Vehicle Symbol Development

Serhat Guven
Senior Consultant

September 2009

Vehicle Symbol Analysis

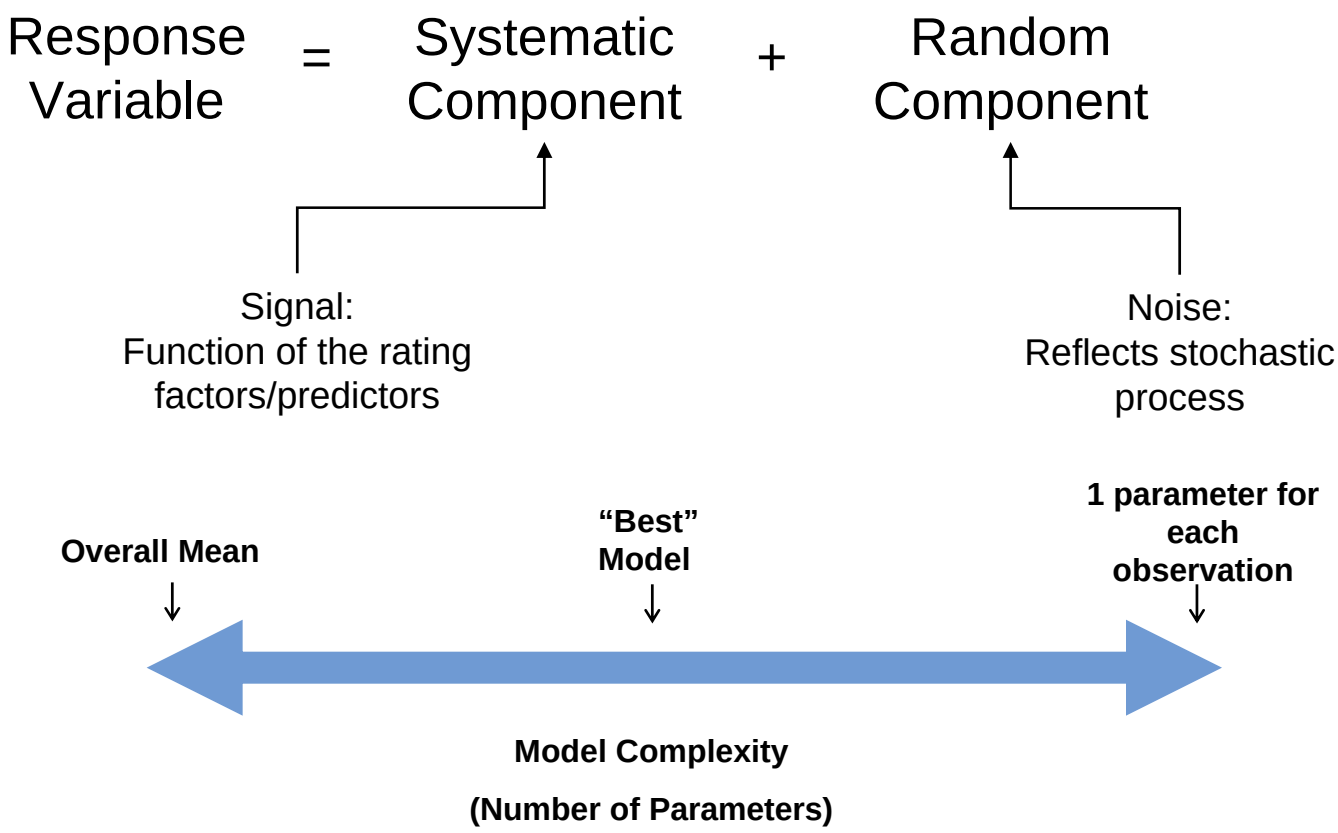
PURPOSE: To discuss techniques for creating proprietary symbols within the context of multivariate review, including proper tools and diagnostics

OUTLINE

- Background
- Vehicle Estimator
  - Initial Estimator
  - Diagnostics
  - Vehicle Spatial Correction
- Vehicle Symbols
- Vehicle Relativities
- Summary

Background

Basics of Predictive Modeling:



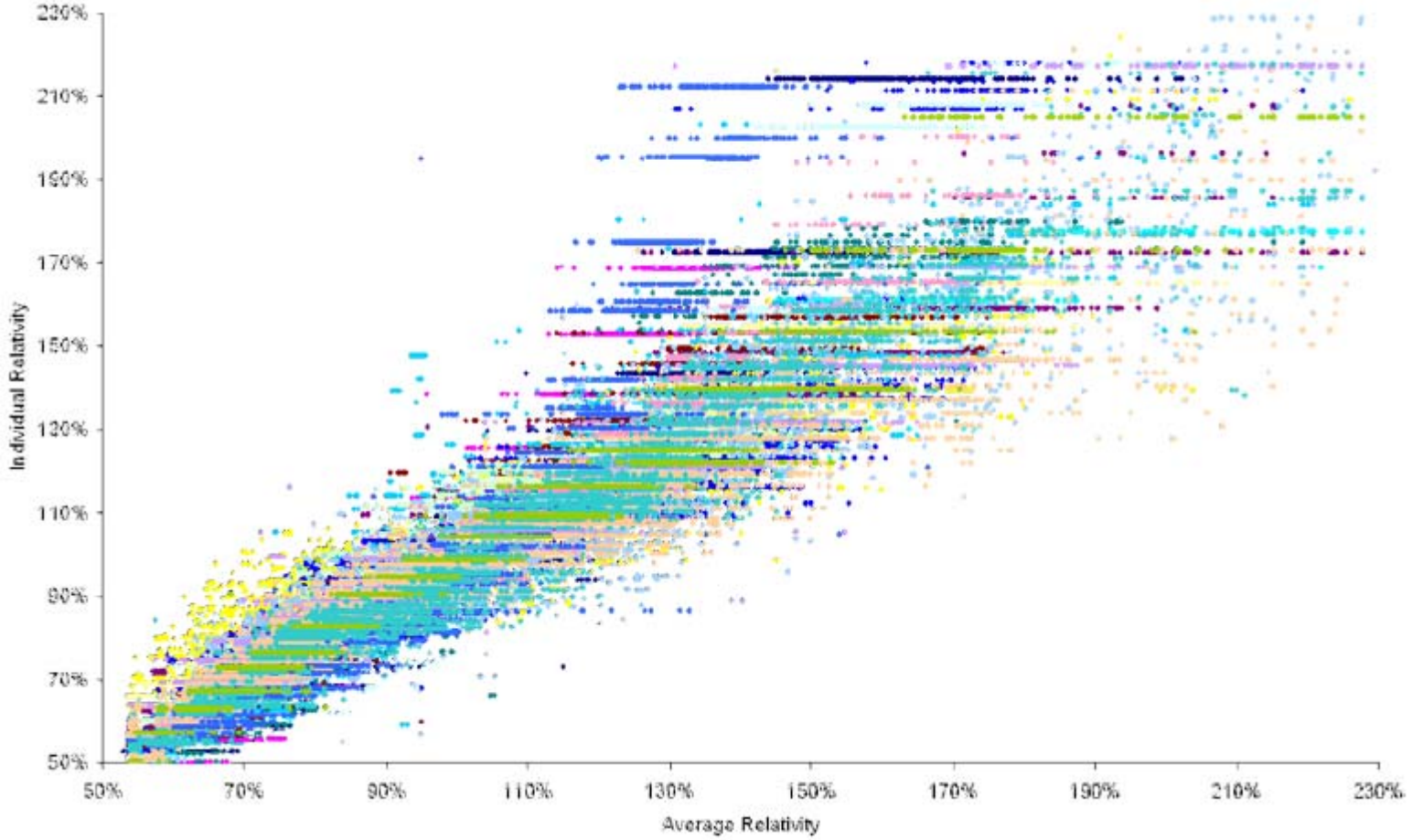
Background

Vehicle Symbol/Relativity Analysis:

- Vehicle is a major risk driver and accounts for much rate variation
- Two elements:
  - Symbols
  - Relativities
- Historically, focus on relativities and not symbols
  - Symbols on limited data, anecdotal evidence, competitors, bureaus, and judgment
  - Liability symbols relatively recent phenomenon
  - Regular reviews of relativities with “tweaking” symbols

Background

Significant Variation in Pricing by Vehicle:



Background

Issues with Symbol Analysis:

- High-dimensionality
  - Vehicle symbol requires a large number of small units as building blocks
  - Companies have limited data, so many building blocks have little or no claims experience
- High correlation
  - Symbols tends to be highly correlated with other rating variables (e.g., age, credit, tier, etc)
  - Multivariate framework required to handle highly correlated variables

Background

Vehicle Symbol/Relativity Analysis

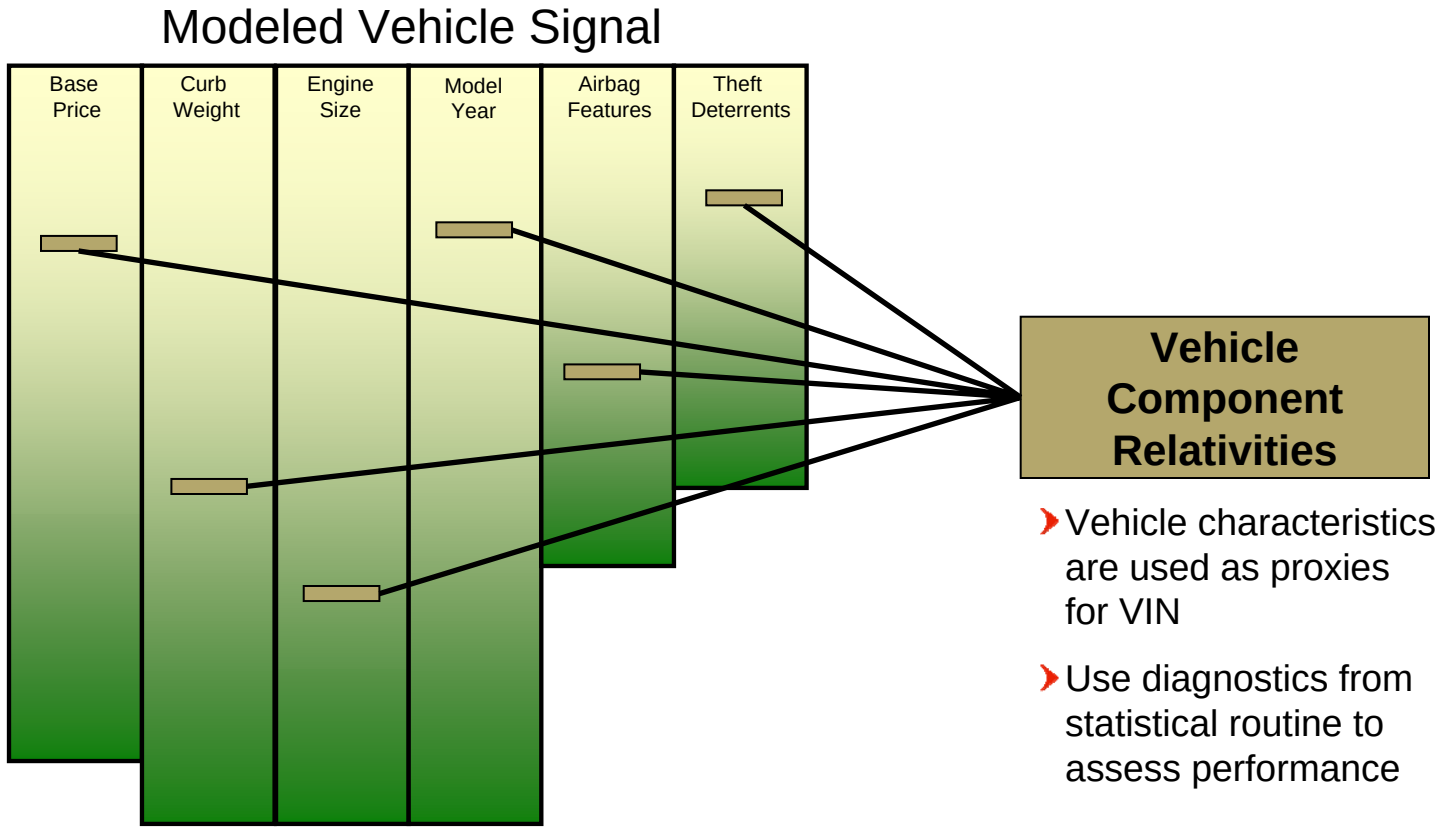
- Vehicle symbol analysis is a multi-stage process.

- Standardized fitted value from multivariate model
- By coverage or cause of loss.
- Frequency and severity determined separately
- Spatial corrections
- Component estimators combined
- Overall estimators clustered to form symbols
- Relativities calculated for each symbol

Vehicle Estimator

Initial Estimator:

- Component models built using vehicle characteristics



- Vehicle characteristics are used as proxies for VIN
- Use diagnostics from statistical routine to assess performance

Vehicle Estimator

Diagnostics:

- Need to identify additional signal in residual that can be attributed to the vehicle dimension
- Basic residual definition

$$\text{Residual}_{\text{VIN}} = \frac{\sum_{i \in \text{VIN}} \text{Actual}_i}{\sum_{i \in \text{VIN}} \text{Expected}_i}$$

- Issues with raw residual
  - VINs are very small building blocks with limited data
  - Unordered categorical data creates interpretation problems
- Use ‘spatial smoothing’ concepts to investigate signals

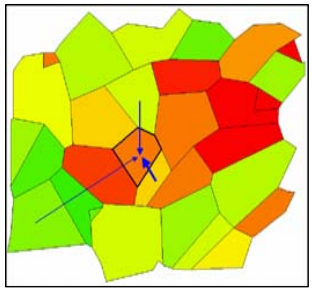
Vehicle Estimator

Diagnostics:

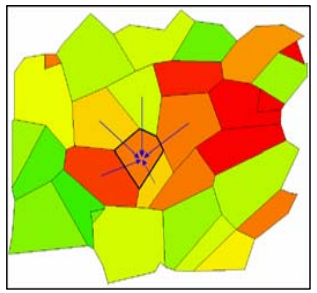
- High number of dimensions (e.g. make/model)
- May by only a few observations for any one dimension

Similar issue for boundary analysis solved via smoothing of results

- Distance-based smoothing
- Adjacency-based smoothing



Distance

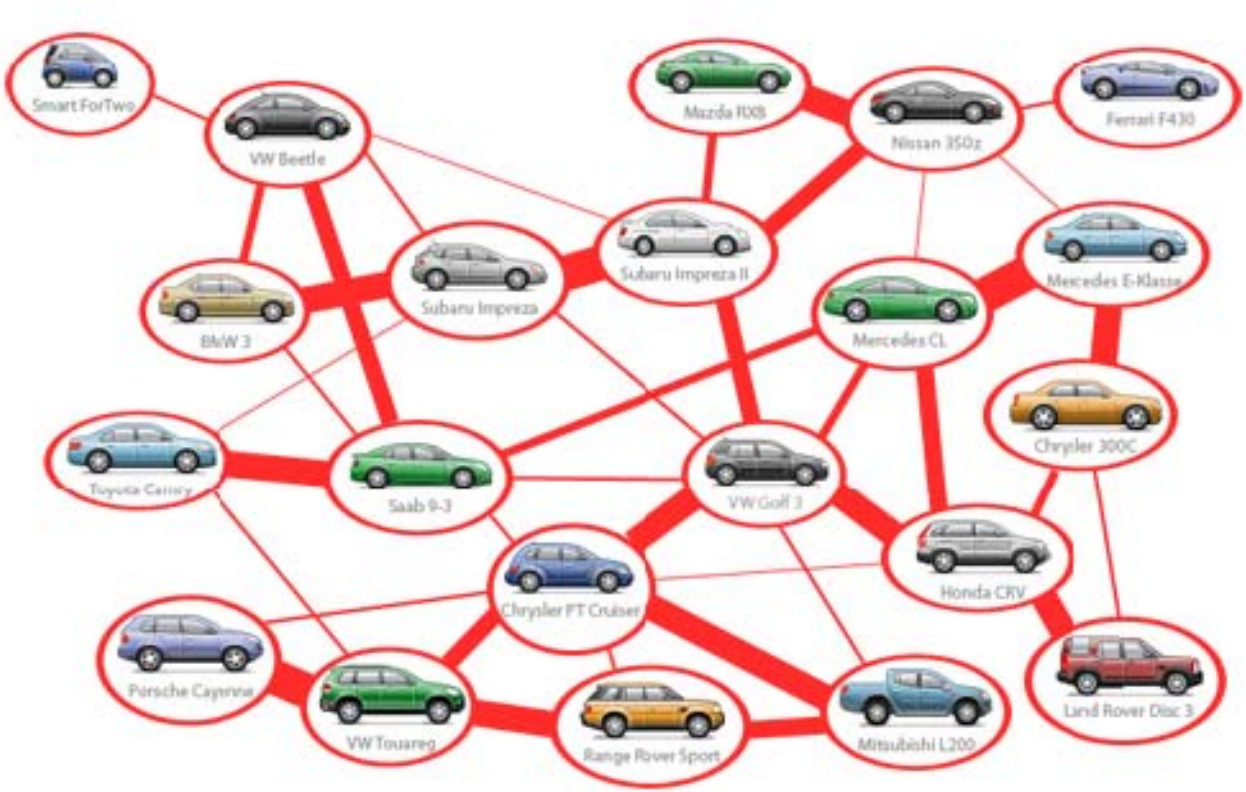


Adjacency

Smoothing techniques can be applied to vehicle residuals to identify signal

“Neighbor” Vehicles

Instead of using latitude/longitude to build adjacency relationships, use vehicle dimensions

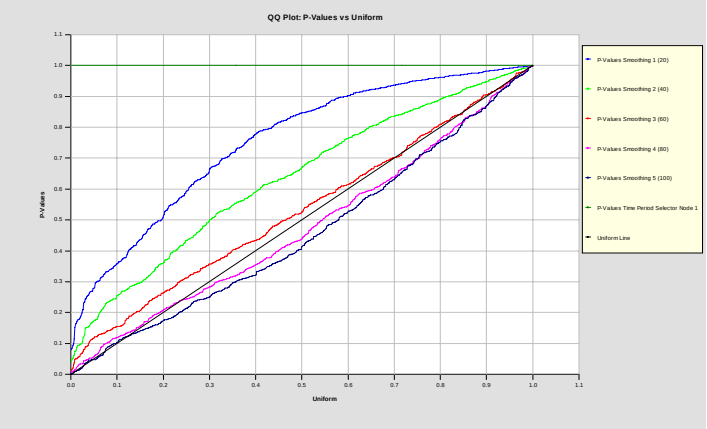


Once the “neighbors” are determined, the same techniques used for territorial analysis can be applied

Vehicle Estimator

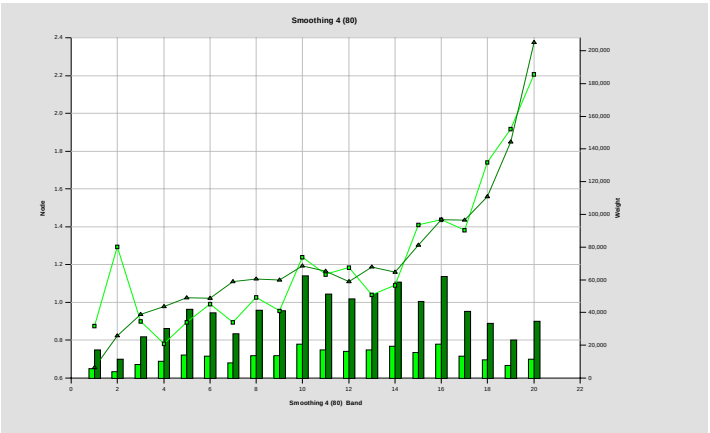
Diagnostics:

- Smoothing and diagnostics used to identify signal in the vehicle residual



- QQ plot on smoothed residual indicates potential signal
- If no signal, then all estimated cdfs would be above uniform cdf

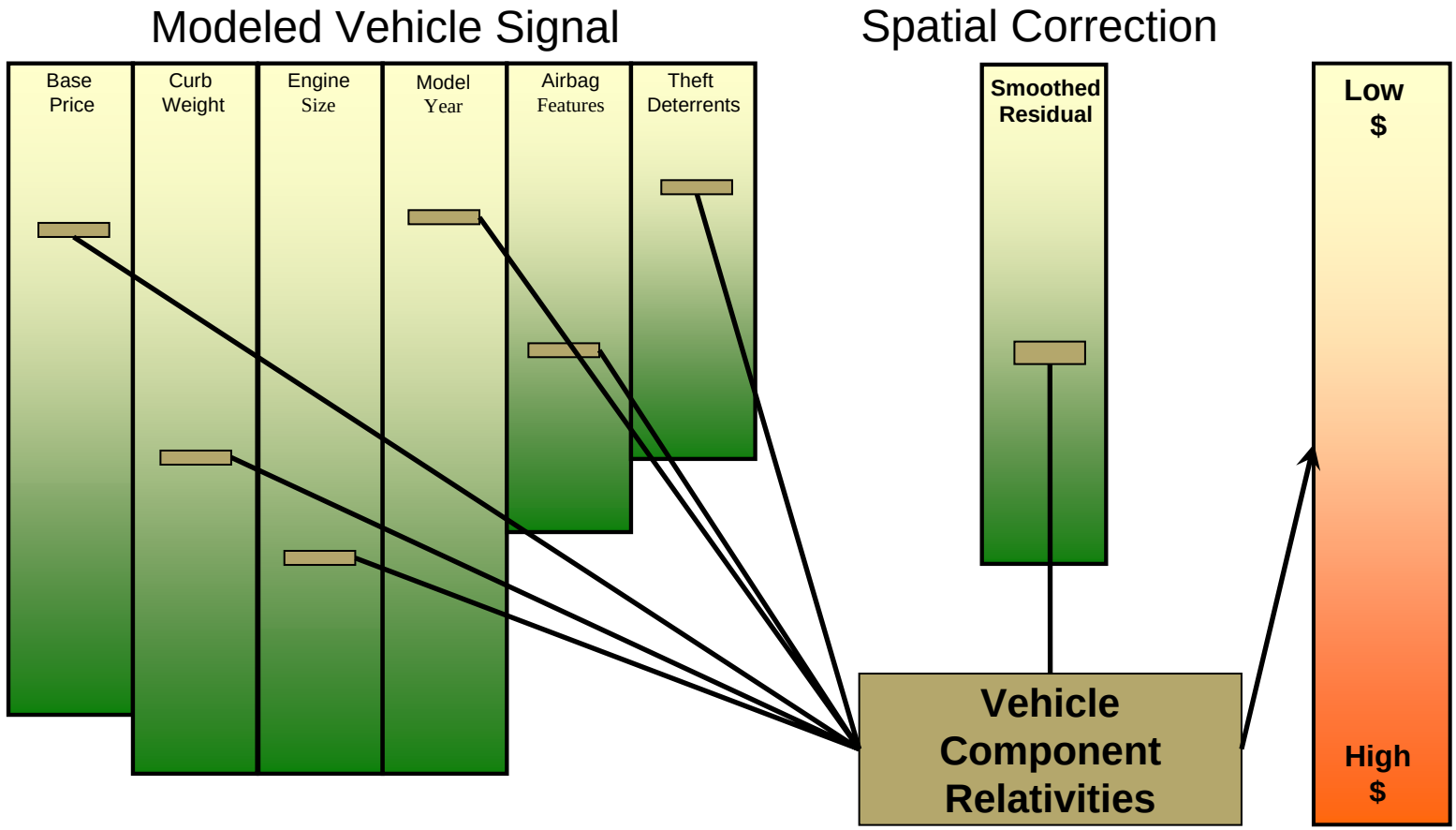
- Consistency tests further support signal identification
- Both time and data consistency tests should be performed



Vehicle Estimator

Vehicle Spatial Correction:

- Residual signal used to adjust scores from multivariate model



Vehicle Symbols

Clustering Basics

- Clustering used to produce groupings that are predictive of the future:
  - Minimize within-group heterogeneity.
  - Maximize cross-group heterogeneity.
- Commonly-used clustering methods:
  - Quantiles
  - Equal Weight
  - Similarity Methods
    - Average Linkage
    - Centroid

Vehicle Symbols

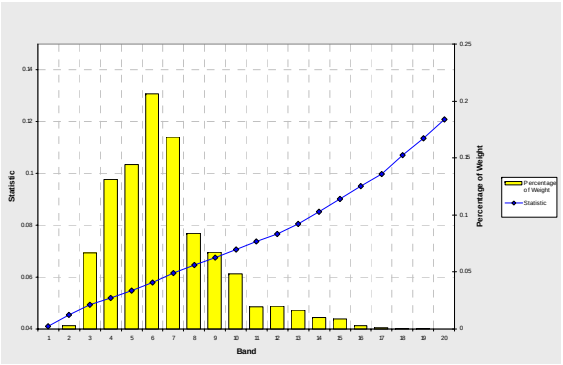
Clustering Methodologies:

- Quantiles
  - Create groups with an equal number of observations
- Equal Weight
  - Create groups which have an equal amount of weight
- Similarity Methods:
  - Rank the data set by the statistic you wish to cluster
  - Decide on which pair of records are the ‘most similar’
  - Group these records
  - Repeat until left with the desired number of groups

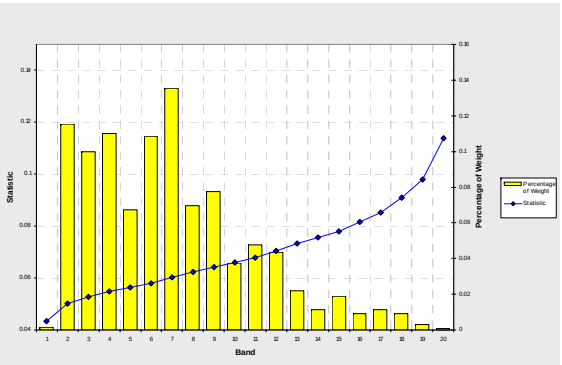
Vehicle Symbols

Similarity Methods

- Average Linkage
  - Difference between clusters is the average difference between pairs of observations, one in each cluster
  - Tends to join clusters with small variances
- Centroid
  - Difference between clusters is the difference between the mean values of the clusters squared
  - VINs allocated to more extreme groups



Average Linkage

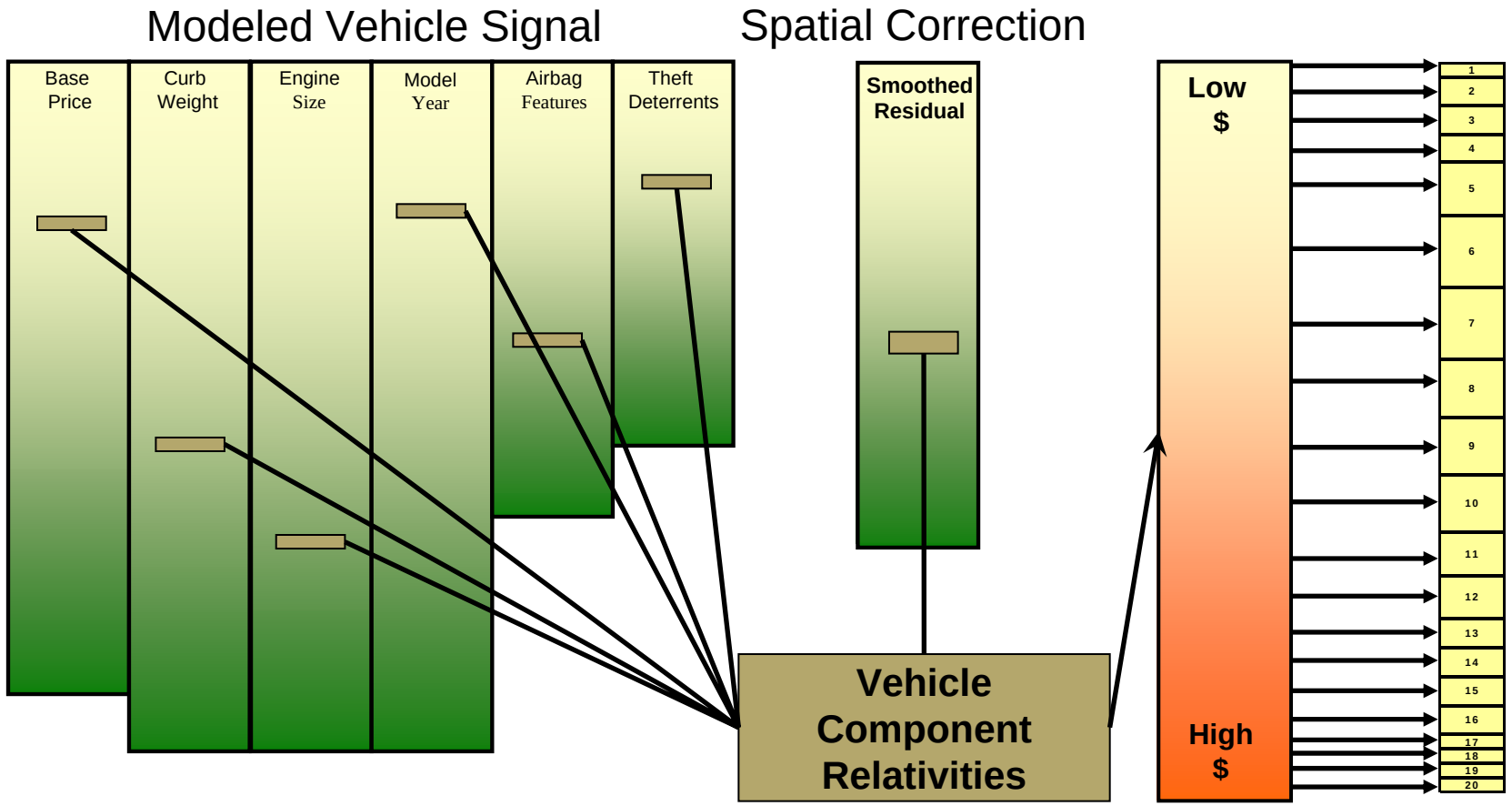


Centroid

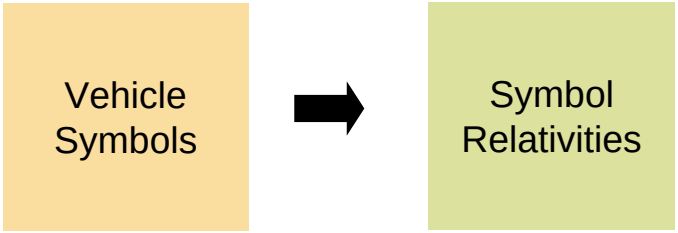
Vehicle Symbols

Clustering:

- Vehicle estimator is clustered into new symbols

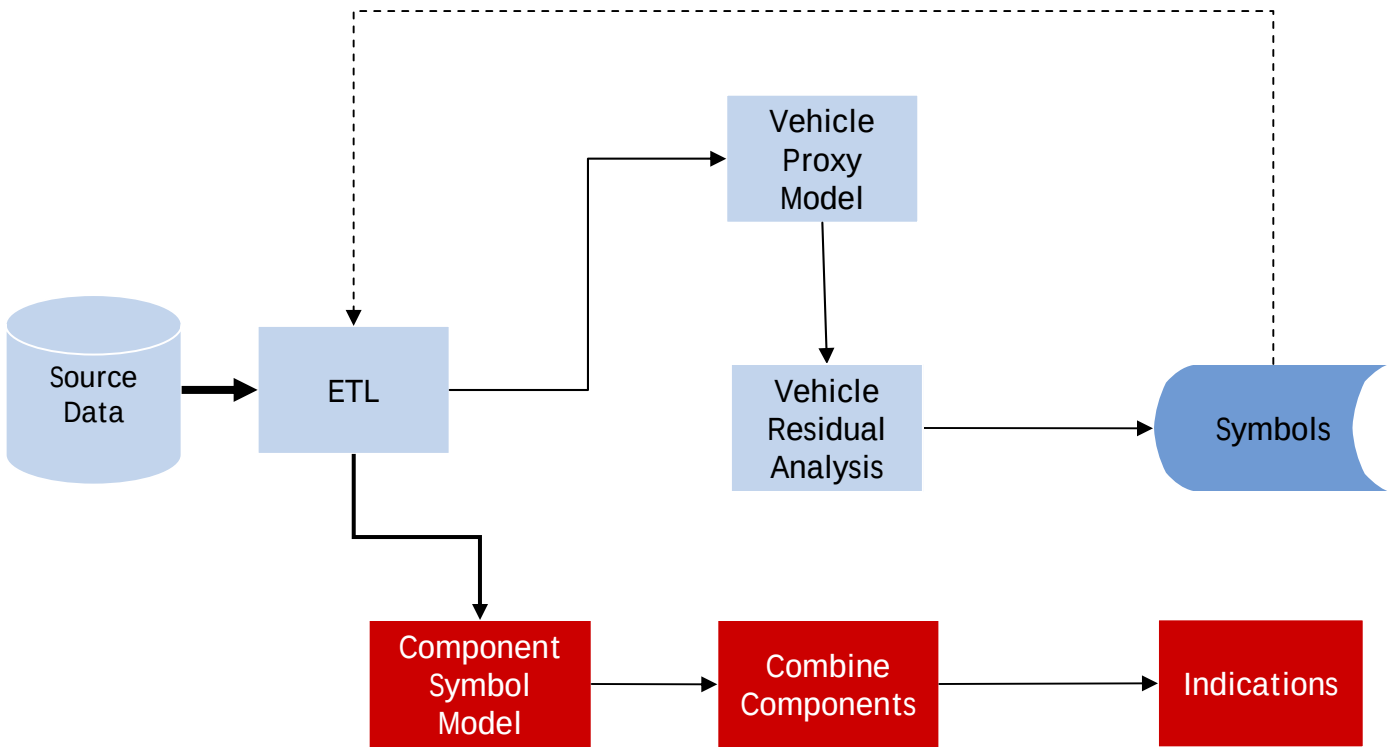


Symbol Relativities



- Model fit using new symbol definitions instead of vehicle proxies
- Test relativities using standard tests
  - Lift statistics used to judge explanatory power
  - Consistency tests used to judge predictive power
- Refine symbols/relativities as appropriate
  - Incorporate rule-based restrictions
  - Apply actuarial knowledge
  - Investigate “neighbors” with very different relativities

Symbol Relativities



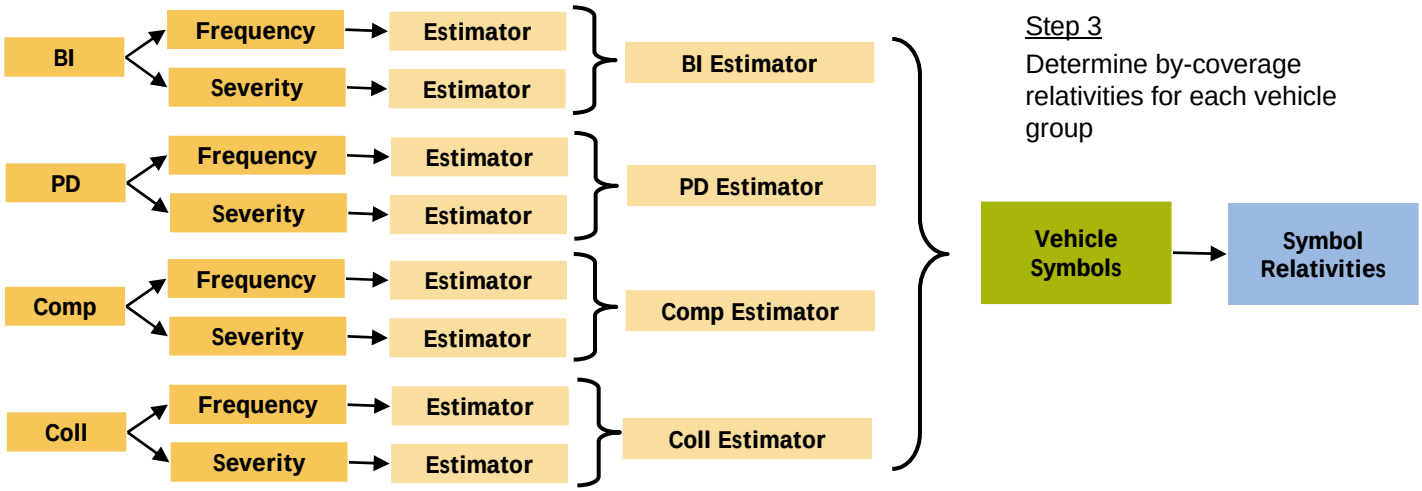
Summary

Vehicle Symboling Overview

- Accurate estimation of underlying risk associated with the vehicle

Step 1

Obtain a separate estimator by claim type and by frequency and severity for each vehicle building block. Combine estimators, as appropriate.



Step 3

Determine by-coverage relativities for each vehicle group

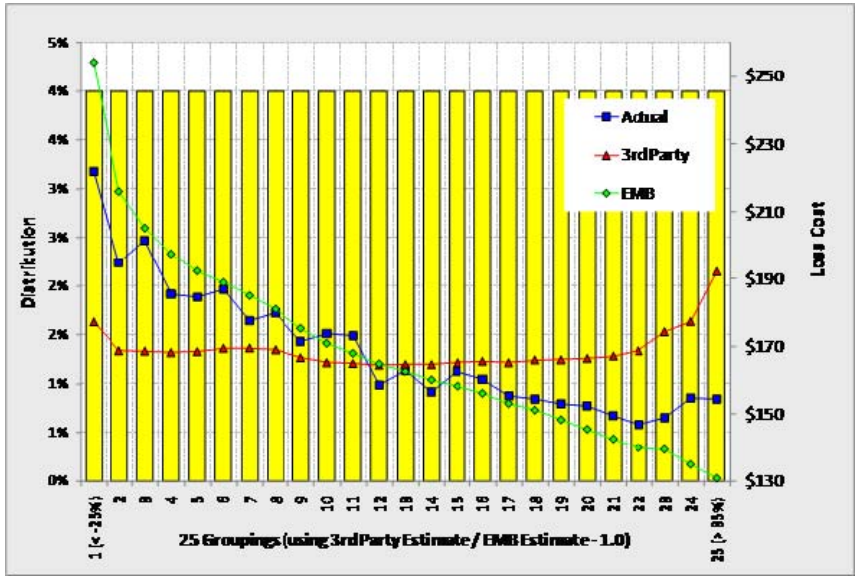
Step 2

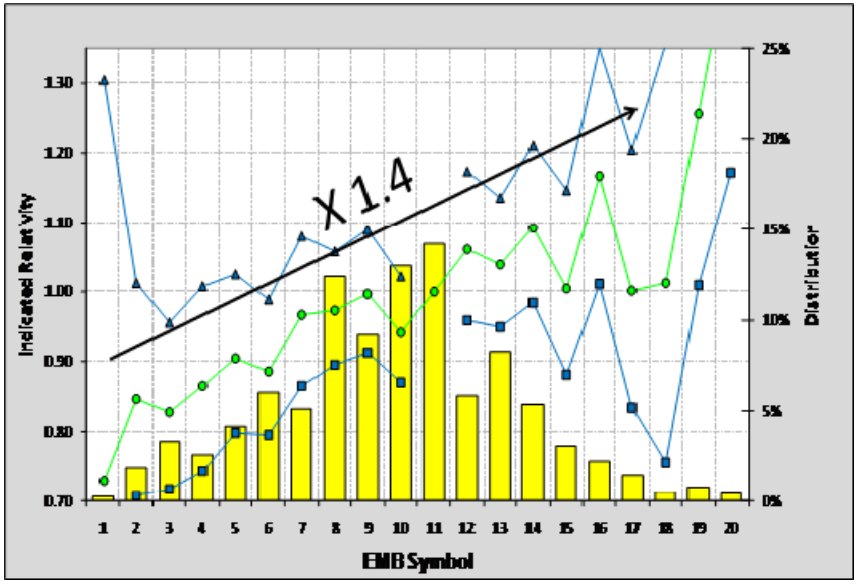
Cluster vehicle building blocks to develop symbols separately by coverage(s) and component

Summary

Proven Results

- US insurer
- Canadian insurer
- German insurance dataset





Technique has proven very successful based on proper hold-out sampling validation

Summary

- Vehicle groupings are a major driver of risk, thus it is critical that companies review symbols and relativities regularly.
- Issues exist that create special challenges with regards to symbol analysis.
  - High-dimensionality
  - Heavily correlated
- Vehicle symbols require a range of different approaches and tools (as there are different loss drivers)
- Diagnostics needed to ensure best model possible

