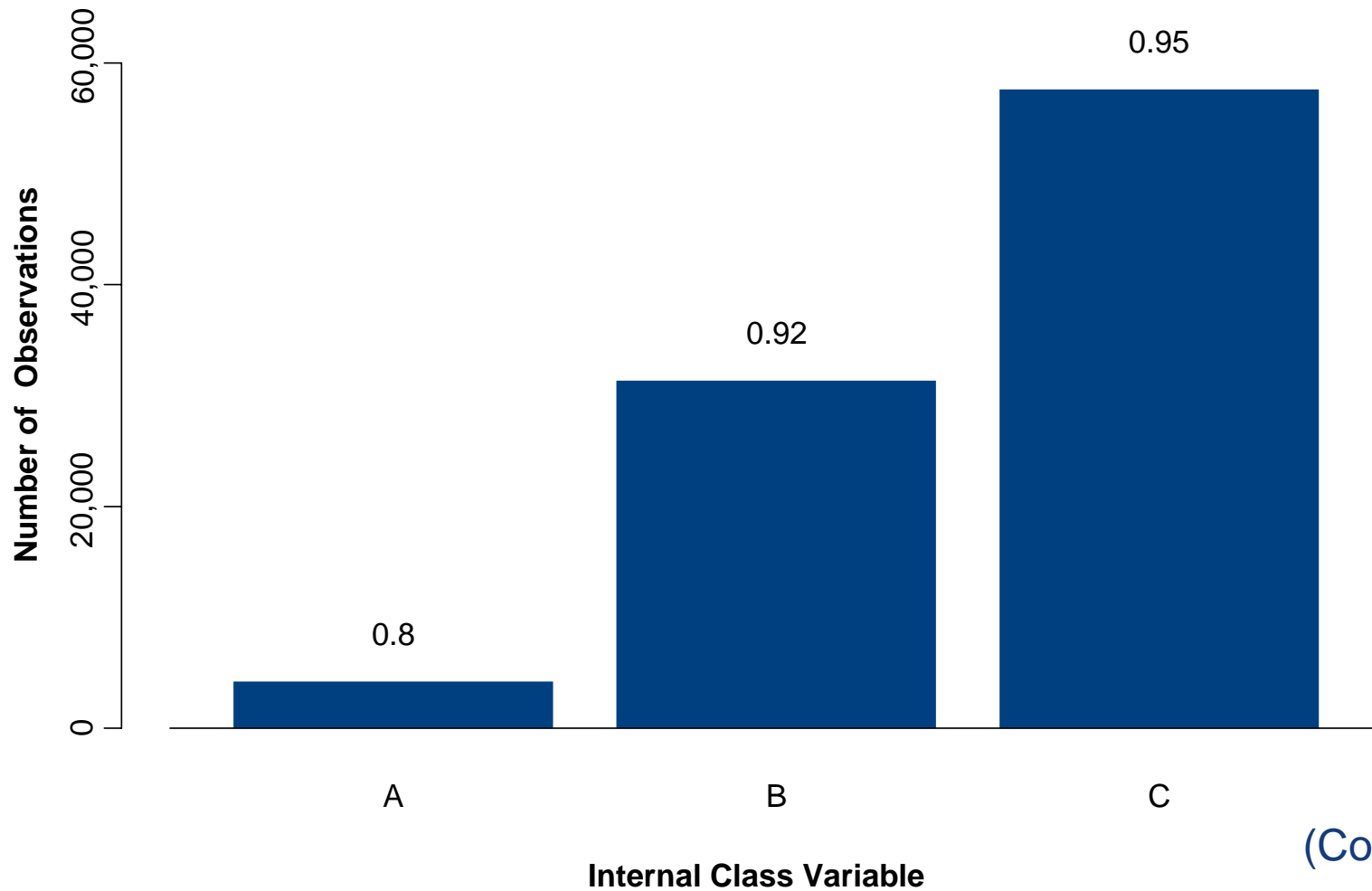
(Continued)

Similarity Model Results – Average Scores



(Continued)

Similarity Model Results – Average Scores



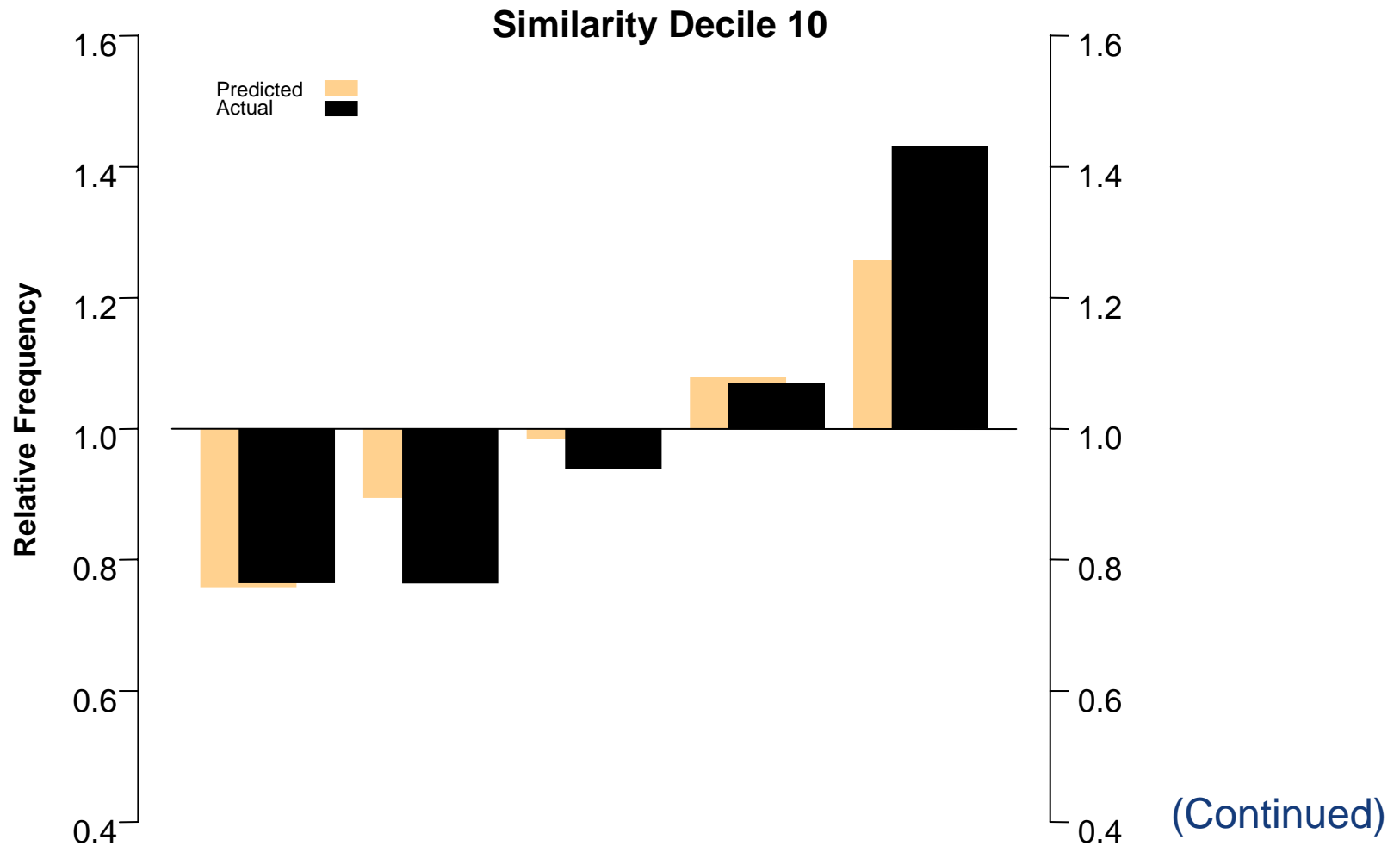
(Continued)

Similarity Model – Summary

- Advantages of this process
 - Produces results consistent with desired characteristics
 - It is a structured way of sorting risks from the fringes of the distribution
 - It is straightforward conceptually
 - It employs a Generalized Linear Model
- This is not the only process that can be employed

(Continued)

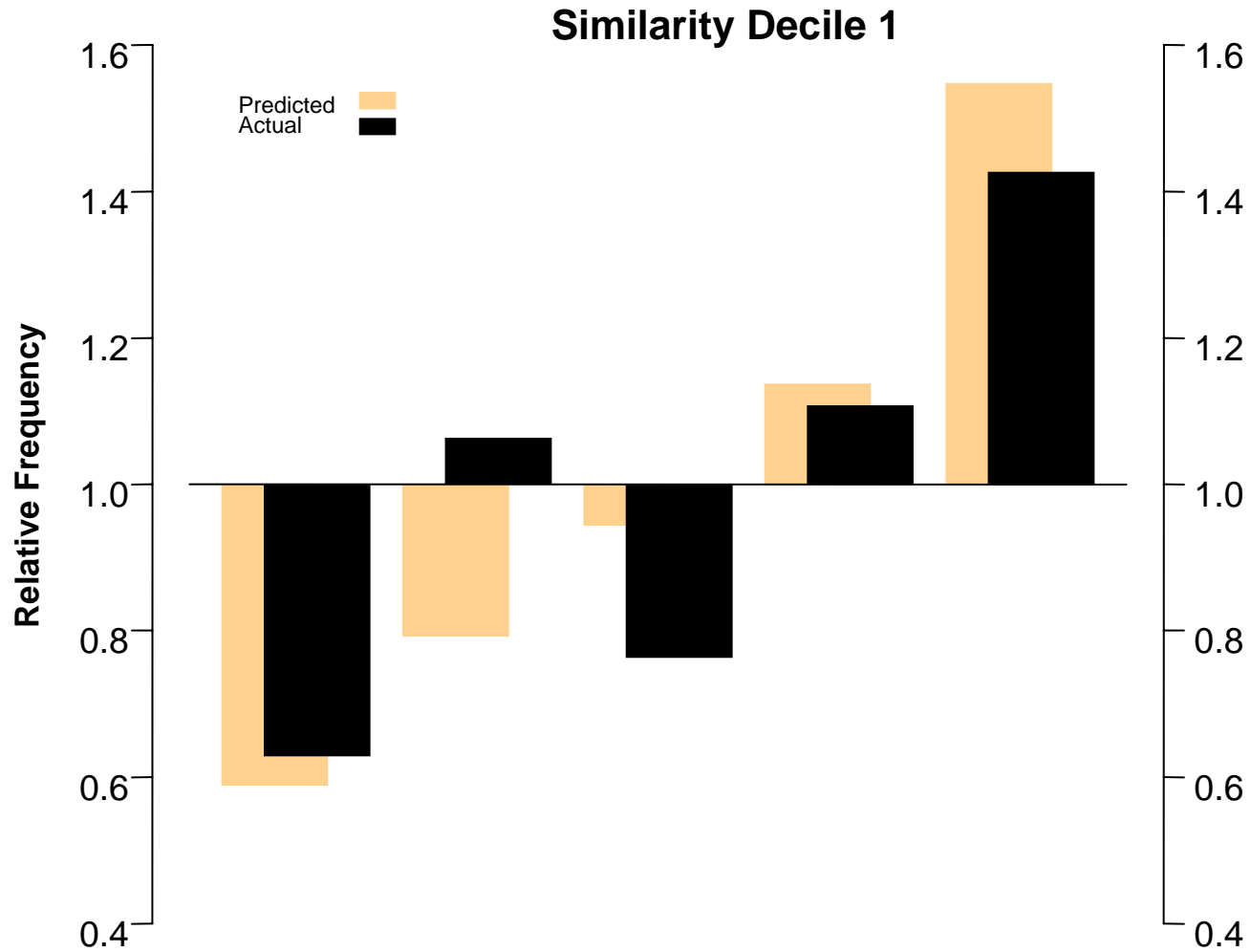
Similarity Model – Impact on Lift Chart



Similarity Model – Impact on Lift Chart



Similarity Model – Impact on Lift Chart



(Continued)

Similarity Model

- Observations

- Policyholders in similarity decile 1 should show more variability in actual results
- The predicted spread in claim frequency is smaller in decile 10 than in decile 1
 - The policyholders in decile 10 tend to be alike

Market Deployment

- Applications
 - Select a cutoff similarity score
 - Based on variability in lift chart
 - Professional judgment of underwriting staff
 - Policyholders with lower similarity scores will receive greater underwriting attention
 - Monitor results by similarity group