CAS RPM SEMINAR

PL-6: Statistically Based Territorial Modeling

TOWERS WATSON

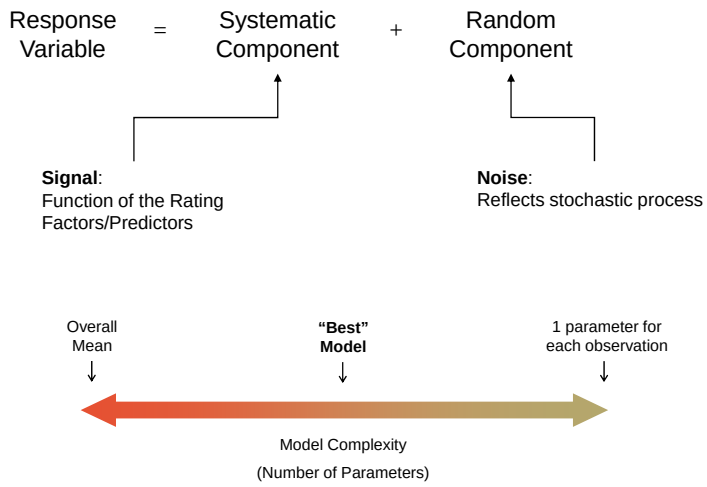
Agenda

PURPOSE To demonstrate techniques for modeling high dimensional variables within the context of a multivariate territorial analysis

OUTLINE

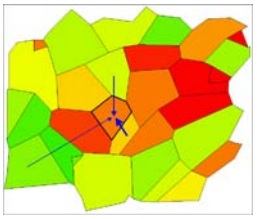
- Background
- Challenge of capturing geographical risk
- Solutions
- Overview of spatial smoothing
- Questions

Predictive Modeling



Territorial Boundary/Relativity Analysis

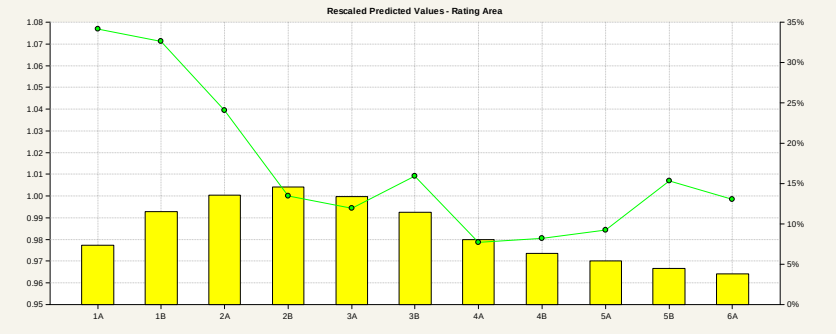
- Location is critical as a major risk driver and accounts for a substantial portion of the variation in insurance risk
- Two elements:
  - Segmentation of the risk (territorial boundaries)
  - Quantification of the risk (territorial relativities)



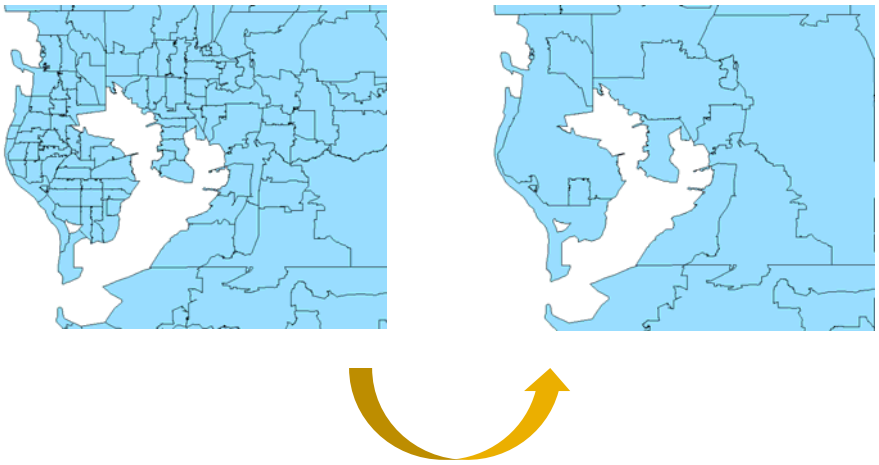
- Historically, the market focus has been on relativities
  - Initial boundaries typically based on limited data, anecdotal evidence, competitors, bureaus, and judgment
  - Regular reviews of relativities, while "tweaking" the boundaries when necessary

Relativities vs. Boundaries

- If concerned only with relativities
  - GLM can capture quantification element
  - Existing territory levels included in model as independent variable



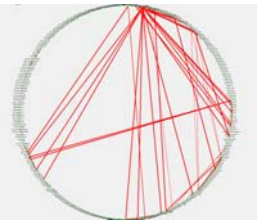
Developing Boundaries



Challenges with Boundary Analysis

- High dimensionality
  - Geography comprised of many small units
  - Individual units often have little or no experience data

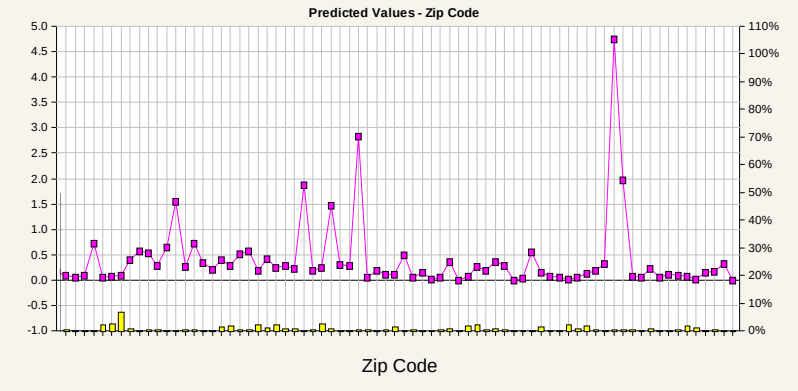
Zip Code
32003
32004
32006
32007
32008
32009
32011
32013
32024
32025
32026
32030
32033
32034



- High correlation
  - Location often highly correlated with other rating variables
    - (e.g., AOI and location, age and location)
  - Multivariate framework required to properly reflect correlations

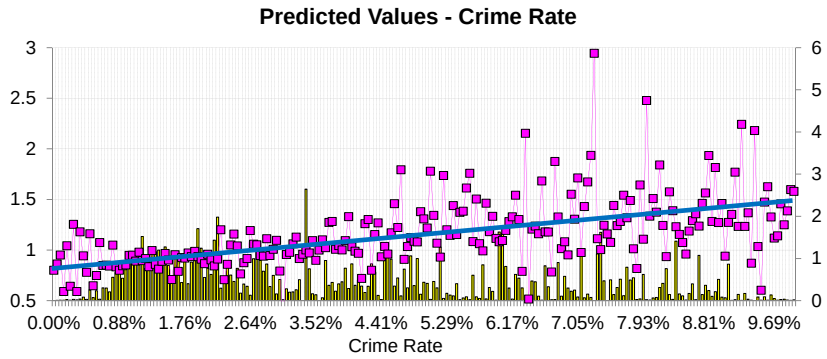
High Dimensional Categorical Variables

- Standard dimension reduction techniques fall short
  - Grouping difficult to evaluate
  - Cannot "order" geographic units, so curves not an option



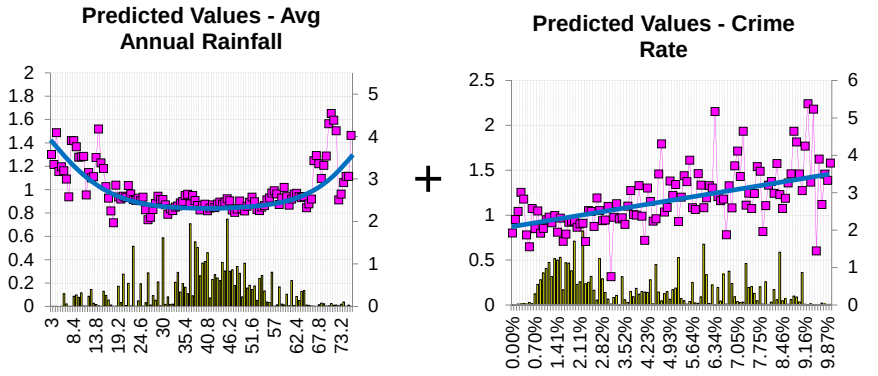
Solution 1: Use Proxies

- Proxies attach at the code level
- Geo-demographics such as:
  - Population density
  - Crime rate
- High-dimensional, but ordered; so we can fit curves



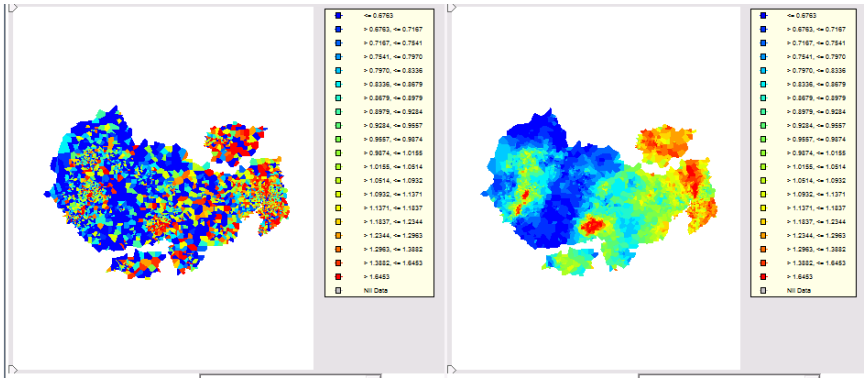
Problem with Proxies Only

- How to determine right proxies (or combinations thereof) have been used?



Solution 2: Use Proxies with Spatial Correction

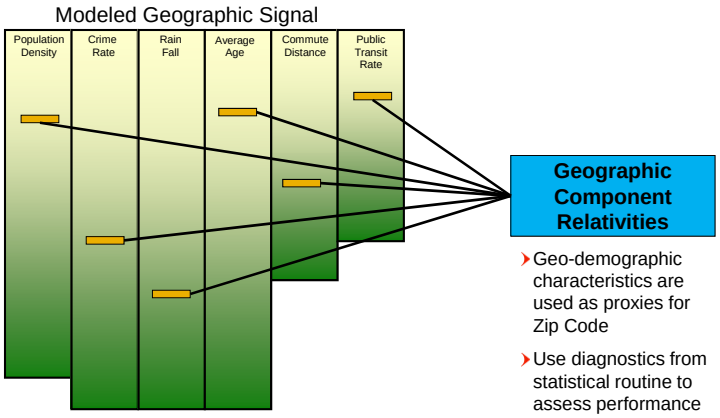
1. Include proxies in GLM
2. Then apply geo-spatial smoothing



Geographic Estimator

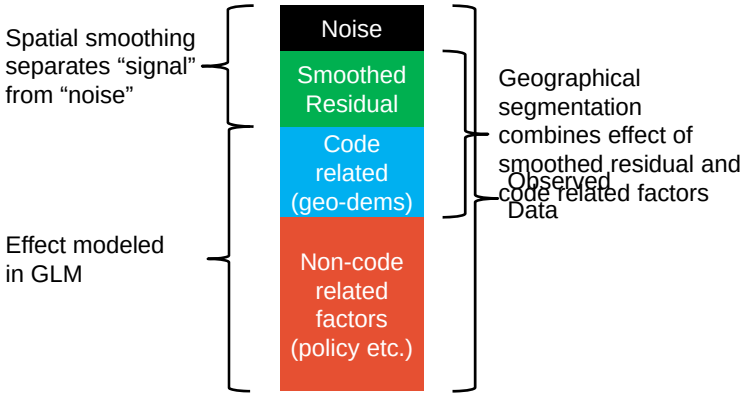
Initial Estimator:

- Component models built using geographic proxies



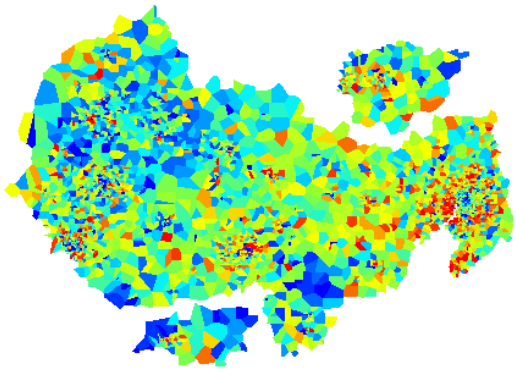
- Geo-demographic characteristics are used as proxies for Zip Code
- Use diagnostics from statistical routine to assess performance

Spatial Correction Approach



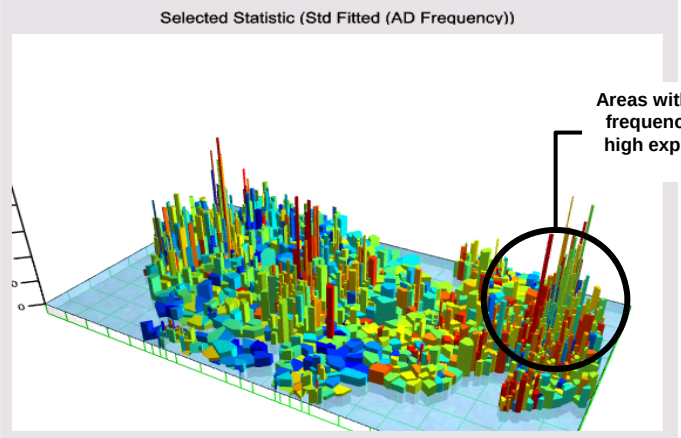
Analyzing Standardized Fitted

- Plot the combined geographic estimator from our predictive model graphically



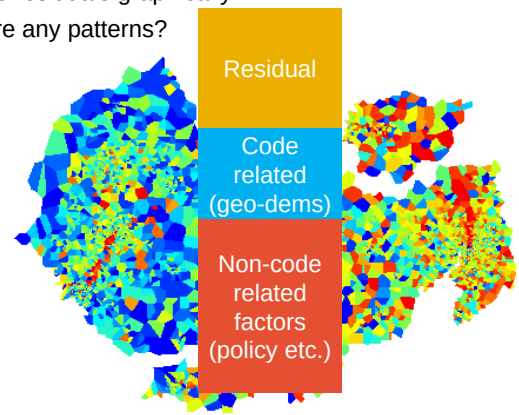
Analyzing Standardized Fitted

- 3-d view with bar heights representing exposure



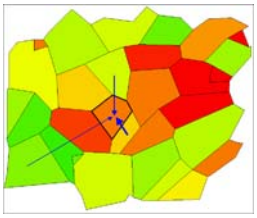
Mapping the residuals

- View the residuals graphically
- Are there any patterns?



Spatial Smoothing Methods

- Uses knowledge of surrounding areas to enhance estimates of the underlying risk in each area based on “Principle of Locality”.

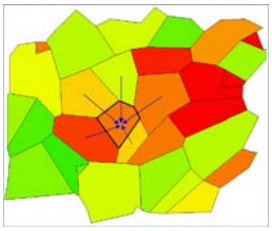


DISTANCE-BASED

- Simpler to implement and interpret.
- Does not consider natural boundaries e.g. rivers, coastline.
- May over smooth urban areas and under smooth rural areas.
- Potential Lines: Windstorm (Homeowners).

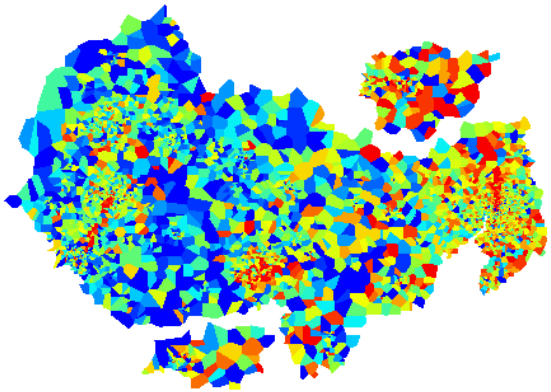
ADJACENCY-BASED

- Distribution assumptions allow prior knowledge of claims process to be incorporated.
- Distance can be built in.
- Considers natural boundaries, but not artificial boundaries.
- Potential Lines: Auto, HO Theft.



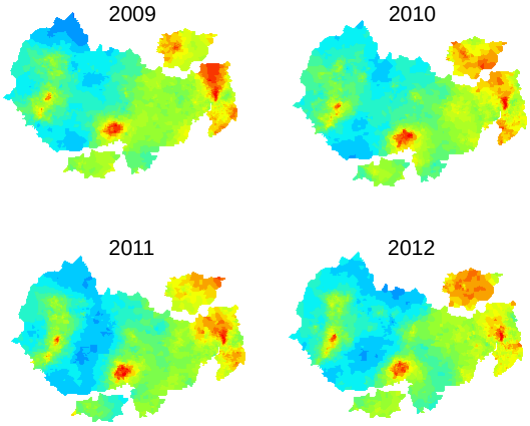
Smoothing the residuals

- View the residuals graphically
- Are there any patterns?



Consistency Over Time

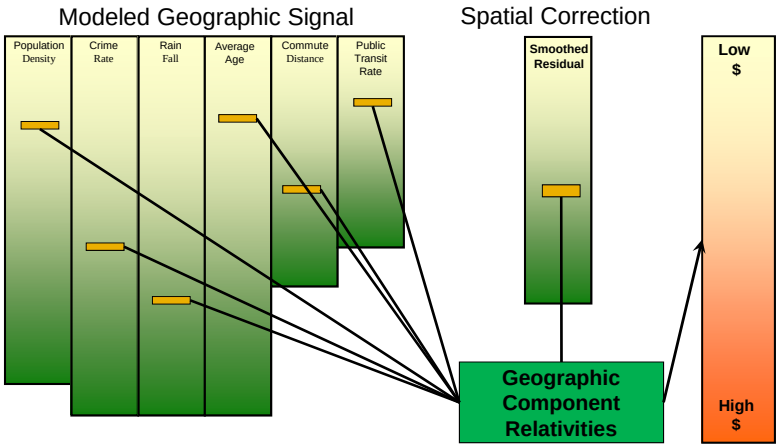
- Are the residual patterns consistent for each year?



Geographic Estimator

Geographic Spatial Correction:

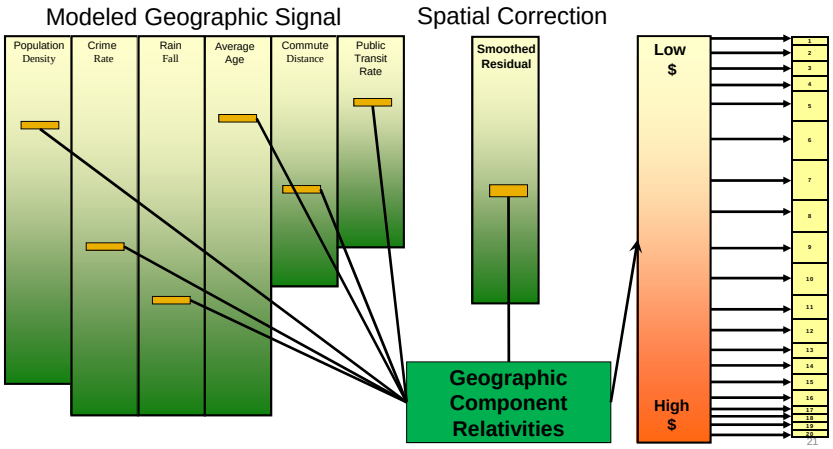
- Residual signal used to adjust scores from multivariate model



Territories

Clustering:

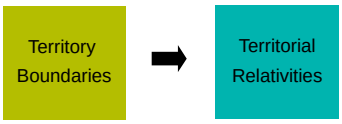
- Geographic signal is clustered into territories



Clustering

- Clustering used to produce grouping 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

Determining Territorial Relativities

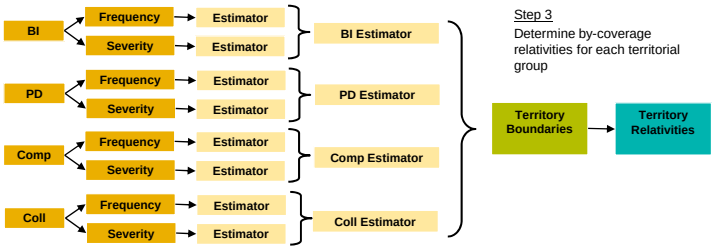


- GLM model fit using data grouped by new territorial boundaries
- Test relativities using standard GLM tests
  - Predictive in GLM model
  - Consistent over time in GLM model
- Refine boundaries/relativities as appropriate
  - Incorporate rules-based restrictions
  - Apply actuarial knowledge
  - Investigate “neighbors” with very different relativities

Territory Rating - Overview

- Accurate estimation of underlying risk associated with geography is a three stage process

Step 1
Obtain a separate estimator by claim type and by frequency and severity for each geographic building block. Combine estimators, as appropriate.



Step 3
Determine by-coverage relativities for each territorial group

Step 2
Cluster geographic building blocks to develop boundaries separately by coverage or for several coverages combined

Summary

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