Statistically Based Territory Modeling

Drew Lawyer – Sr. Professional Services Consultant
Agenda

- History of territorial modeling
  - Defining the business problem
- Modeling in practice
  - Residual analysis for creating rating territories and pricing
  - Directly modeling geo-effects
- Case Study – comparison of methods
History of Territory Modeling

- As time has progressed territorial segmentation has gotten more granular

- This causes an issue when working in the current multi-variate GLM framework
Difficulties in Territory Modeling

- Estimating loss cost for a granular location
- Creating territorial groupings for rating
- Variable have two levers (the price and the assignment)
- There is not a single agreed upon approach for defining and pricing territory
  - Low vs. High segmentation
  - Credibility weighting
  - Integrated competitor pricing
  - GLM vs. GAM
Advantages to Granular Segmentation

- Many of the largest insurers are filing rates by...
  - Zip code, Census tract, or Census block

- Avoid large rate differences between adjacent territories

- Avoid analytical issues with defining classic “territorial boundaries”

- More refined estimate of risk is a competitive advantage
  - Write and retain good risks
  - Send bad risks to the competition
Industry Survey
Rate Change Drivers: Strategic Goals

One way to achieve greater profitability is through more refined segmentation.

- Increased retention: 4%
- Consistent product design: 4%
- Customer acquisition: 8%
- Greater profitability: 84%

Survey responses were collected online from 99 insurance professionals representing companies that sell Homeowners coverage in the United States and Canada.
Agenda

- History of territorial modeling
  - Defining the business problem

- Modeling in practice
  - Residual analysis for creating rating territories and pricing
  - Directly modeling geo-effects

- Case Study – comparison of methods
Approaches to Territory Modeling

Residual Analysis:
1. Build base GLM loss cost model
2. Use Residuals to define rating territories
3. Refit GLM with new territory definitions

Loss Cost Data

Direct Estimate:
1. Build GLM loss cost model and directly incorporate geo effect

Final Risk Models

≈
Approaches to Territory Modeling

- Residual Analysis
  - High level of control over estimates
  - Clean fit into a multiplicative rating structure
  - Time consuming multi-step process

- Direct Estimate
  - Simplified modeling process
  - Intuitive interpretation of results
  - Lack of control over estimates
  - Additional work is required to create a multiplicative structure
Data for Analysis

- Homeowners loss data in Illinois provided by large insurer
- Exposure years 2007-2011 & 2013
- 826,000 exposure years
  - 2007-2011 used for model development
  - 2013 used for comparison of results
- Risk models developed non-weather peril
  - Fire, Theft, Water, and Other
- Tweedie GLM used to model pure premium

Methodology is applicable for other business lines
Agenda

- History of territorial modeling
  - Defining the business problem
- Modeling in practice
  - Residual analysis for creating rating territories and pricing
  - Directly modeling geo-effects
- Case Study – comparison of methods
Approaches to Territory Modeling

Residual Analysis

1. Build base GLM loss cost model
2. Use Residuals to define rating territories
3. Refit GLM with new territory definitions
4. Final Risk Models

Loss Cost Data

Direct Estimate

Build GLM loss cost model and directly incorporate geo effect

Final Risk Models
Residual Territory Modeling

- Develop initial countrywide loss cost models by peril
- Models include principal components (PCA) of geo-demographic data not used in rating
- Starting point for all state specific models
Residual Territory Modeling

- The residuals for a specific state are tabulated by census tract.
- Unsmoothed, the residual output appears as noise.
- It is possible that not all tracts have exposures.
- A smoothing function is applied to the residual.
Residual Territory Modeling

- The smoothing algorithm removes noise and draws out the signal.
- The resulting estimates by census tract are then placed into 100 noncontiguous groups.*

* modeler/company preference dictates smoothing method, number of groups, and other inputs into the smoothing.
Residual Territory Modeling

- The ordered groups are now returned to the risk model
- The other betas are fixed (offset) and the PCA’s are removed
- The territorial effect is then fit with some type of variate

- The final result is 100 price points by census tract
Residual Territory Modeling

Theft Territory

Theft Territory Model Fit

Theft Factor
Residual Territory Modeling

Theft Territory

Theft Factor
Agenda

- History of territorial modeling
  - Defining the business problem
- Modeling in practice
  - Residual analysis for creating rating territories and pricing
    - Directly modeling geo-effects
- Case Study – comparison of methods
Approaches to Territory Modeling

**Residual Analysis**

- Build base GLM loss cost model
- Use Residuals to define rating territories
- Refit GLM with new territory definitions
- Final Risk Models

**Loss Cost Data**

**Direct Estimate**

- Build GLM loss cost model and directly incorporate geo effect
- Final Risk Models
Direct Estimate of Geo Effect

- After developing the initial countrywide loss cost models
- Again, remove the PCA’s and fix (offset) other rating factors
- Add the geo parameter to account for the territorial effect
  - Geo parameter is built using latitude and longitude
  - Can either be defined using customer geo-coding (specific location for each customer) or mapping lat/long to the geo root level (e.g. census tract)
Direct Estimate of Geo Effect

- The smoothing algorithm is applied to the geo parameter to draw out the signal
- Can be done in different software; methods vary slightly
- Earnix uses thin-plate splines for smoothing

Generate knots by random sampling will add knots randomly proportional to observation density

Cross-validation ensures that the geo effect does not overfit the data
Direct Estimate of Geo Effect

50 Knots

100 Knots
Direct Estimate of Geo Effect

- Determining the proper number of knots is an iterative process.
- Cross-validation reduces the chances of overfitting the geo effect; however, it is still possible.
- Each census tract is defined as its own territory. If desired, neighboring tracts can be grouped together.
  - Useful if extreme values are identified
- Due to the nature of thin-plate splines the GLM loss cost model is actually transformed into a GAM
- The functional form can easily be converted back to multiplicative where a rating factor is assigned to each census tract
Agenda

- History of territorial modeling
  - Defining the business problem
- Modeling in practice
  - Residual analysis for creating rating territories and pricing
  - Directly modeling geo-effects
- Case Study – comparison of methods
Comparison of Results – Theft Peril

Residual Territory Modeling

Direct Estimate

≈
Comparison of Results – Theft Peril

Residual Territory Modeling

Difference in Theft Premium

Direct Estimate
Comparison of Results – Theft Peril

Range Analysis | Residual Modeling | Direct Estimation
---|---|---
1st Percentile | $2 | $2
99th Percentile | $96 | $107
Range (inner 98%) | $94 | $105

50% exposures within +-$2

DISTRIBUTIONAL DIFFERENCE RESIDUAL - DIRECT
Comparison of Results – Water Peril

Residual Territory Modeling

Direct Estimate
Comparison of Results – Water Peril

Residual Territory Modeling

Difference in Water Premium

Direct Estimate
Comparison of Results – Water Peril

<table>
<thead>
<tr>
<th>Range Analysis</th>
<th>Residual Modeling</th>
<th>Direct Estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Percentile</td>
<td>$6</td>
<td>$5</td>
</tr>
<tr>
<td>99th Percentile</td>
<td>$158</td>
<td>$168</td>
</tr>
<tr>
<td>Range (inner 98%)</td>
<td>$152</td>
<td>$163</td>
</tr>
</tbody>
</table>

50% exposures within +-$5

DISTRIBUTIONAL DIFFERENCE
RESIDUAL - DIRECT

Earnix Copyright 2015
Comparison of Results – Combined Peril

<table>
<thead>
<tr>
<th>Statistics (from holdout)</th>
<th>Residual Modeling</th>
<th>Direct Estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Percentile</td>
<td>$42</td>
<td>$38</td>
</tr>
<tr>
<td>99th Percentile</td>
<td>$808</td>
<td>$872</td>
</tr>
<tr>
<td>Range (inner 98%)</td>
<td>$766</td>
<td>$834 (9%)</td>
</tr>
</tbody>
</table>

Additional segmentation is useless if segments do not result in better risk classification.
Comparison of Results

- Lift charts used to compare the results

- **Out-of-time** (2013) premiums were compared ~ Direct / Residual

- The ordered values are bucketed into 5 equal exposure quintiles

- The loss ratio was then observed by comparing the observed losses to the current average premium within the group – Residual Premium

- Bars to the left depict where Direct Estimation approach predicts lower than Residual Estimation

- Bars to the right predicts higher than residual

- If direct estimation method provides lift, loss ratios should trend upward

- Lift is calculated as (Highest Quintile LR / Lowest Quintile LR – 1)
Lift Chart Analysis

Fire Peril

Fire Lift = \( \frac{127\%}{80\%} - 1 = 80\% \)

Other Peril

Other Lift = \( \frac{100\%}{87\%} - 1 = 15\% \)

Theft Peril

Theft Lift = \( \frac{105\%}{84\%} - 1 = 25\% \)

Water Peril

Water Lift = \( \frac{127\%}{78\%} - 1 = 62\% \)
Total – Combined Peril
Fire, Other, Theft, Water

Loss Ratio Relativity Quintiles
Direct Estimation / Residual Estimation

Lift = \( \frac{132\%}{81\%} - 1 \) = 63%

Positive, but not monotonic
Rate Comparison

Direct Approach
>
>10% Lower
LR = 99%

Residual Approach
>
>10% Lower
LR = 1.20%
Future Analysis

- Out of time dataset limitations
  - Limited number of observations for homeowners modeling
  - Recent year has limited development (should be minimally bias with territory)

- Test factors without initial beta offset
  - Larger dataset required
  - Estimating geo and other factors simultaneously eliminates the need for PCA, thus simplifying the process more
## Conclusion:

- Both modeling techniques perform similarly on out-of-time sample

<table>
<thead>
<tr>
<th>Residual Modeling</th>
<th>Direct Estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long / complex process</td>
<td>Quick / simple process</td>
</tr>
<tr>
<td>2 weeks for analysis</td>
<td>2 days for analysis*</td>
</tr>
<tr>
<td>Less Segmentation</td>
<td>More Segmentation</td>
</tr>
<tr>
<td>Full control of process</td>
<td>Put faith into statistics</td>
</tr>
<tr>
<td>Results in a discrete territory groups</td>
<td>Results in an individual rate for each geo root level</td>
</tr>
<tr>
<td>GLM</td>
<td>GAM</td>
</tr>
</tbody>
</table>

*once initial process is defined
Thank You

Drew Lawyer  
Professional Services Consultant, Earnix  
+1-309-530-2360  
drew.lawyer@earnix.com

For complete Homeowners Insurance Ratemaking Applications Survey results, visit earnix.com
Additional Research

Background – Territorial Ratemaking

- Common techniques for reflecting geography in insurance models:
  - Odontic models
  - Adding geographic, crime, weather, traffic... variables to models
  - Spatial smoothing concepts

- Generalized Additive Models are a practical way to incorporate spatial smoothing in one's model.

- Some advantages:
  - Fitting paradigm: GAM is a generalization of GLM
  - Latitude and longitude can be used as model inputs
  - All/long can be incorporated alongside demographic variables
  - Use of offsets enables "modular" approach

Standard references:
- Generalized Additive Models by Hastie and Tibshirani (just last the author's name)
- Generalized Additive Models by Simon Wood (paraphraged here)

Deloitte.

Geo-spatial Analysis with Generalized Additive Models

CAS Annual Meeting
Chicago
November, 2011

Jim Guszcza
Deloitte Consulting LLP
The University of Wisconsin-Madison

PL-7
Putting Your Company on the Map:
Determination of Statistically Indicated Territory Boundaries
2006 CAS Seminar on Ratemaking
Duncan Anderson MA FIA
Watson Wyatt Worldwide

Two approaches to spatial smoothing

- Estimate effect of non-territory factors and then smooth residuals to derive new zones
- very practical
- can include differing distance metrics
- can incorporate credibility in a straightforward way
- distorted by non-systematic element of experience
- slight distortion from correlated factors

- Fit surface directly using maximum likelihood as part of GLM (ideally with splines)
- MLE
- harder to fit
- prone to over-smooth