2008 Predictive Modeling Seminar

Estimating Loss Cost at the Address Level

Glenn Meyers
ISO Innovative Analytics

Territorial Ratemaking

- Territories should be big
  - Have a sufficient volume of business to make credible estimates of the losses.
- Territories should be small
  - "You live near that bad corner!"
  - Driving conditions vary within territory.

Some Environmental Features Related to Auto Accidents

- Proximity to Business Districts
  - Workplaces
    - Busy at beginning and end of work day
  - Shopping Centers
    - Always busy (especially on weekends)
  - Restaurants
    - Busy at mealtimes
  - Schools
    - Busy and beginning and end or school day

Some Environmental Features Related to Auto Accidents

- Weather
  - Rainfall
  - Temperature
  - Snowfall (especially in hilly areas)
- Traffic Density
  - More traffic sharing the same space increases odds of collision
- Others

Combining Environmental Variables at a Particular Garage Address

- Individually, the geographic variables have a predictable effect on accident rate and severity.
- Variables for a particular location could have a combination of positive and negative effects.
- ISO has built a model to calculate the combined effect of all variables.
  - Based on countrywide data – Actuarially credible

View as Case Study in Model Development

- Reduction in number of variables
  - Necessary for small insurers
- Special circumstances in fitting models to individual auto data.
- Diagnostics
  - Graphic and Maps
- Economic value of lift

Data Used in Building Model

- Obtained loss, exposure, classification and address for individual policies from cooperating insurers
- ISO Statistical Plan data
- Third-Party Data
  - Traffic
  - Business Location
  - Demographic
  - Weather
  - etc
- Approximately 1,000 indicators

Environmental Module Examples

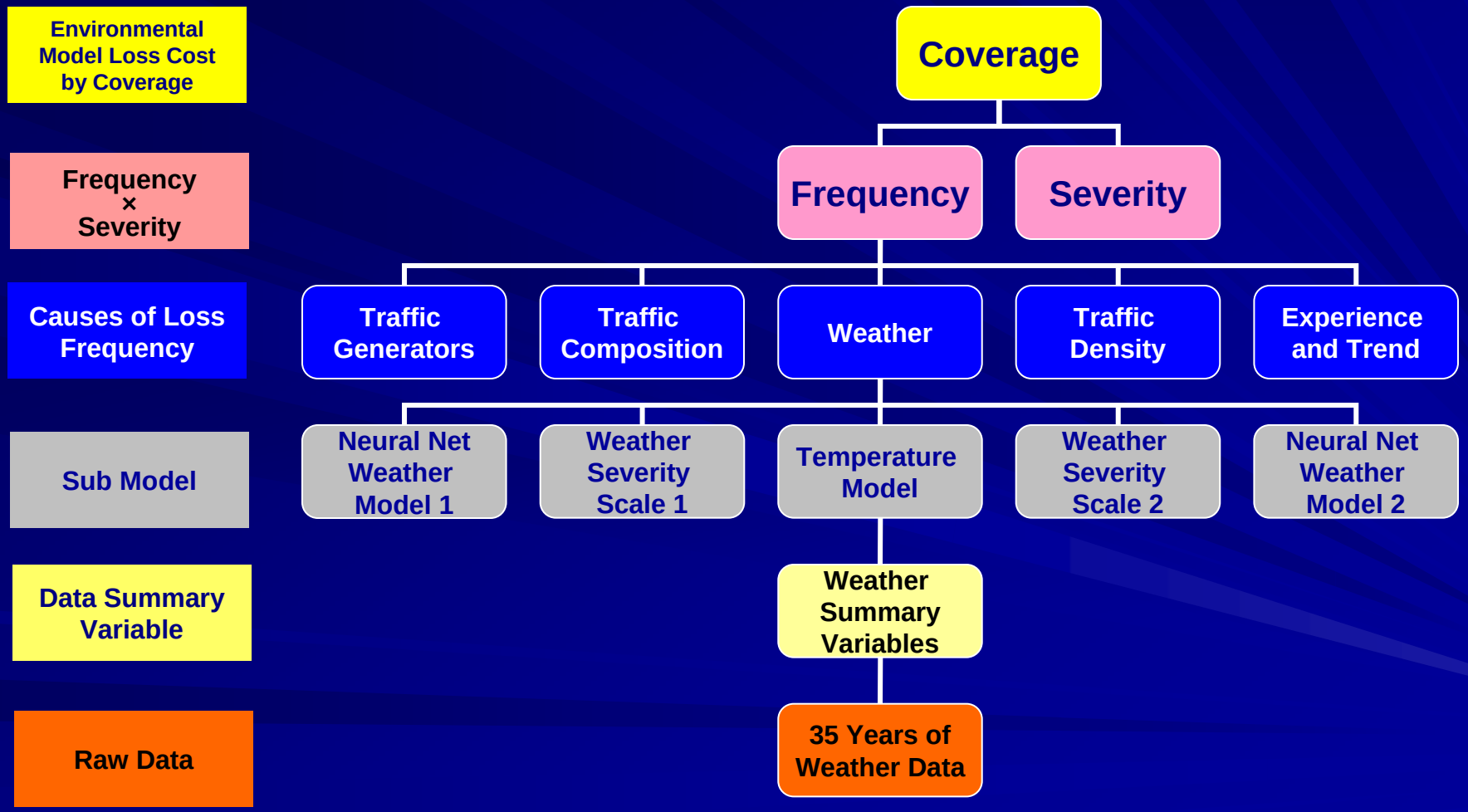
- **Comprised of over 1000 indicators**

- **Weather:**
  - Measures of snowfall, rainfall, temperature, wind and elevation
- **Traffic Density and Driving Patterns**:
  - Commute patterns
  - Public transportation usage
  - Population density
  - Types of housing
- **Traffic Composition**
  - Demographic groups
  - Household size
  - Homeownership
- Traffic Generators
  - Transportation hubs
  - Shopping centers
  - Hospitals/medical centers
  - Entertainment districts
- Experience and trend:
  - ISO loss cost
  - State frequency and severity trends from ISO lost cost analysis

Techniques Employed in Variable Reduction

- Variable Selection – univariate analysis, transformations, known relationship to loss
- Sampling
- Sub models/data reduction – neural nets, splines, principal component analysis, variable clustering
- Spatial Smoothing – with parameters related to auto insurance loss patterns

In Depth for Weather Component



Environmental Model

Loss Cost = Pure Premium

= Frequency x Severity

$$\text{Frequency} = \frac{e^{\lambda}}{1+e^{\lambda}}$$

λ = Intercept

+ Weather

+ Traffic Density

+ Traffic Generators

+ Traffic Composition

+ Experience and Trend

Environmental Model

Loss Cost = Pure Premium

= Frequency x Severity

Severity = e^{μ}

μ = Intercept

+ Weather

+ Traffic Density

+ Traffic Generators

+ Traffic Composition

+ Experience and Trend

Environmental Model

Loss Cost = Pure Premium

= Frequency x Severity

- **Separate Models by Coverage**
  - Bodily Injury Liability
  - No-Fault
  - Property Damage Liability
  - Collision
  - Comprehensive

Constructing the Components Frequency Model as Example

$\lambda =$ Intercept

$+ \alpha_1 \cdot x_1 + \ldots + \alpha_{n_1} \cdot x_{n_1}$	= Weather
$+ \alpha_{n_1+1} \cdot x_{n_1+1} + \ldots + \alpha_{n_2} \cdot x_{n_2}$	= Traffic Density
$+ \alpha_{n_2+1} \cdot x_{n_2+1} + \ldots + \alpha_{n_3} \cdot x_{n_3}$	= Traffic Generators
$+ \alpha_{n_3+1} \cdot x_{n_3+1} + \ldots + \alpha_{n_4} \cdot x_{n_4}$	= Traffic Composition
$+ \alpha_{n_4+1} \cdot x_{n_4+1} + \ldots + \alpha_{n_5} \cdot x_{n_5}$	= Experience & Trend

$+$ Other Classifiers

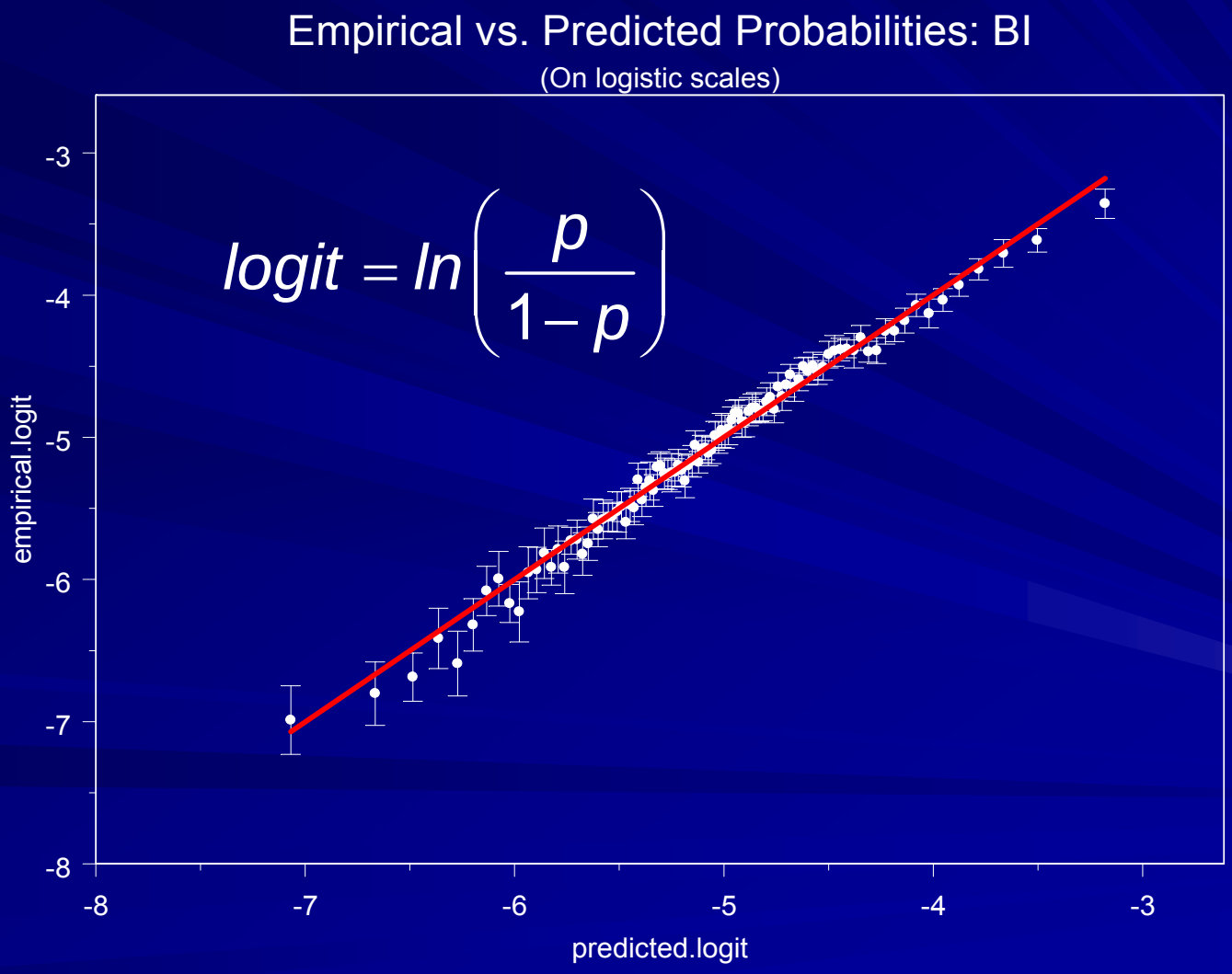
Constructing the Components Frequency Model as Example

- "Other Classifiers" reflect driver, vehicle, limits and deductibles.
- Model output is deployed to a base class, standard limits and deductibles.

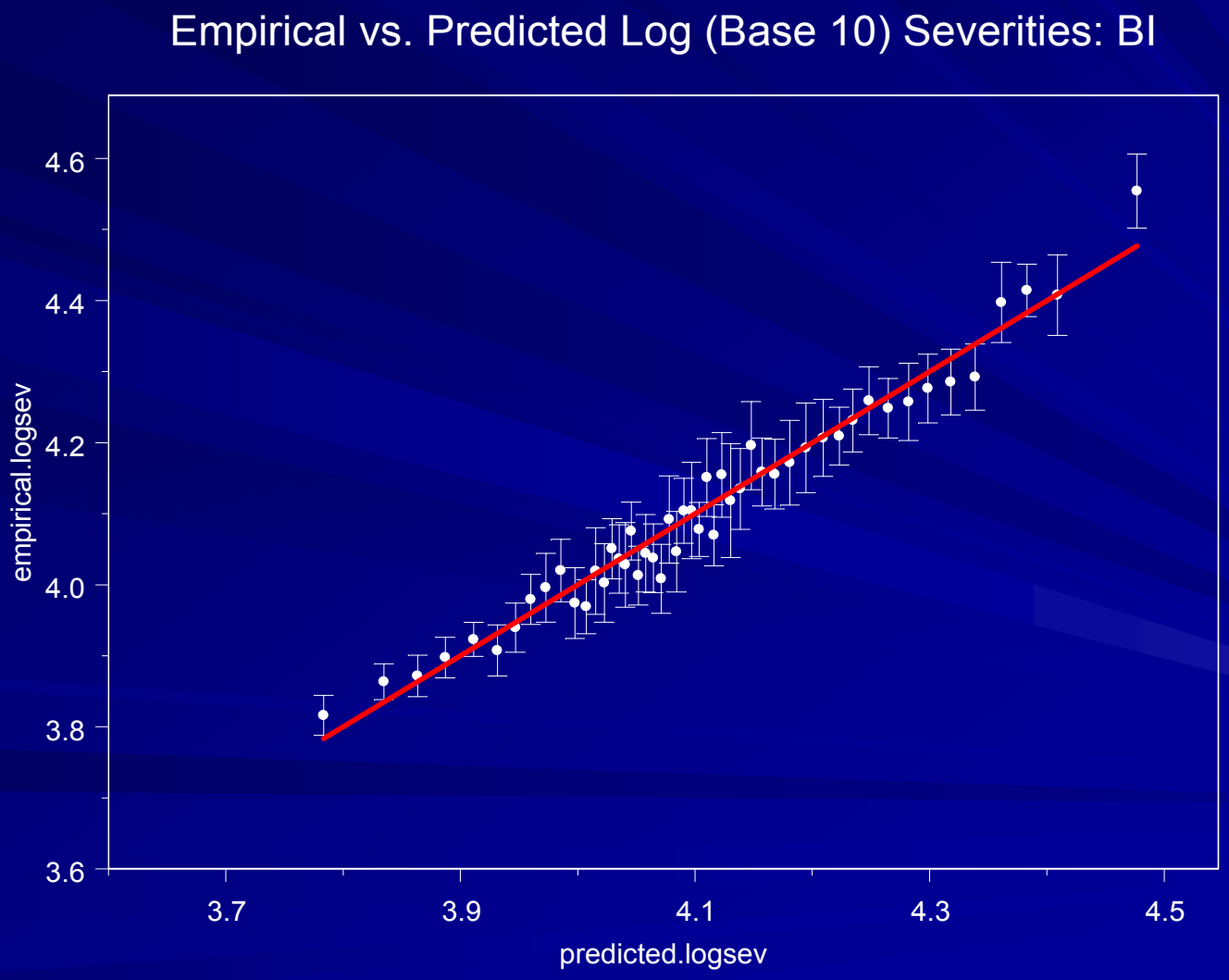
Overall Model Diagnostics

- Sort in order of increasing prediction
  - Frequency & Severity
- Group observations in buckets
  - 1/100th of record count for frequency
  - 1/50th of the record count for severity
- Calculate bucket averages
- Apply the GLM link function for bucket averages and predicted value
  - logit for frequency
  - log for severity
- Plot predicted vs empirical
  - With confidence bands

Overall Diagnostics - Frequency



Overall Diagnostics - Severity



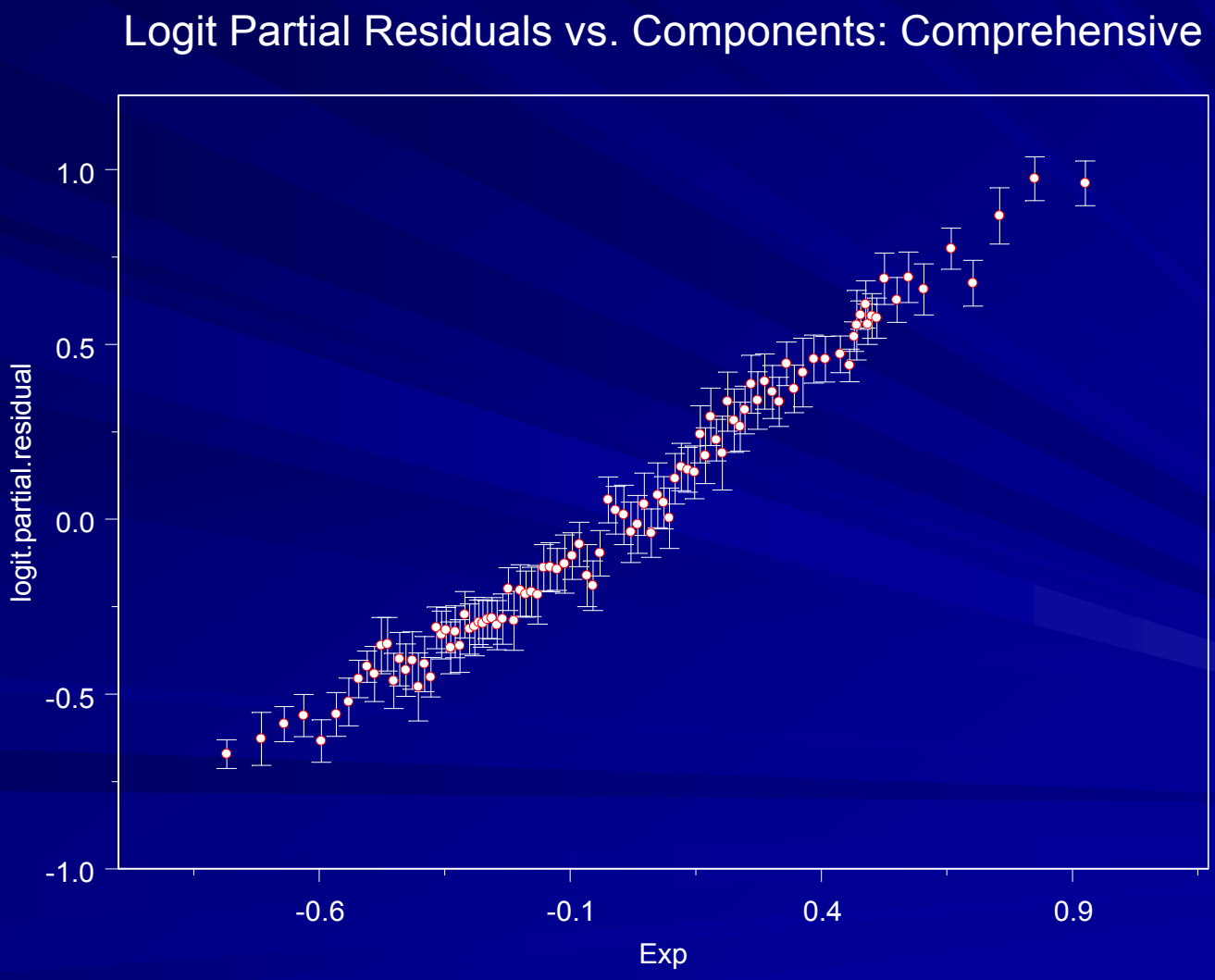
Component Diagnostics Frequency Example

- Sort observations in order of C_i
- Bucket as above and calculate
  - C_{ib} = Average C_i in bucket b
  - p_{ib} = Average p_i in bucket b
  - Partial Residuals

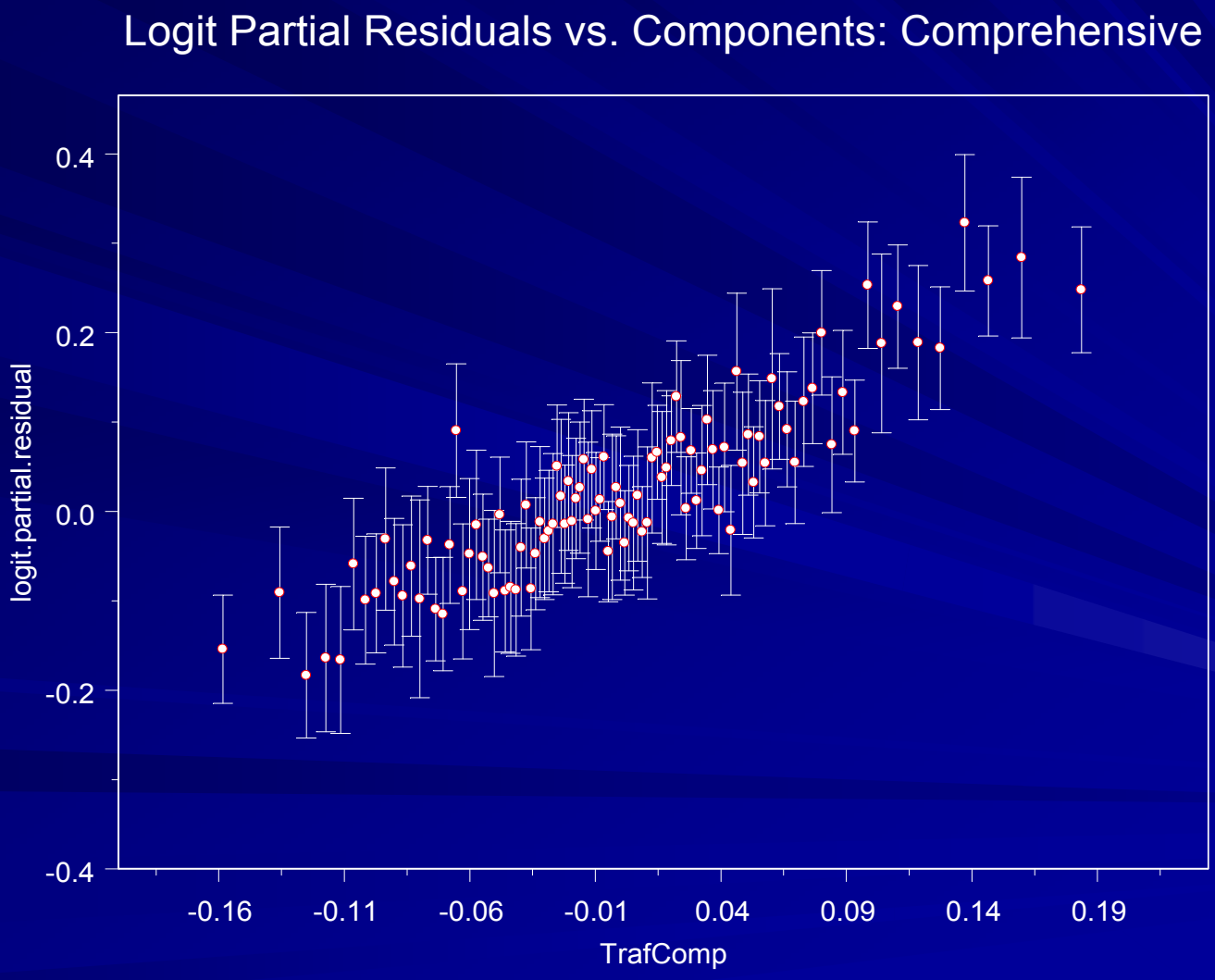
$$R_{ib} = ln\left(\frac{p_{ib}}{1-p_{ib}}\right) - \left(\lambda + \sum_{k \neq i} C_{kb}\right)$$

- Plot C_{ib} vs R_{ib} – Expect linear relationship

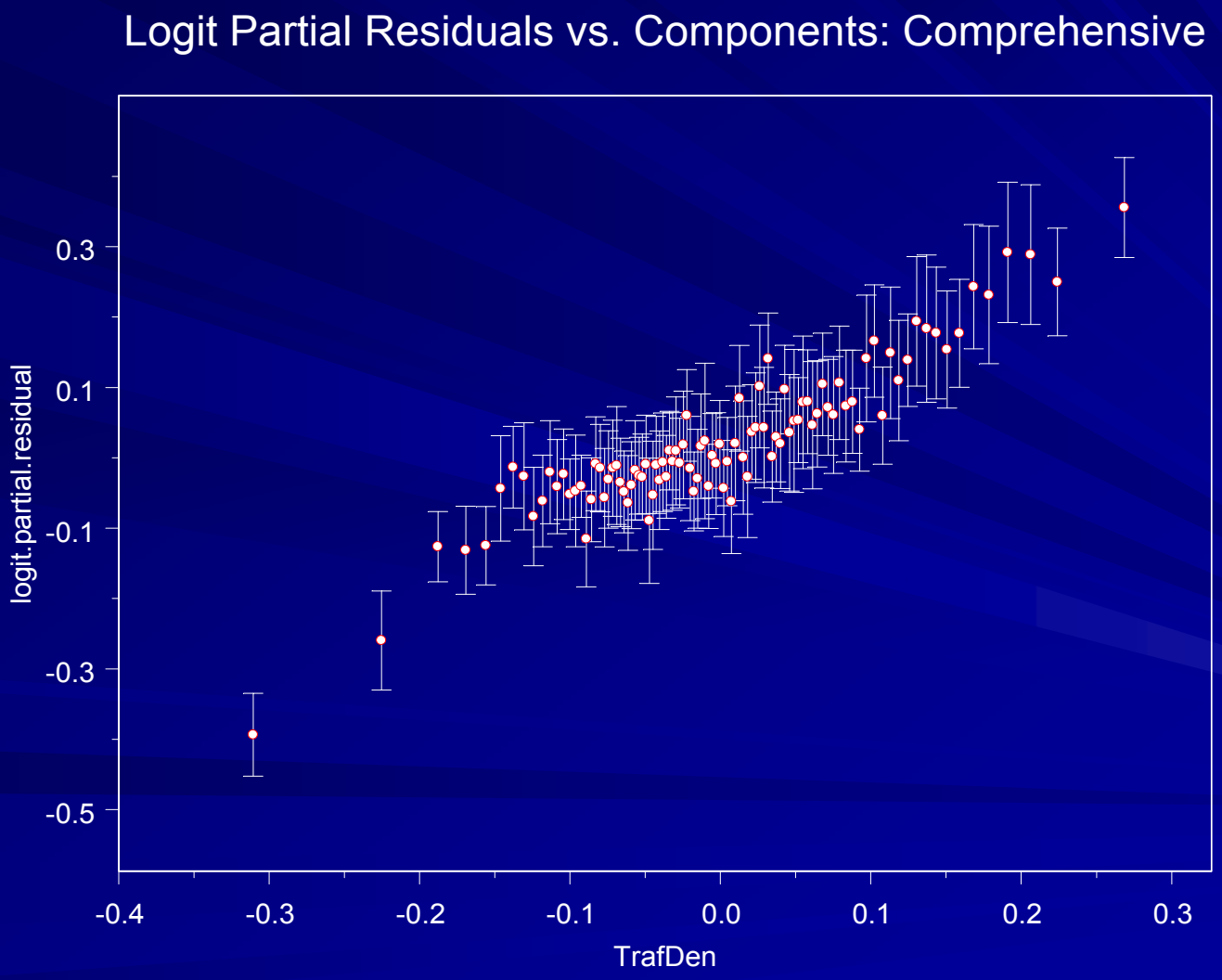
Component Diagnostics Experience and Trend



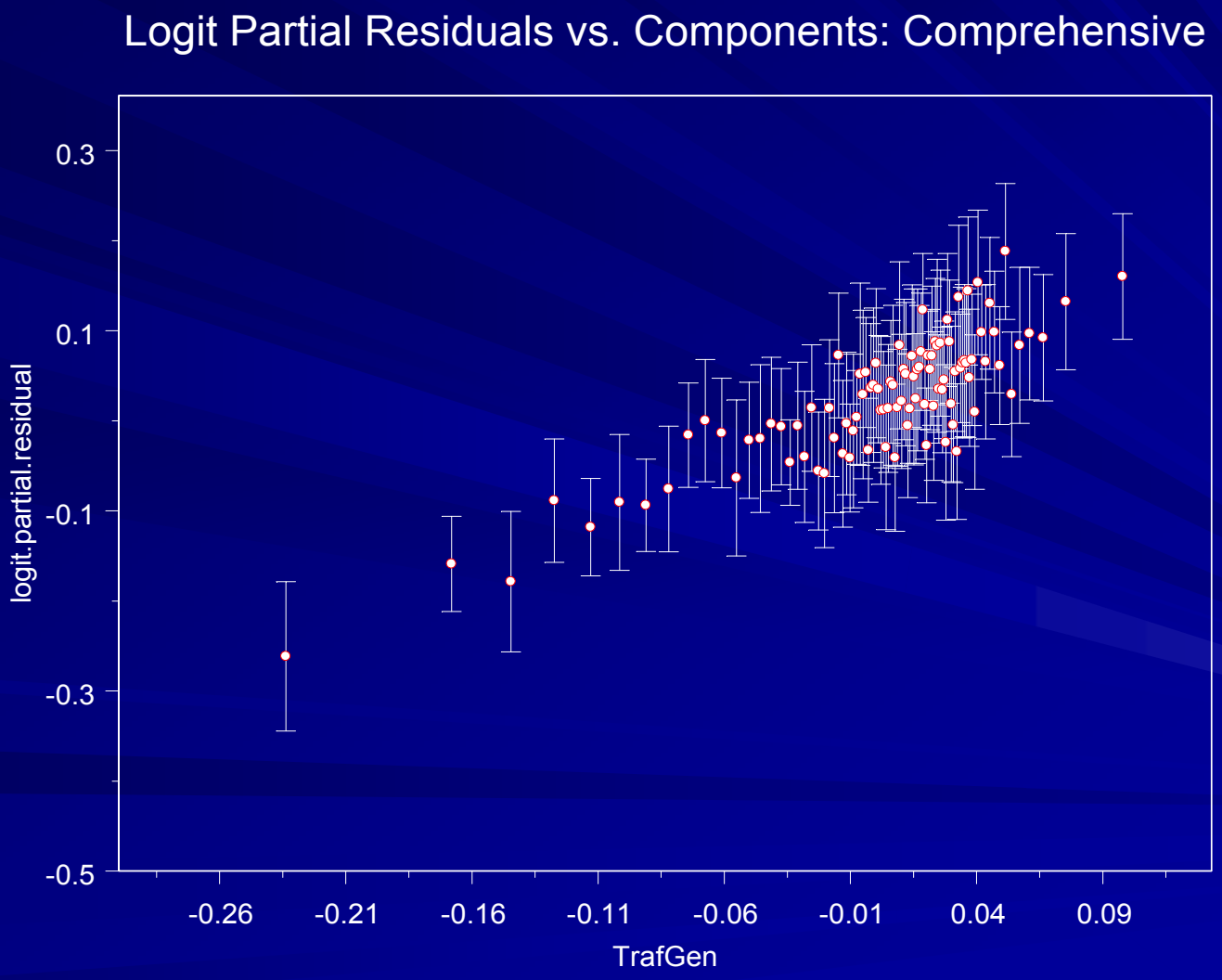
Component Diagnostics
Traffic Composition



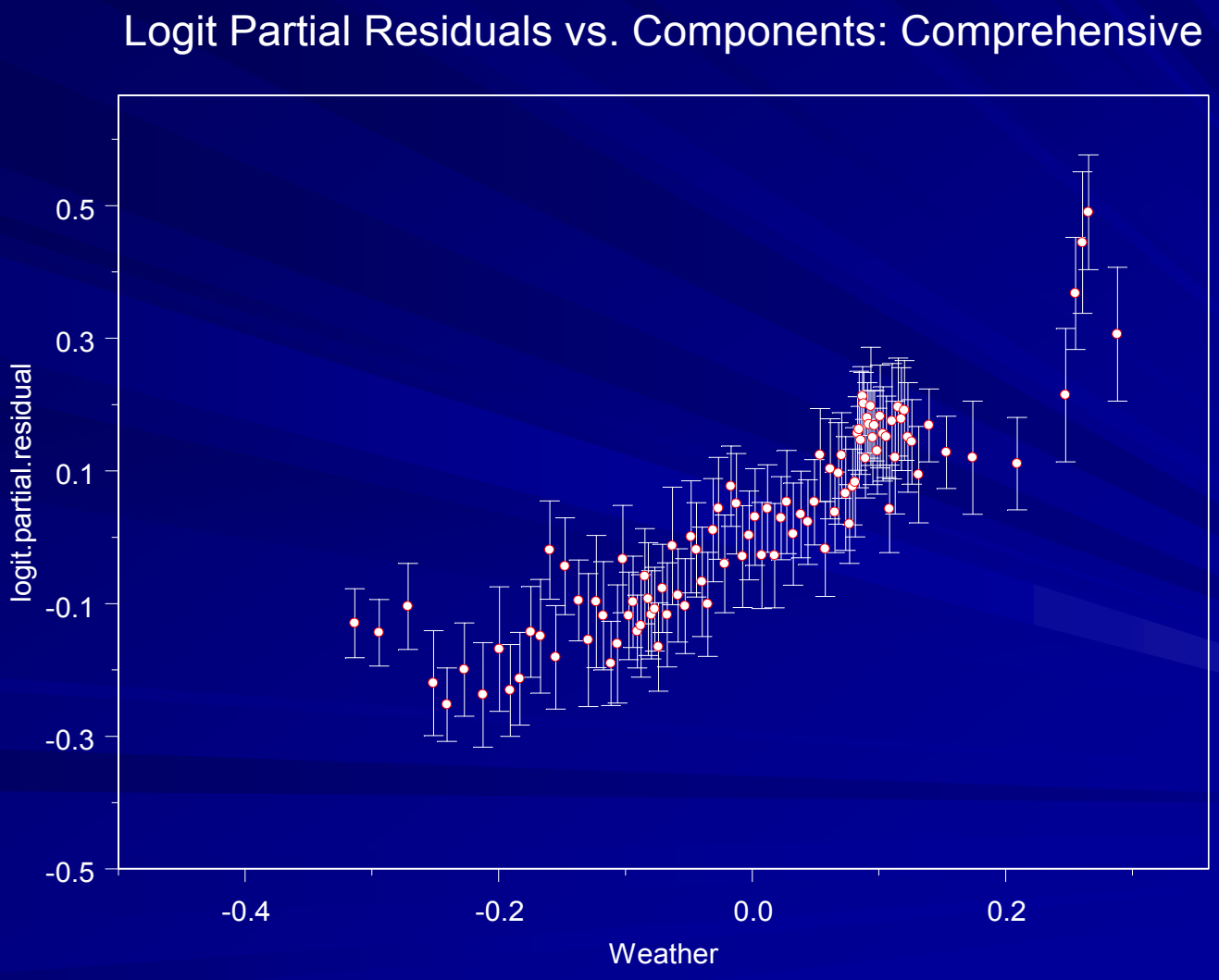
Component Diagnostics
Traffic Density



Component Diagnostics
Traffic Generators



Component Diagnostics Weather



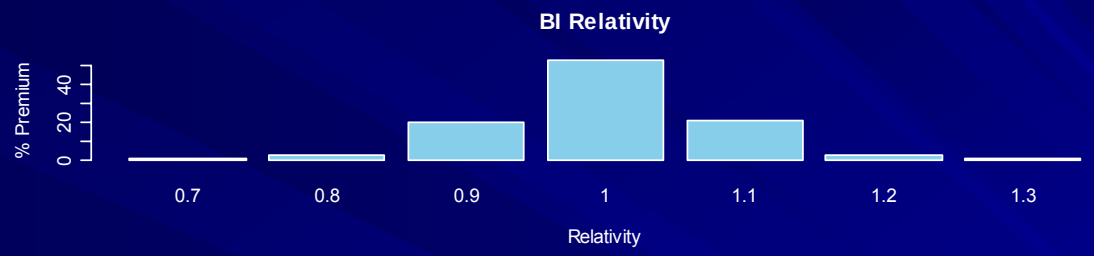
Comparing Model Output to Current Loss Costs

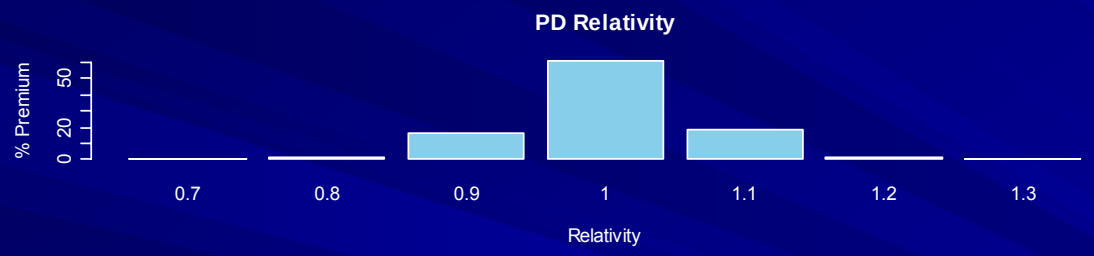
- Model output is deployed to a base class, standard limits and deductibles.
  - Similar to current loss cost, but at garaging address rather than territory.
- Define:

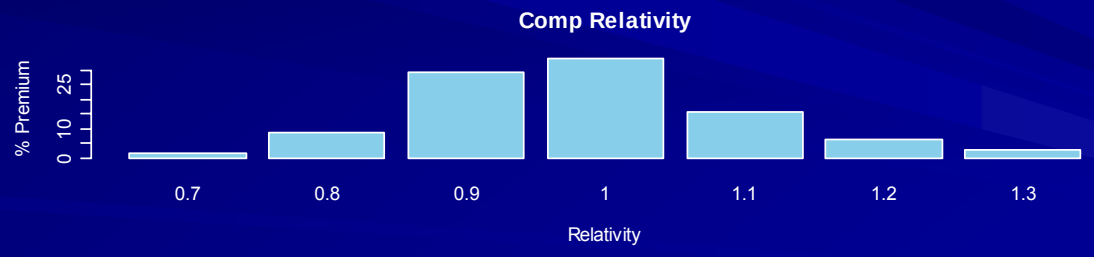
$$\text{Relativity} = \frac{\text{Model Output}}{\text{Current Loss Cost}}$$

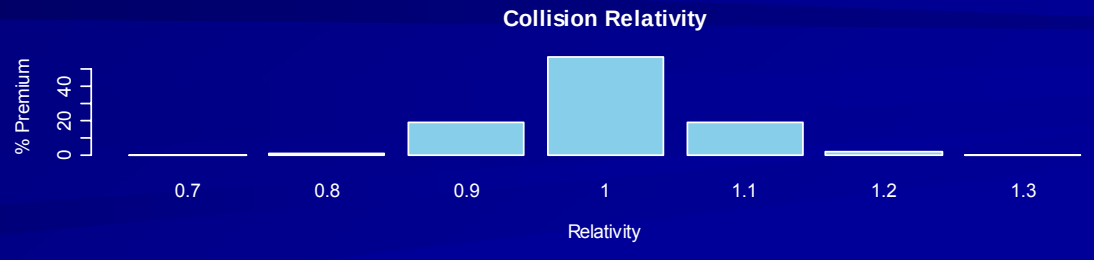
- Relativity is proportional to premium that could be charged with “refined loss costs” using the model output.

Relativities to Current Loss Costs

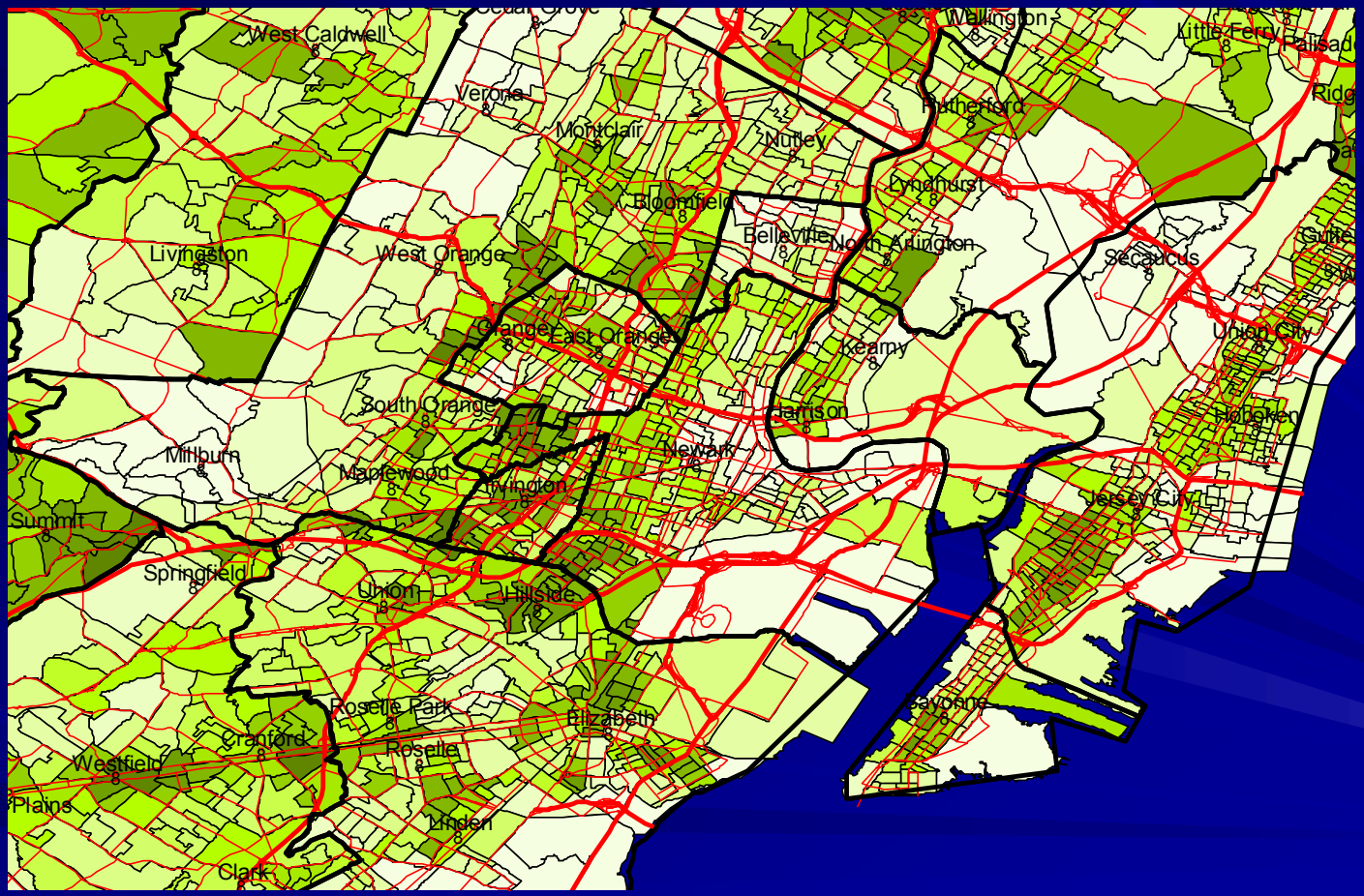








Newark NJ Area Combined Relativity

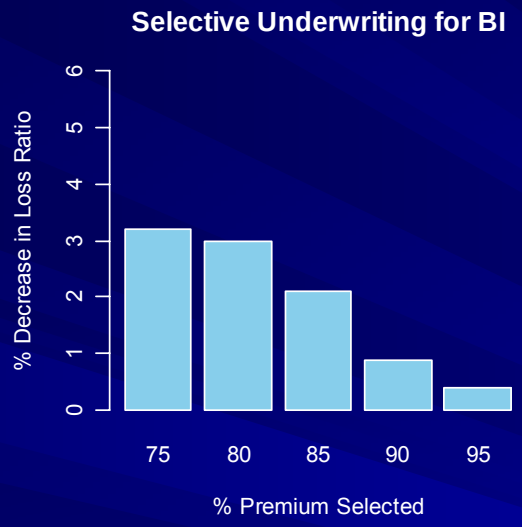


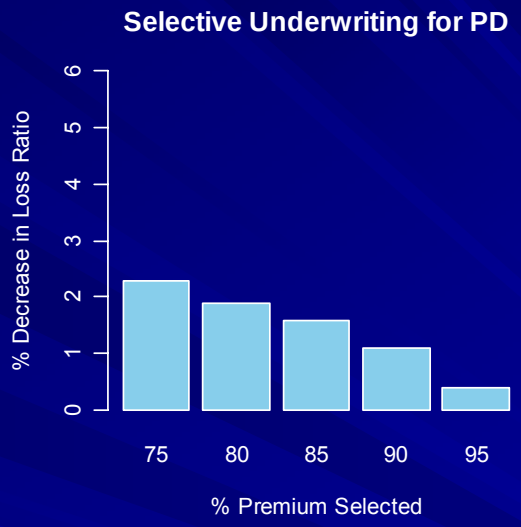
Evaluating the Lift of the Environmental Model

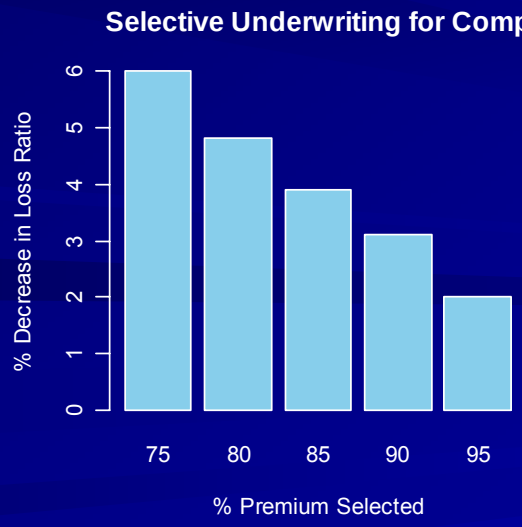
- Demonstrate the ability to select the more profitable risks
- Demonstrate the adverse effect of competitors “skimming the cream”
- Calculate the “Value of Lift” statistic

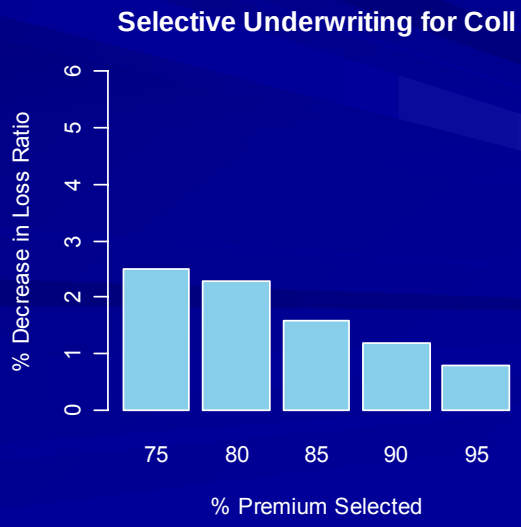
- Once insurers see the value of lift other actions are possible
  - Change prices (etc)

Effect of Selecting Lower Relativities

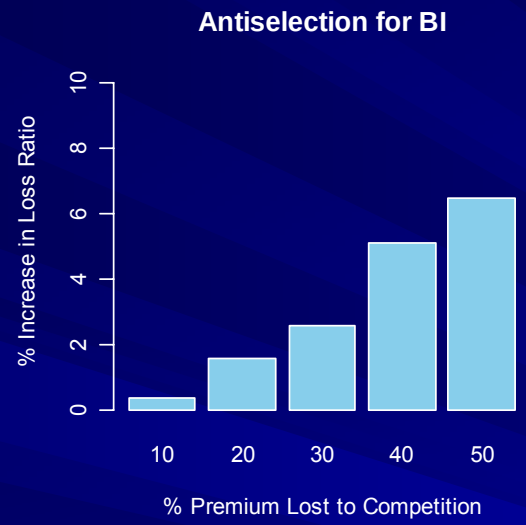


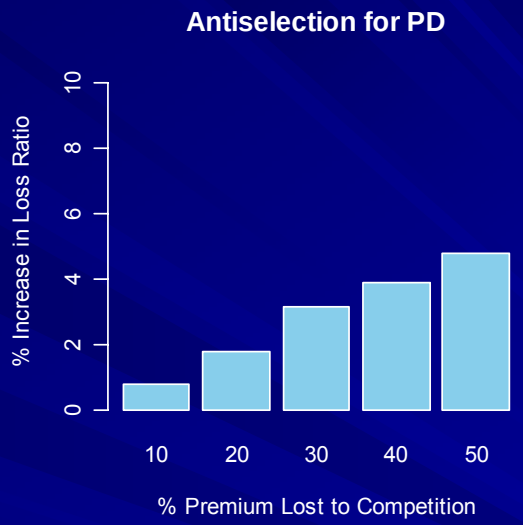


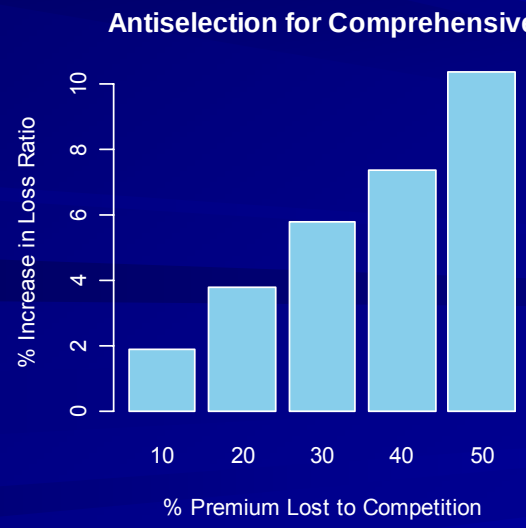


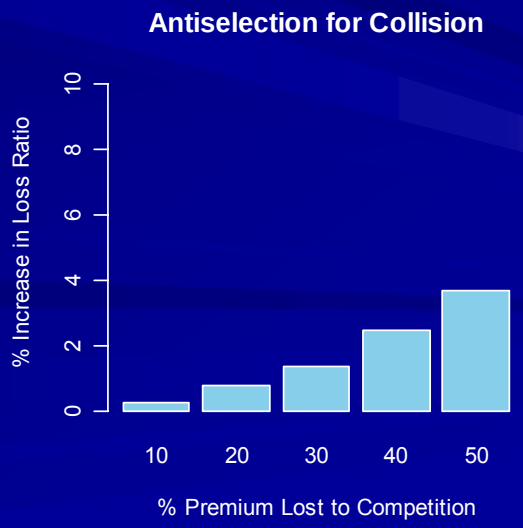


Effect of Competitors Selecting Lower Relativities









Assumptions of The Formula Value of Lift (VoL)

- Assume a competitor comes in and takes away the business that is less than your class average.
- Because of adverse selection, the new loss ratio will be higher than the current loss ratio.
- *What is the value of avoiding this fate?*
- VoL is proportional to the difference between the new and the current loss ratio.
- Express the VoL as a $ per car year.

The VoL Formula

- L_C = Current losses
- P_C = Current Loss Cost
- L_N = New losses of business remaining
  - After adverse selection
- P_N = New Loss Cost
  - After adverse selection
- E_C = Current exposure in car years

The VoL Formula

$$VoL = \frac{\left(\frac{L_N}{P_N} - \frac{L_C}{P_C} \right) \cdot P_N}{E_C}$$

- The numerator represents $ value of the potential cost of competitors skimming the cream.
- Dividing by E_C expresses this value as a $ value per car year.

Value of Lift Results on Pilot Testers

Coverage	Value of Lift
BI Liability	$4.99
PD Liability	$3.63
Collision	$1.61
Comprehensive	$4.85
PIP	$15.04
Combined	$13.29

Customized Model

Loss Cost = Pure Premium

= Frequency x Severity

$$\text{Frequency} = \frac{e^{\lambda}}{1+e^{\lambda}}$$

$\lambda = \alpha_0$

$+ \alpha_1 \cdot$ Weather

$+ \alpha_2 \cdot$ Traffic Density

$+ \alpha_3 \cdot$ Traffic Generators

$+ \alpha_4 \cdot$ Traffic Composition

$+ \alpha_5 \cdot$ Experience and Trend

+ Other Classifiers

$\alpha_1 \ldots \alpha_5 \equiv 1$
in industry model

Severity model
customized similarly

Summary

- Model estimates loss cost as a function of business, demographic and weather conditions.
- Demonstrated model diagnostics
- Demonstrated lift
- Indicated how to customize the model