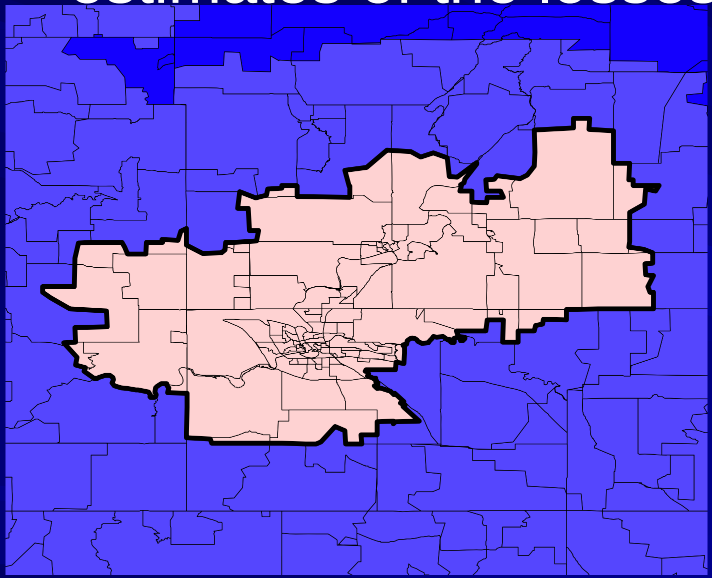
2009 CAS Ratemaking Seminar

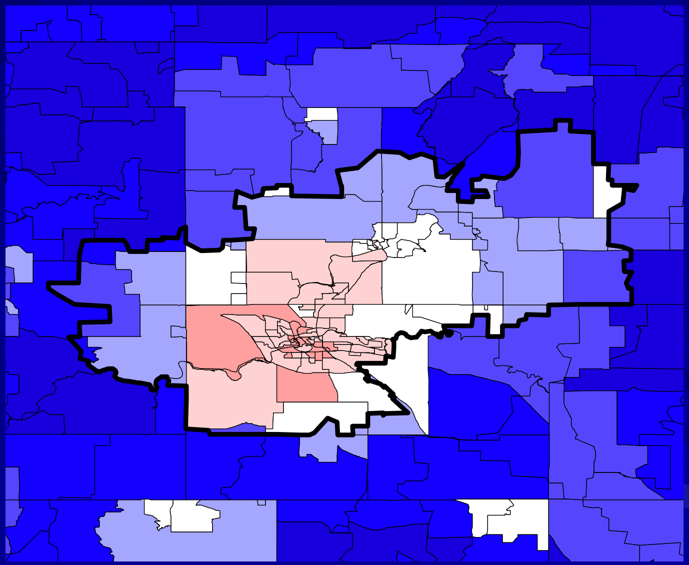
Estimating Loss Cost at the Address Level

Mark S. Richards
ISO Innovative Analytics

Territorial Conundrum

- **Territories should be big**
  - Have a sufficient volume of business to make credible estimates of the losses.





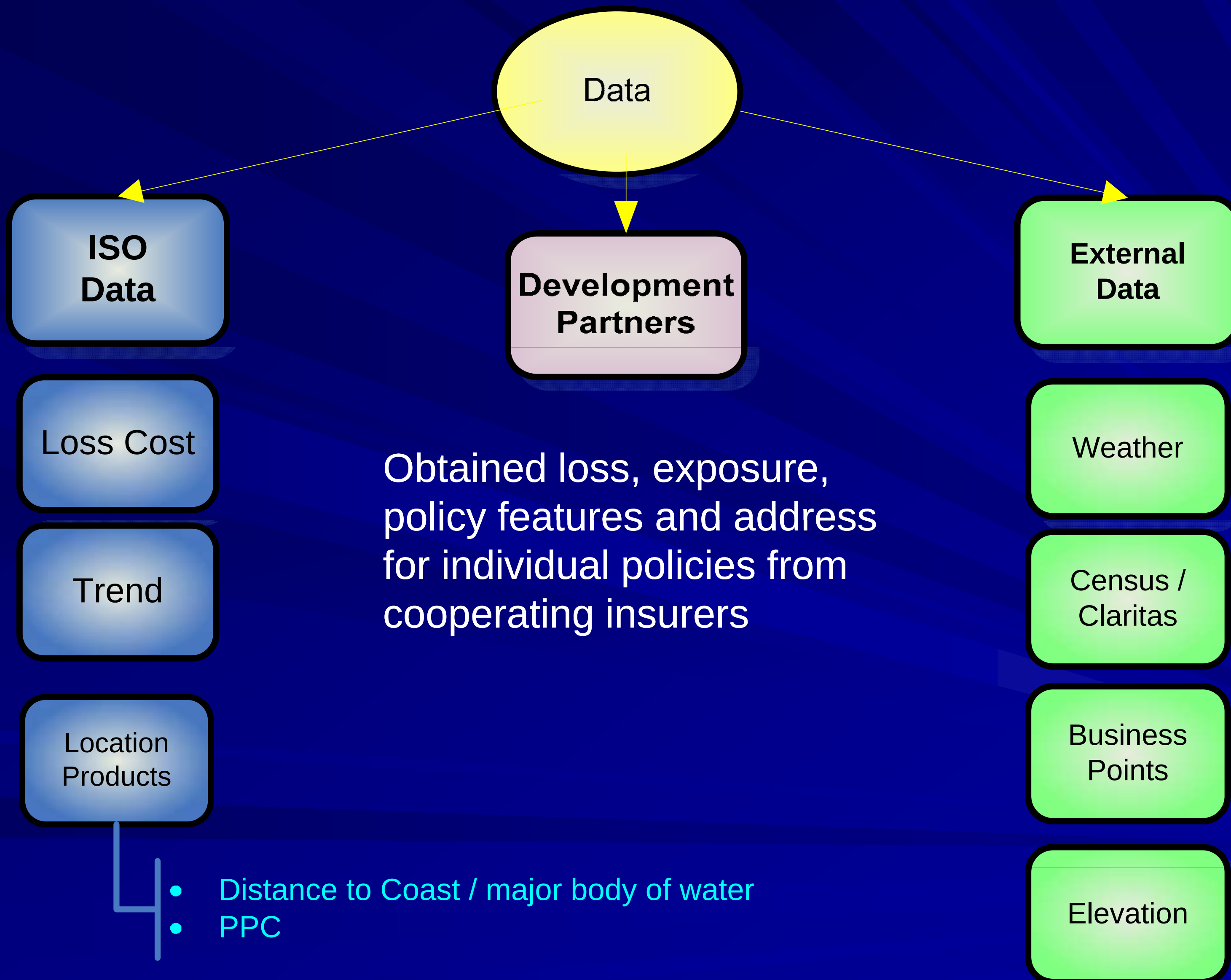
- **Territories should be small**
  - Conditions vary within territory.

View as Case Studies in Model Development

Data Driven Approach

- Reduction in number of variables
  - Necessary for small insurers
- Special circumstances in fitting models to individual auto / home owners data.
- Diagnostics
  - Graphics and Maps

Data Versus the Conundrum



Obtained loss, exposure, policy features and address for individual policies from cooperating insurers

- Distance to Coast / major body of water
- PPC

Some Environmental Features (Possibly) Related to Claims

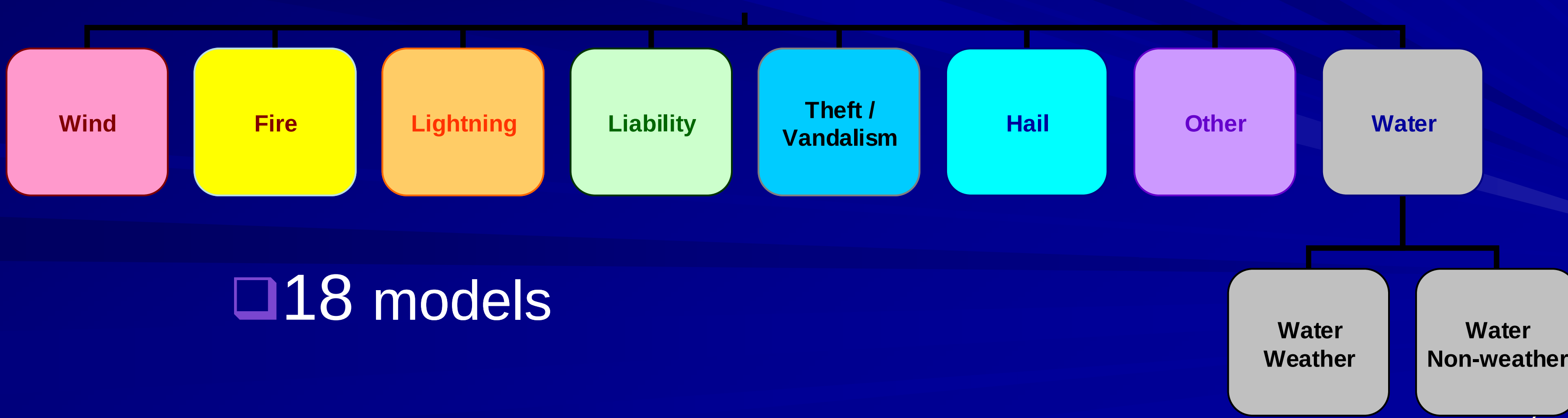
- Proximity to Businesses and Attractions
  - Workplaces, Shopping Centers, Contractors, etc.
- Weather / Terrain: Wind, Temperature, Snowfall, Change in Elevation
- Population (Traffic) Density
- Others : Commuting Patterns, Coastal proximity, etc.

Combining Environmental Variables at a Particular Address

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

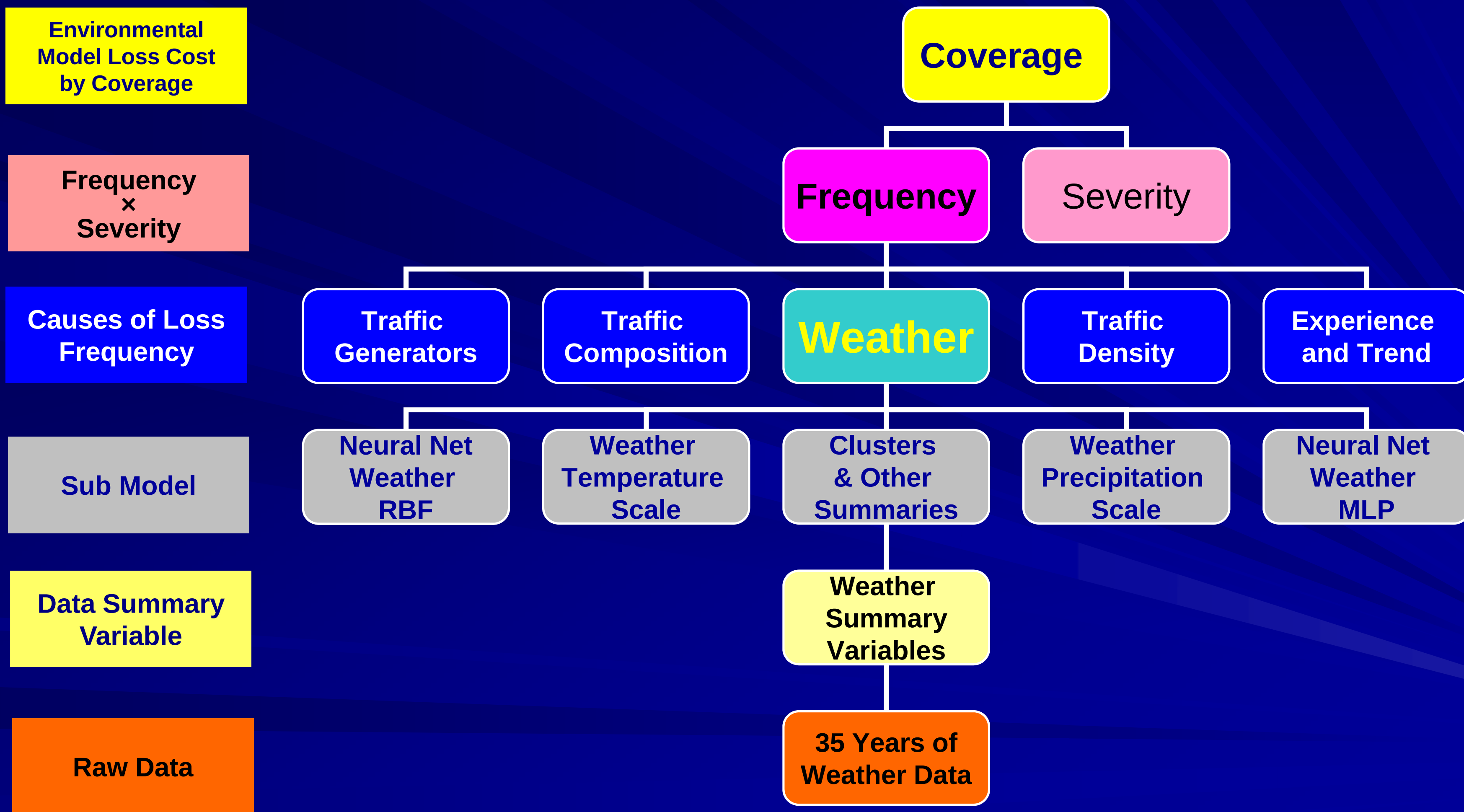
Variable Selection is Multiplied by the Number of Models

- Frequency and Severity are modeled separately
- Models are at coverage / peril level
  - Five auto coverages: BI, PD, PIP, Comp. & Coll.
    - 10 models
  - Nine home owners perils:



- 18 models

In Depth for Auto Weather Component



Environmental Model

Loss Cost = Pure Premium

= Frequency x Severity

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

λ = Intercept
+ Weather
+ Traffic Density
+ Traffic Generators
+ Traffic Composition
+ Experience and Trend

μ = Intercept
+ Weather
+ Traffic Density
+ Traffic Generators
+ Traffic Composition
+ Experience and Trend

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

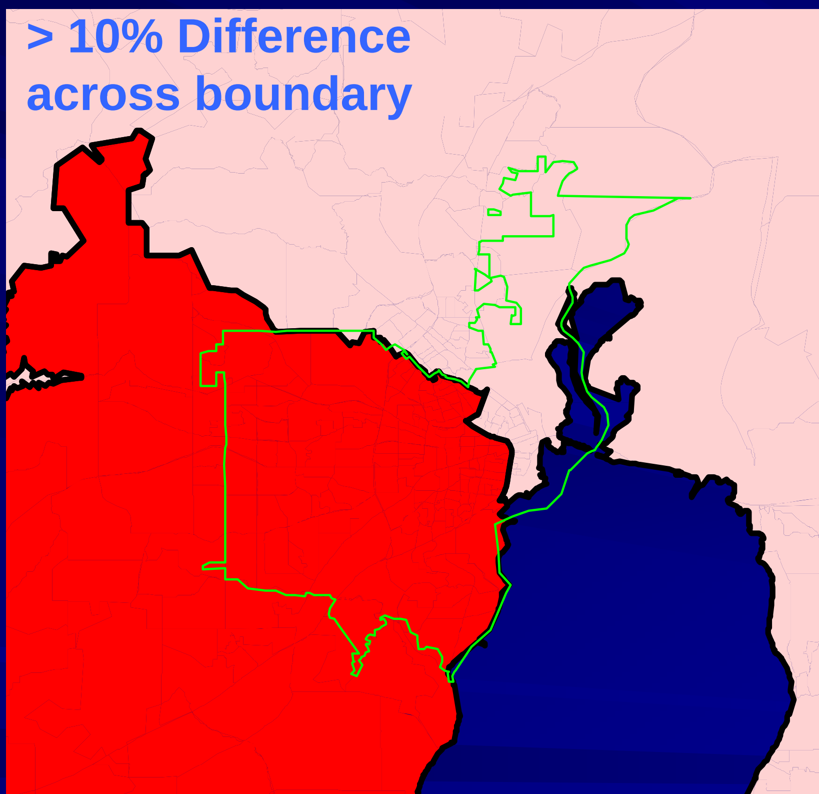
$+ \alpha_{n_2+1} \cdot X_{n_2+1} + \ldots + \alpha_{n_3} \cdot X_{n_3}$ = Traffic Generators

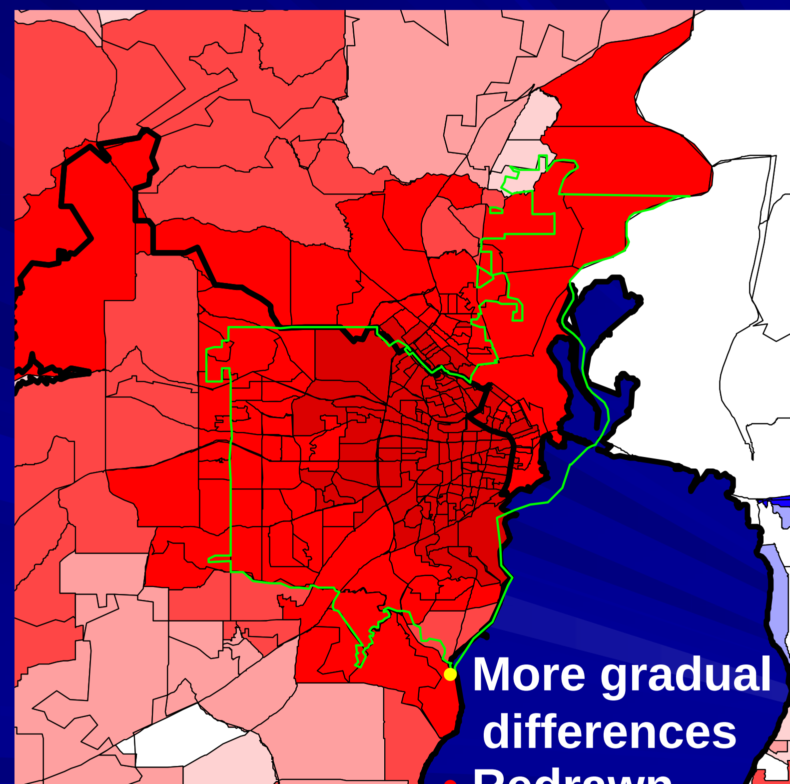
$+ \alpha_{n_3+1} \cdot X_{n_3+1} + \ldots + \alpha_{n_4} \cdot X_{n_4}$ = Traffic Composition

$+ \alpha_{n_4+1} \cdot X_{n_4+1} + \ldots + \alpha_{n_5} \cdot X_{n_5}$ = Experience & Trend

$+$ Other Classifiers

An Example on the Ground



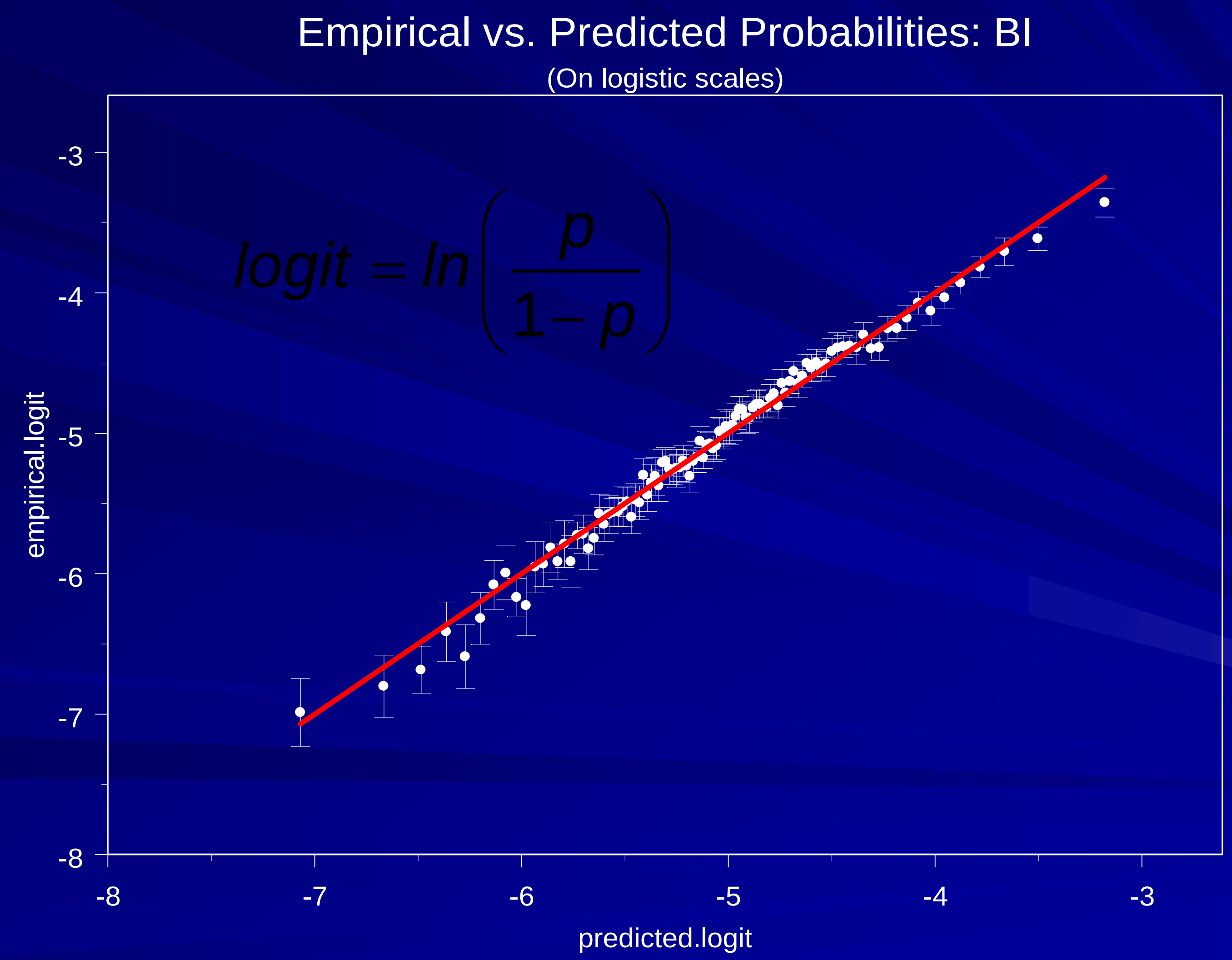


- More gradual differences
- Redrawn boundaries

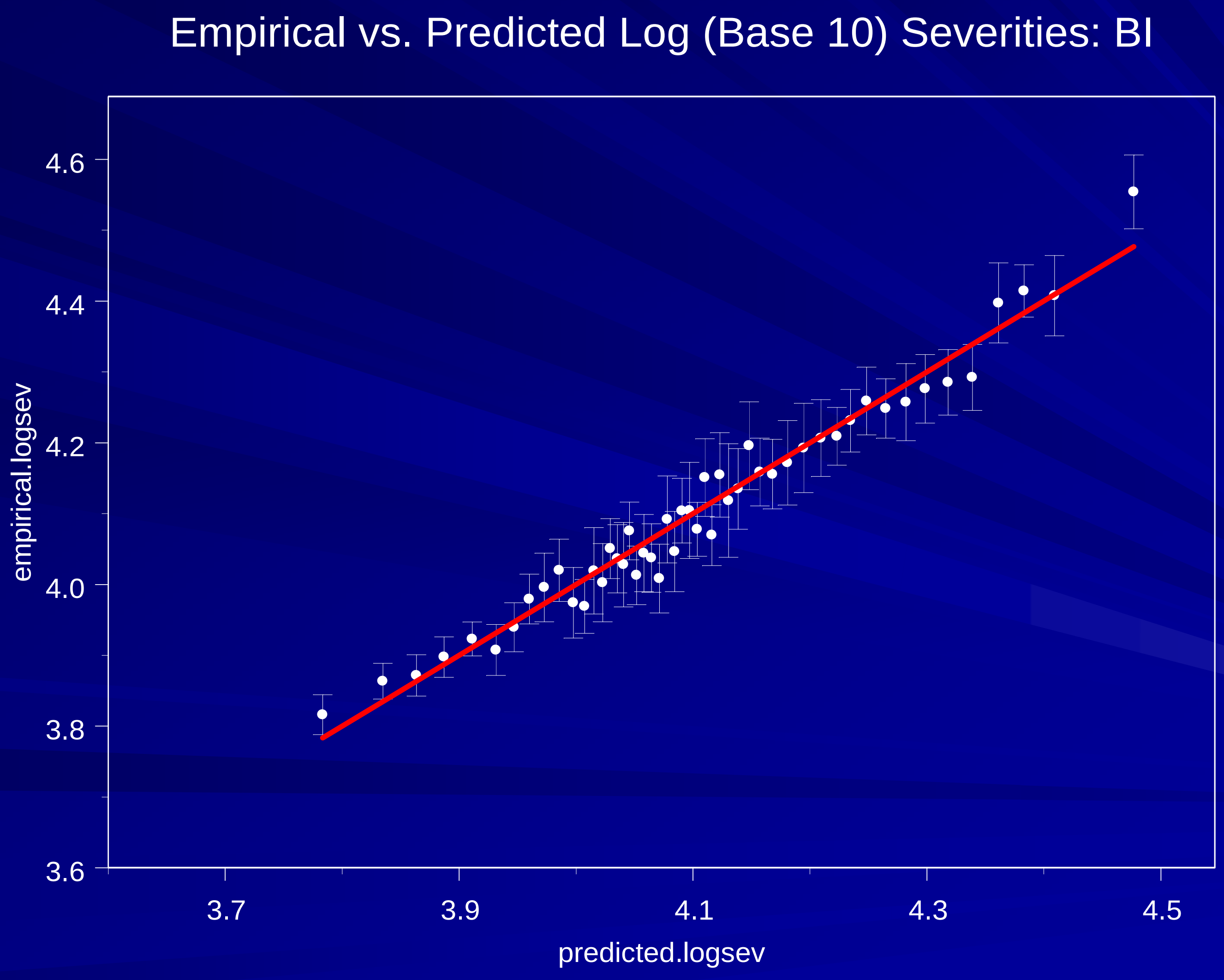
Overall Model Diagnostics

- 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



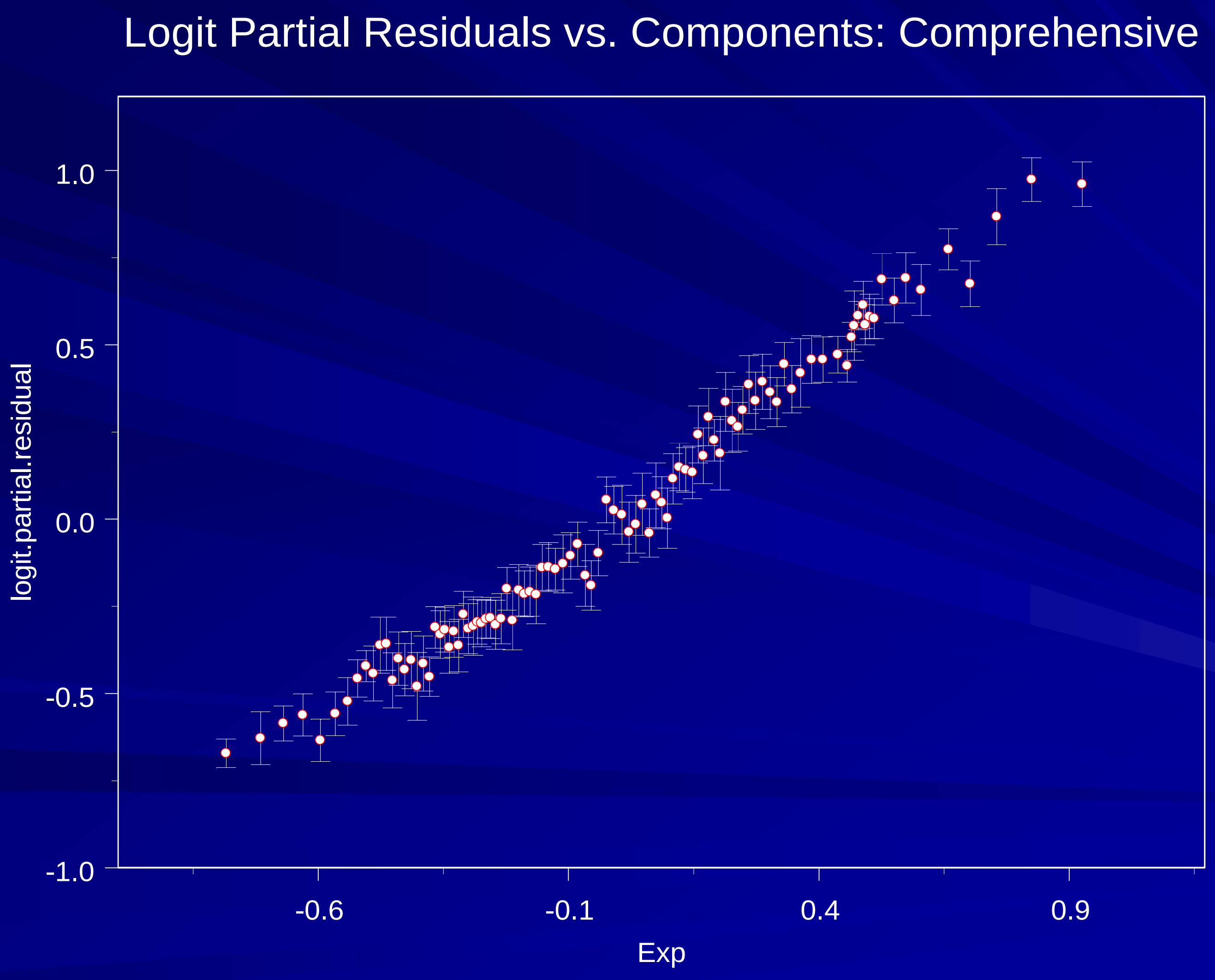
Component Diagnostics Frequency Example

- Sort observations in order of i^{th} component 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

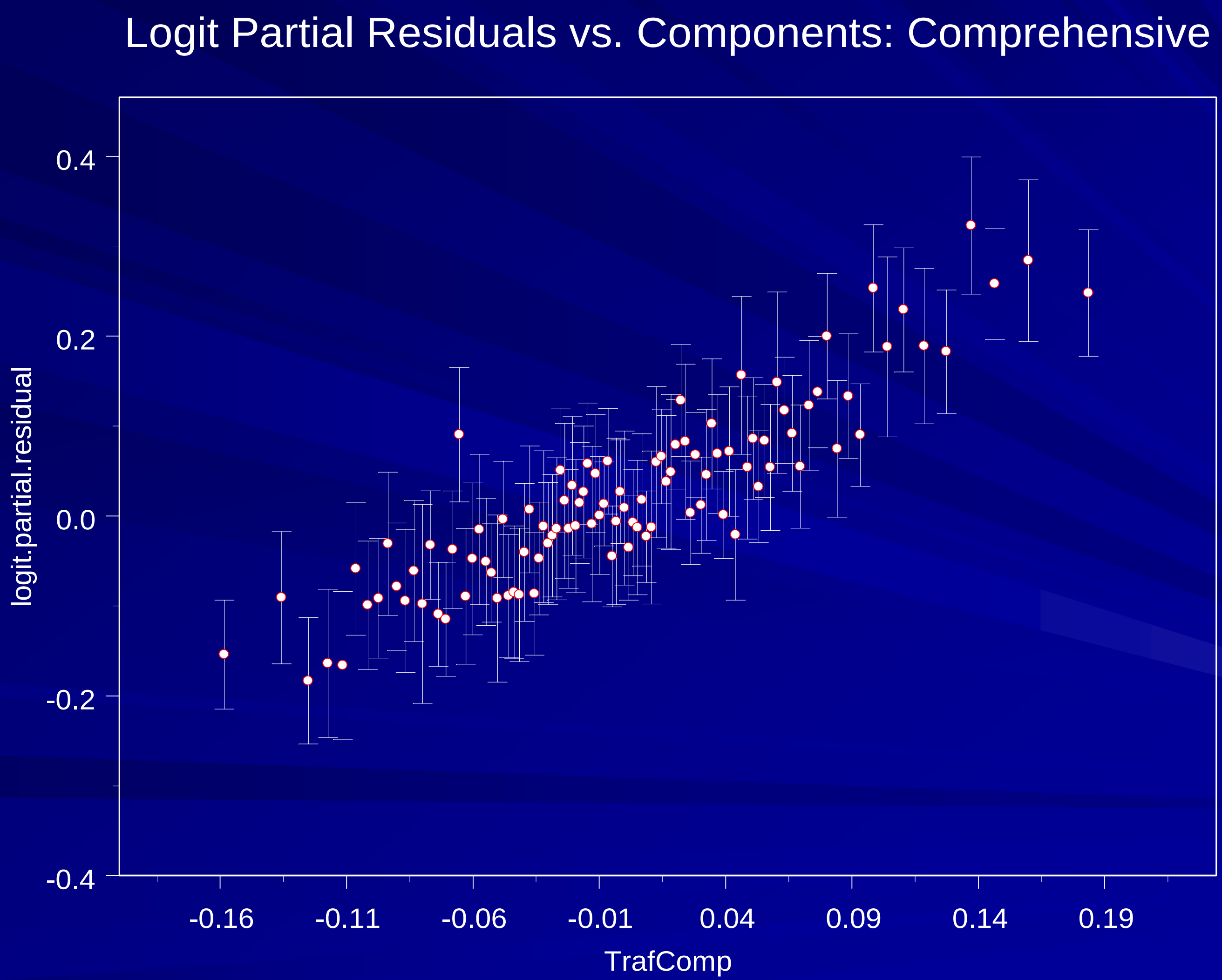
$$R_{ib} = ln\left(\frac{p_{ib}}{1-p_{ib}}\right) - \left(\lambda + \sum_{k \neq i} C_{kb}\right)$$

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

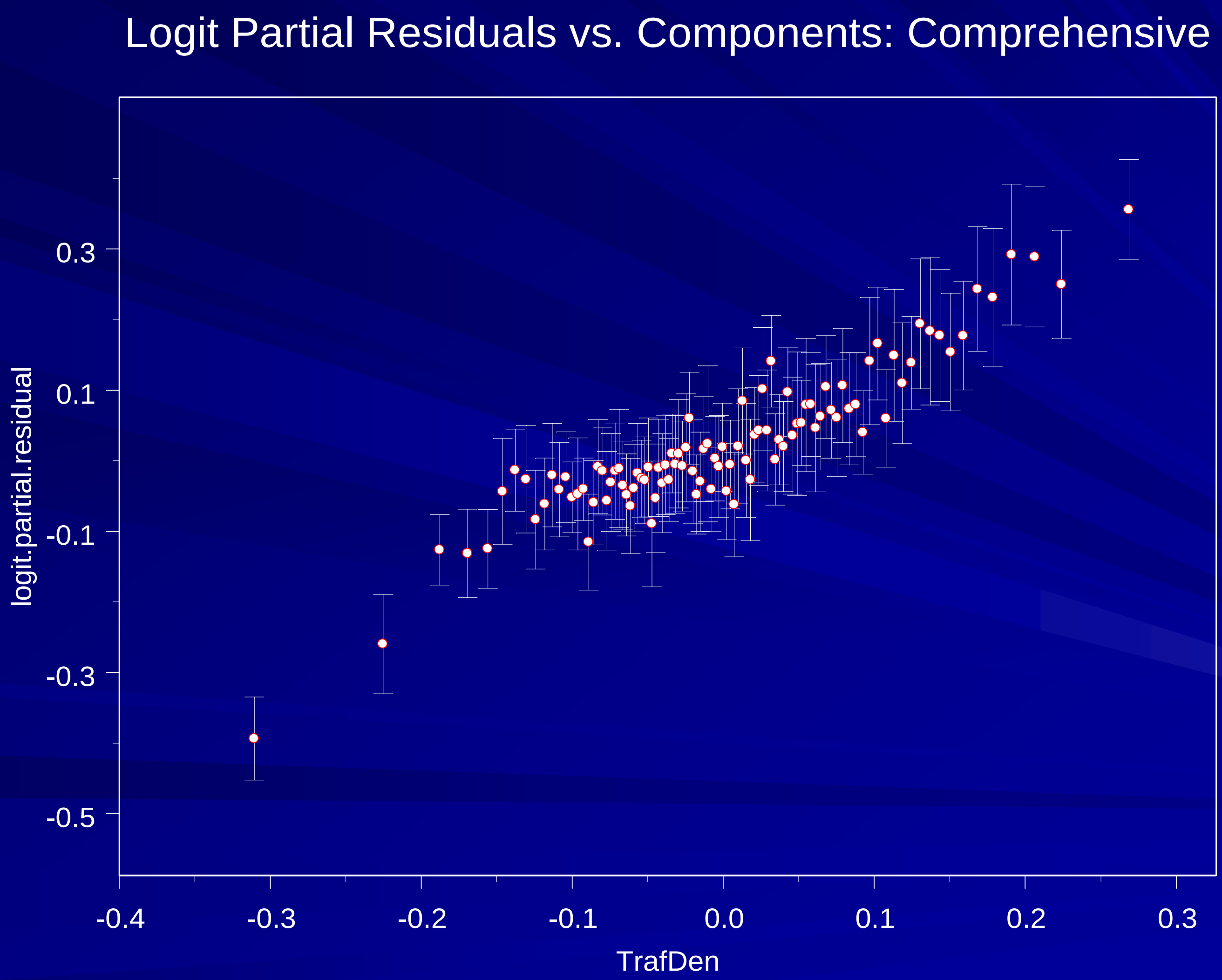
Component Diagnostics
Experience and Trend



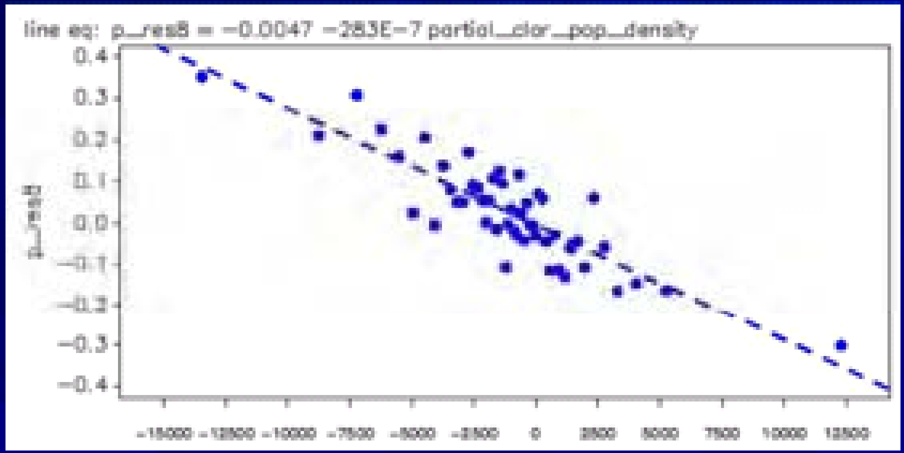
Component Diagnostics
Traffic Composition

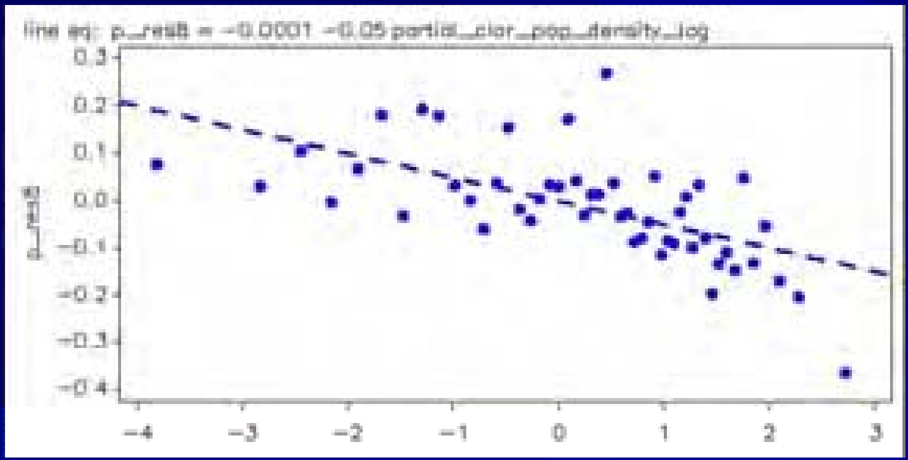


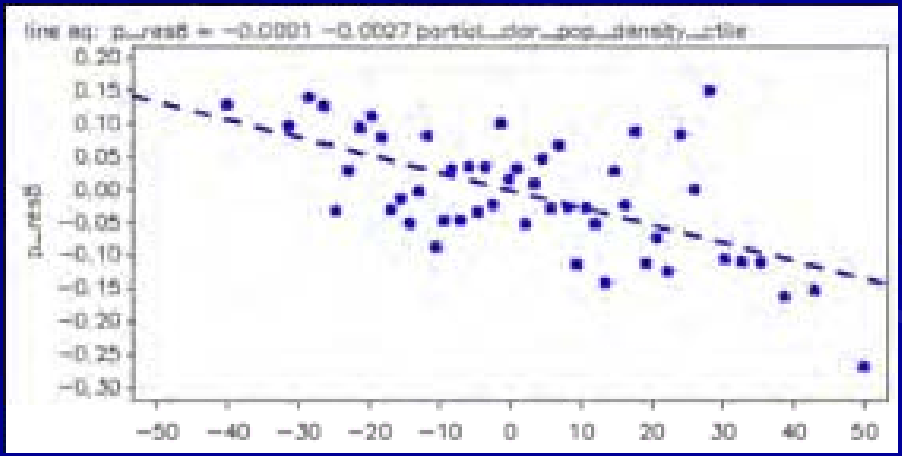
Component Diagnostics
Traffic Density



Partial Residual Plot : Finding Transformations







Example of Diagnostics Collinearity and Multicollinearity

- Correlations Matrix: measures the correlations among each pair of variables in the models, but does not consider multicollinearity.

- Variance Inflation Factors (VIF): A measure of the multicollinearity among independent variables.

Customized Model

Loss Cost = Pure Premium

= Frequency x Severity

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

$$\begin{aligned}\lambda = \alpha_0 & \\ & + \alpha_1 \cdot \text{Weather} \\ & + \alpha_2 \cdot \text{Traffic Density} \\ & + \alpha_3 \cdot \text{Traffic Generators} \\ & + \alpha_4 \cdot \text{Traffic Composition} \\ & + \alpha_5 \cdot \text{Experience and Trend} \\ & + \text{Other Classifiers}\end{aligned}$$

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

Severity model
customized similarly

Summary

- Model estimates loss cost as a function of business, demographic and weather conditions associated with address.
- Preparing data for models based on geography is not a trivial exercise
- Showed fit assessment and model diagnostics
- Indicated how to customize the model