Statistically Based Territory Modeling

Drew Lawyer – Sr. Professional Services Consultant

EARNIX
Predict. Perform.

Agenda

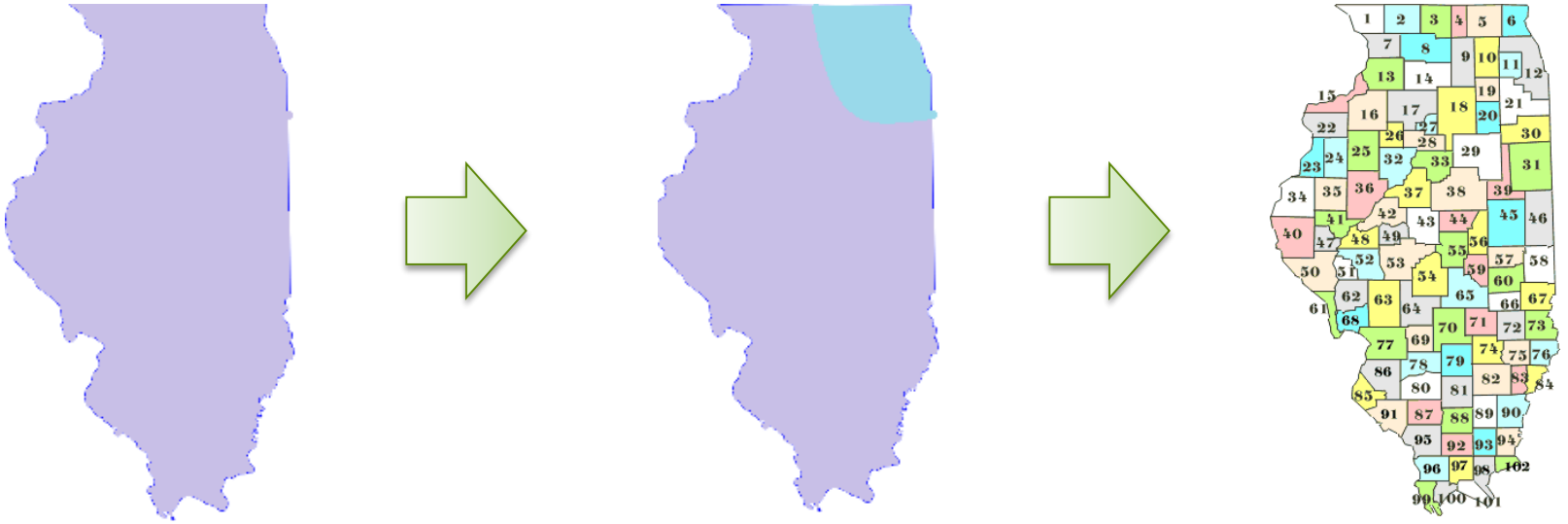
EARNIX

- 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

EARNIX

- 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

EARNIX

- 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

EARNIX

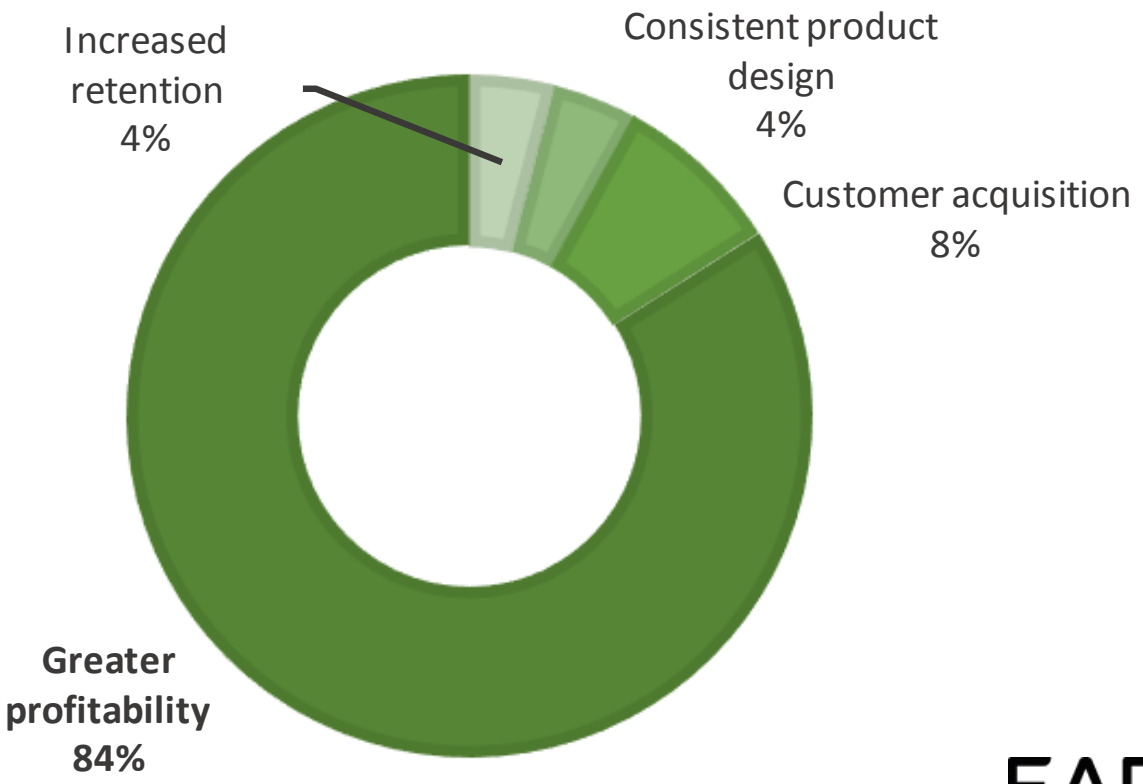
- Many of the largest insurers are filing rates by…
  - Zip code, Census tract, or Census block

- 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**



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

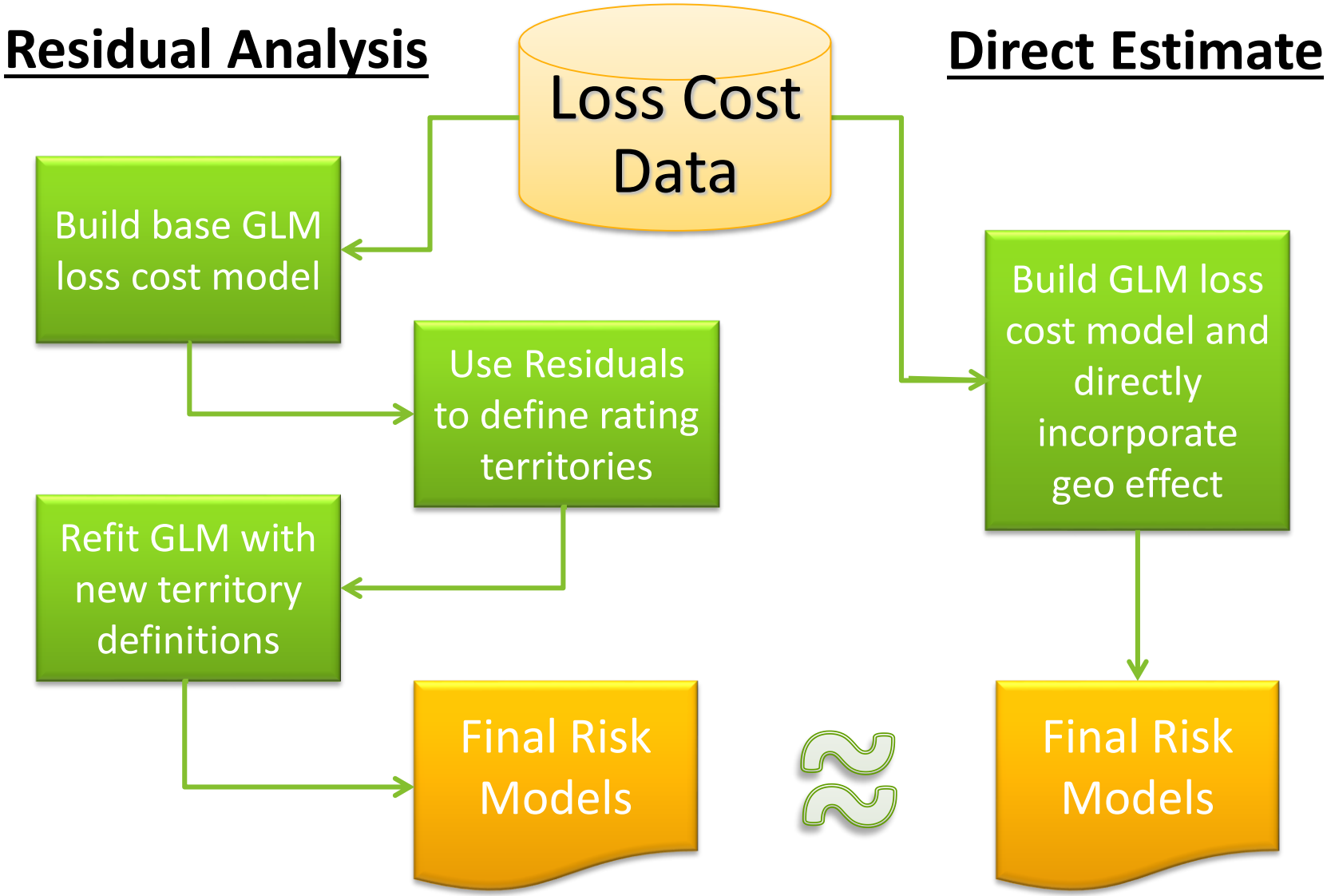
Agenda

EARNIX

- 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

EARNIX



Approaches to Territory Modeling

EARNIX

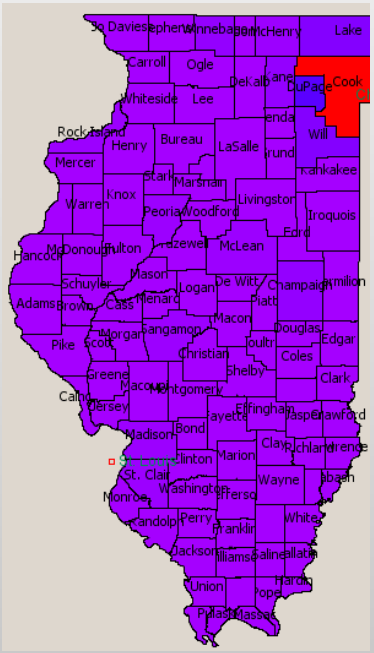
- **Residual Analysis**
  - High level of control over estimates
  - Clean fit into a multiplicative rating structure
  - Time consuming multi-step process
- **Direct Estimate**
  - Simplified modeling process
  - Intuitive interpretation of results
  - Lack of control over estimates
  - Additional work is required to create a multiplicative structure

Data for Analysis

EARNIX

- 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

Pop Density



Methodology is applicable for other business lines

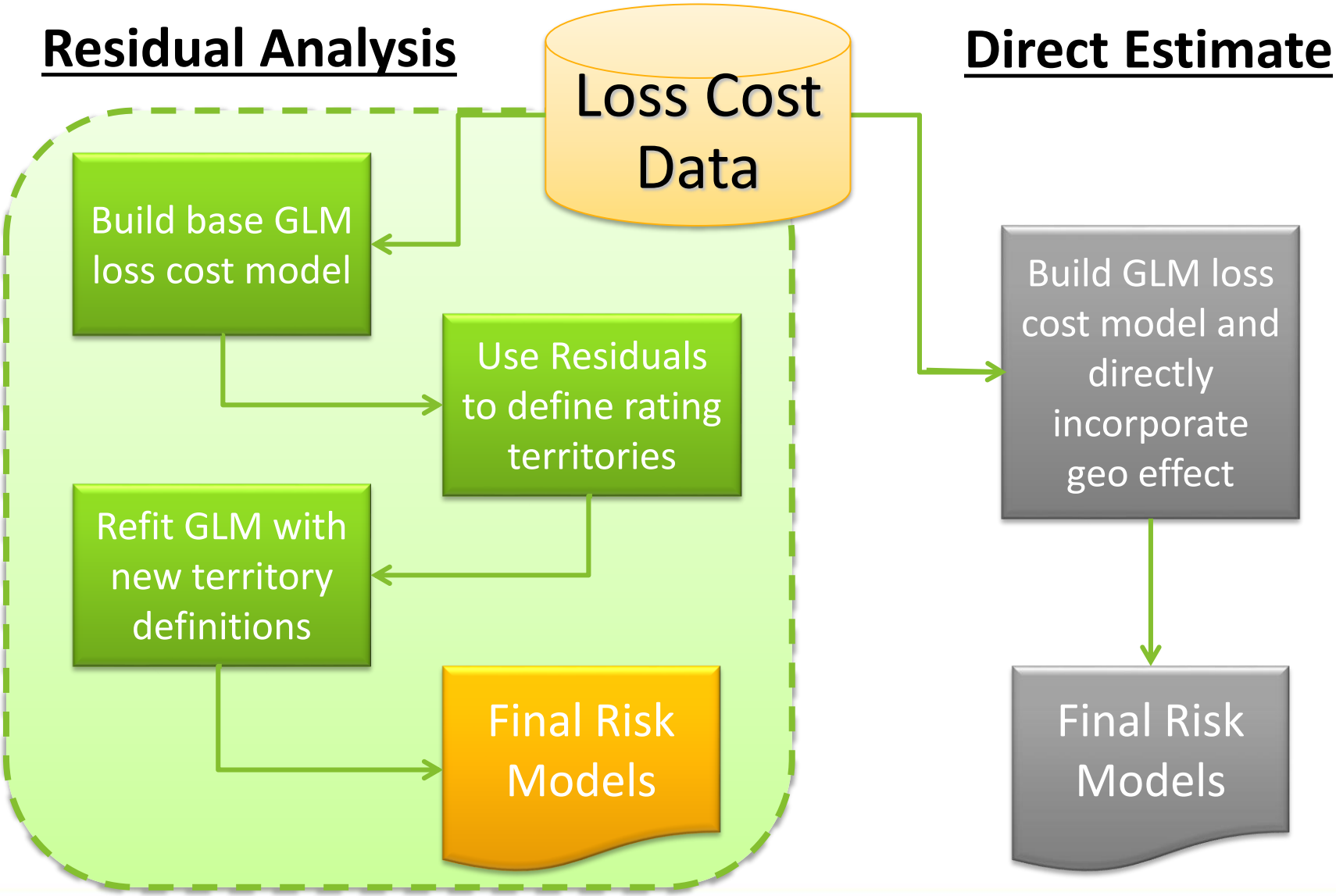
Agenda

EARNIX

- 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

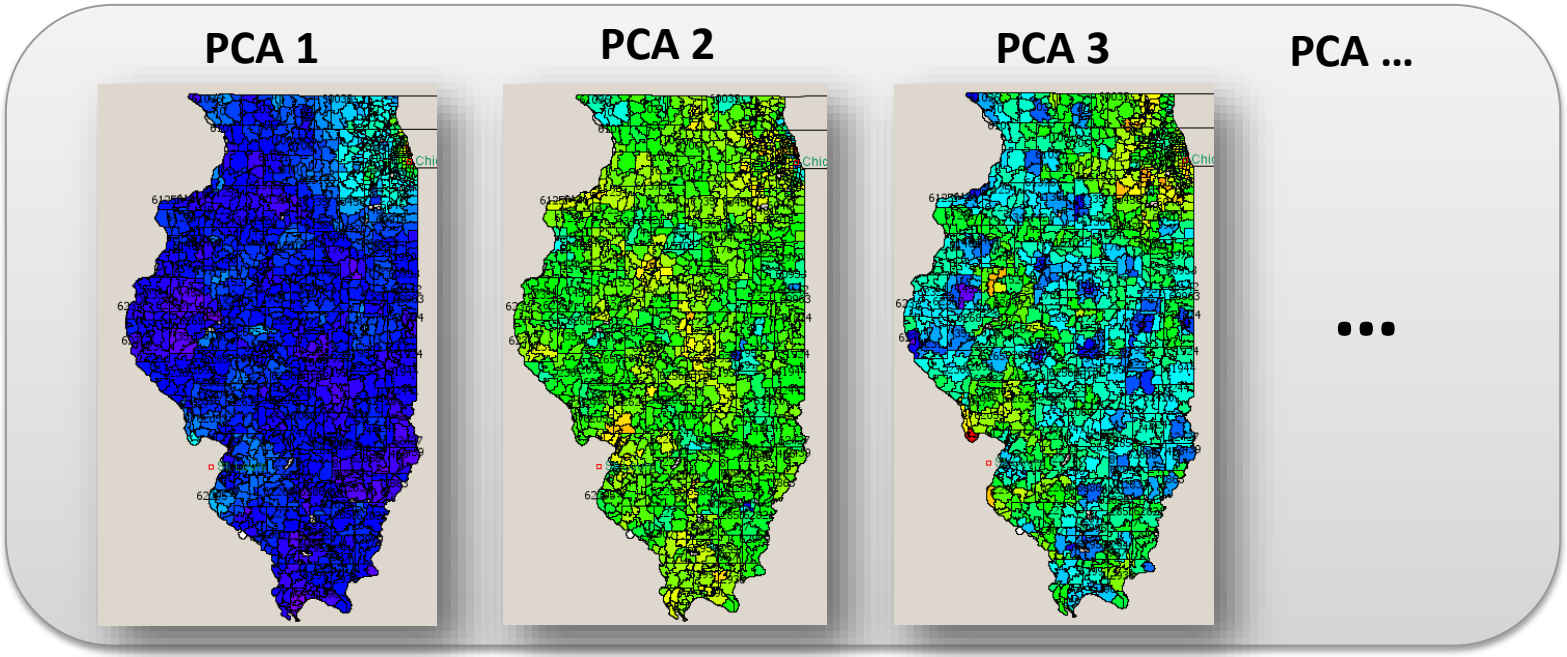
EARNIX



Residual Territory Modeling

EARNIX

- Develop initial countrywide loss cost models by peril
- Models include principal components (PCA) of geo-demographic data not used in rating

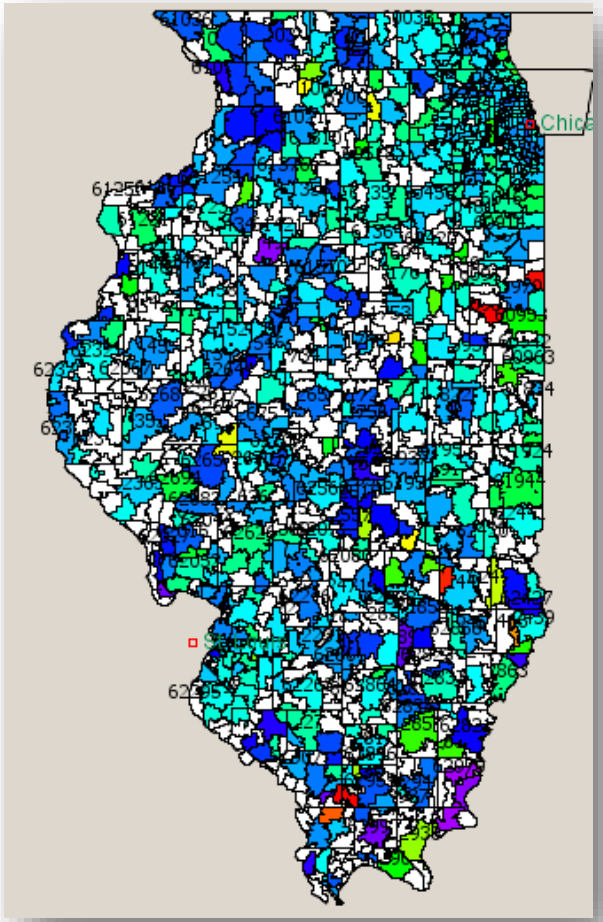


- Starting point for all state specific models

Residual Territory Modeling

EARNIX

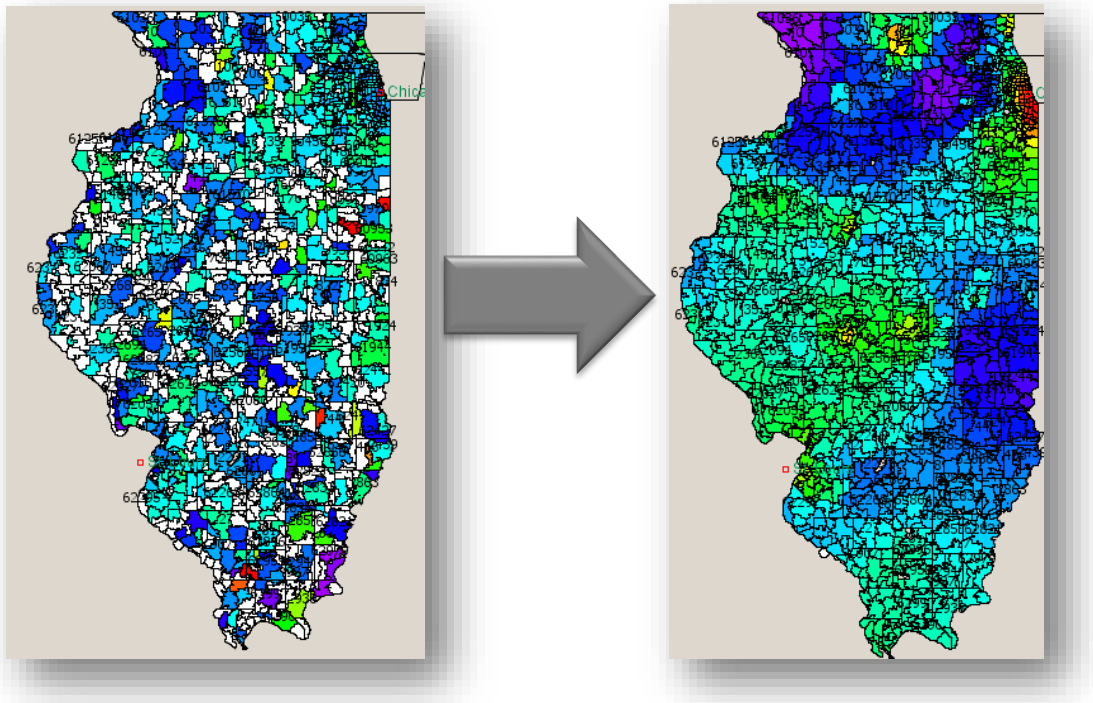
- The residuals for a specific state are tabulated by census tract
- Unsmoothed, the residual output appears as noise.
- It is possible that not all tracts have exposures
- A smoothing function is applied to the residual



Residual Territory Modeling

EARNIX

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

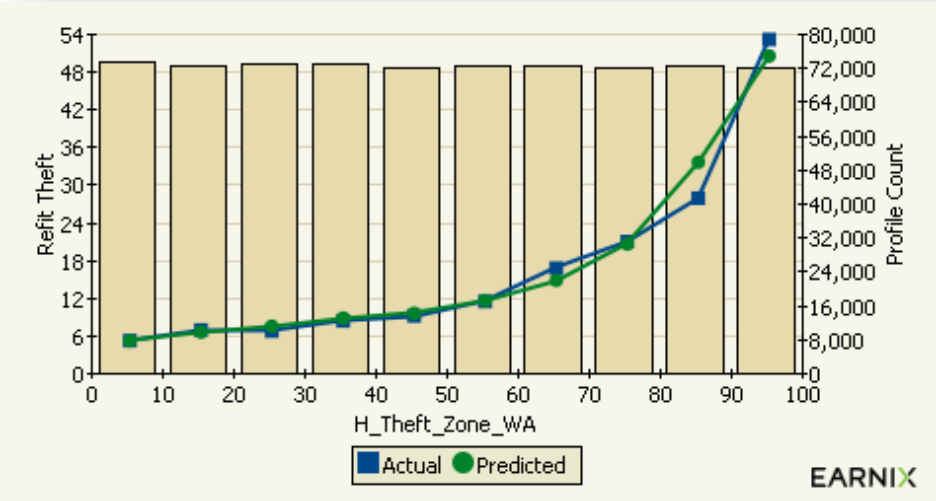


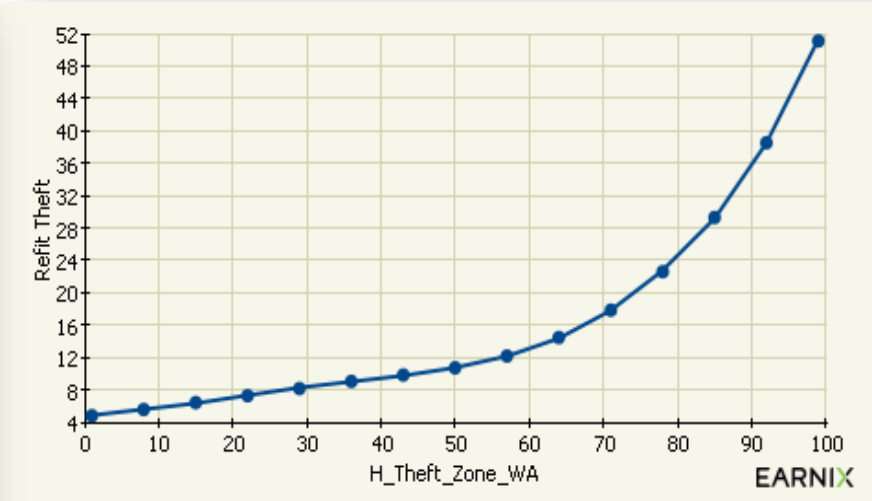
* modeler/company preference dictates smoothing method, number of groups, and other inputs into the smoothing

Residual Territory Modeling

EARNIX

- 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

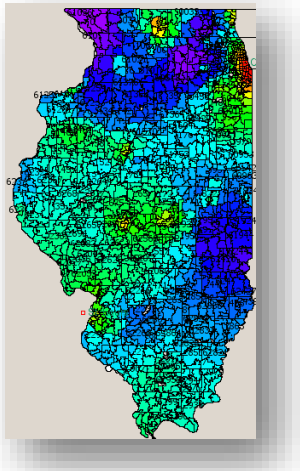
Residual Territory Modeling

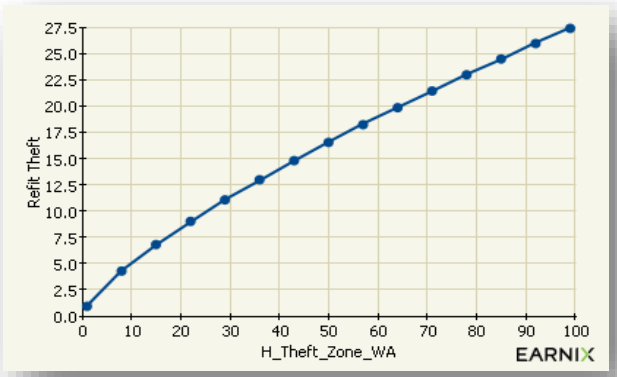
EARNIX

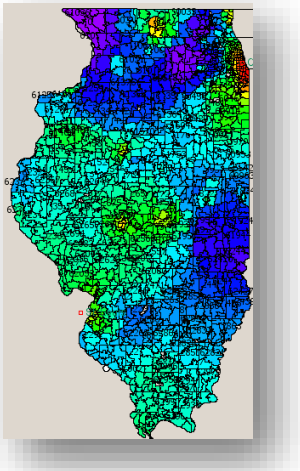
Theft Territory

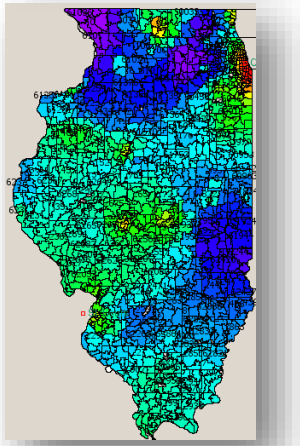
Theft Territory Model Fit

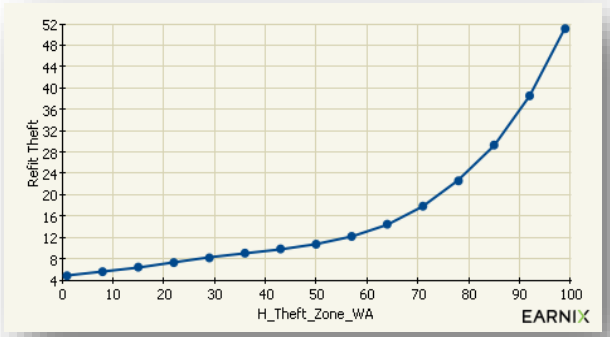
Theft Factor

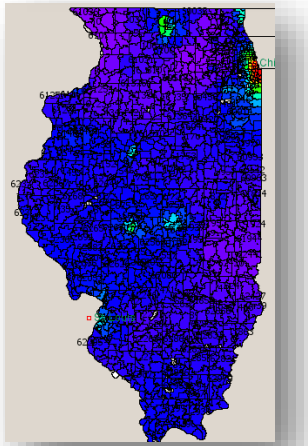








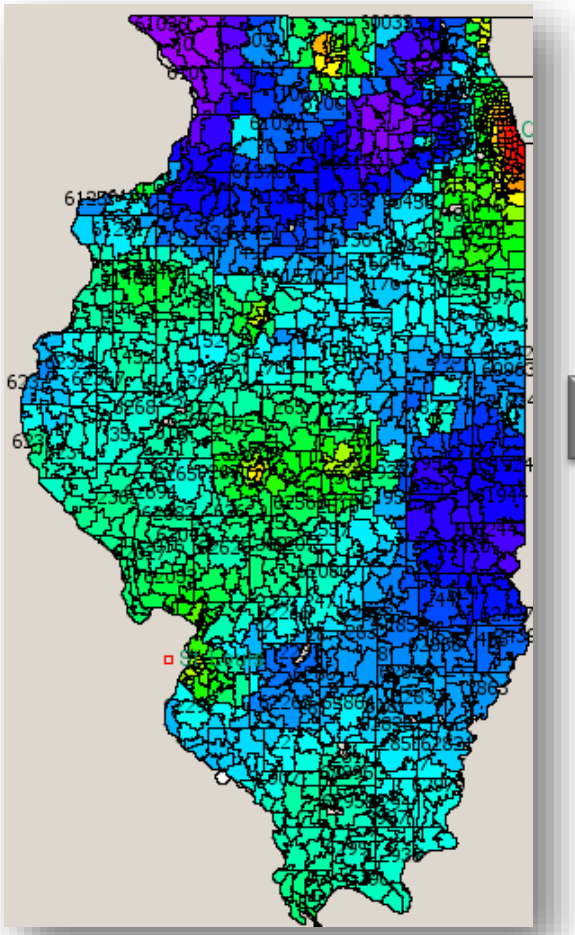




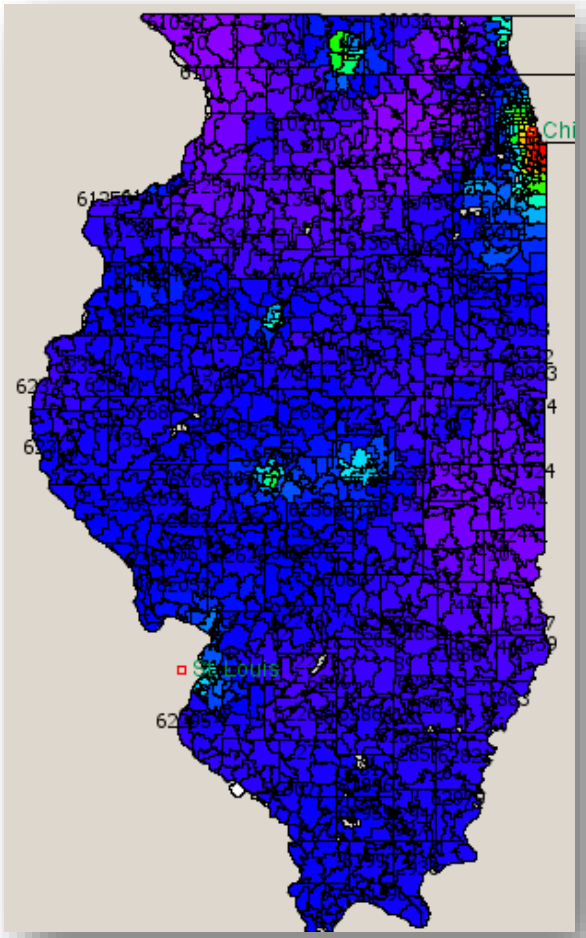
Residual Territory Modeling

EARNIX

Theft Territory



Theft Factor



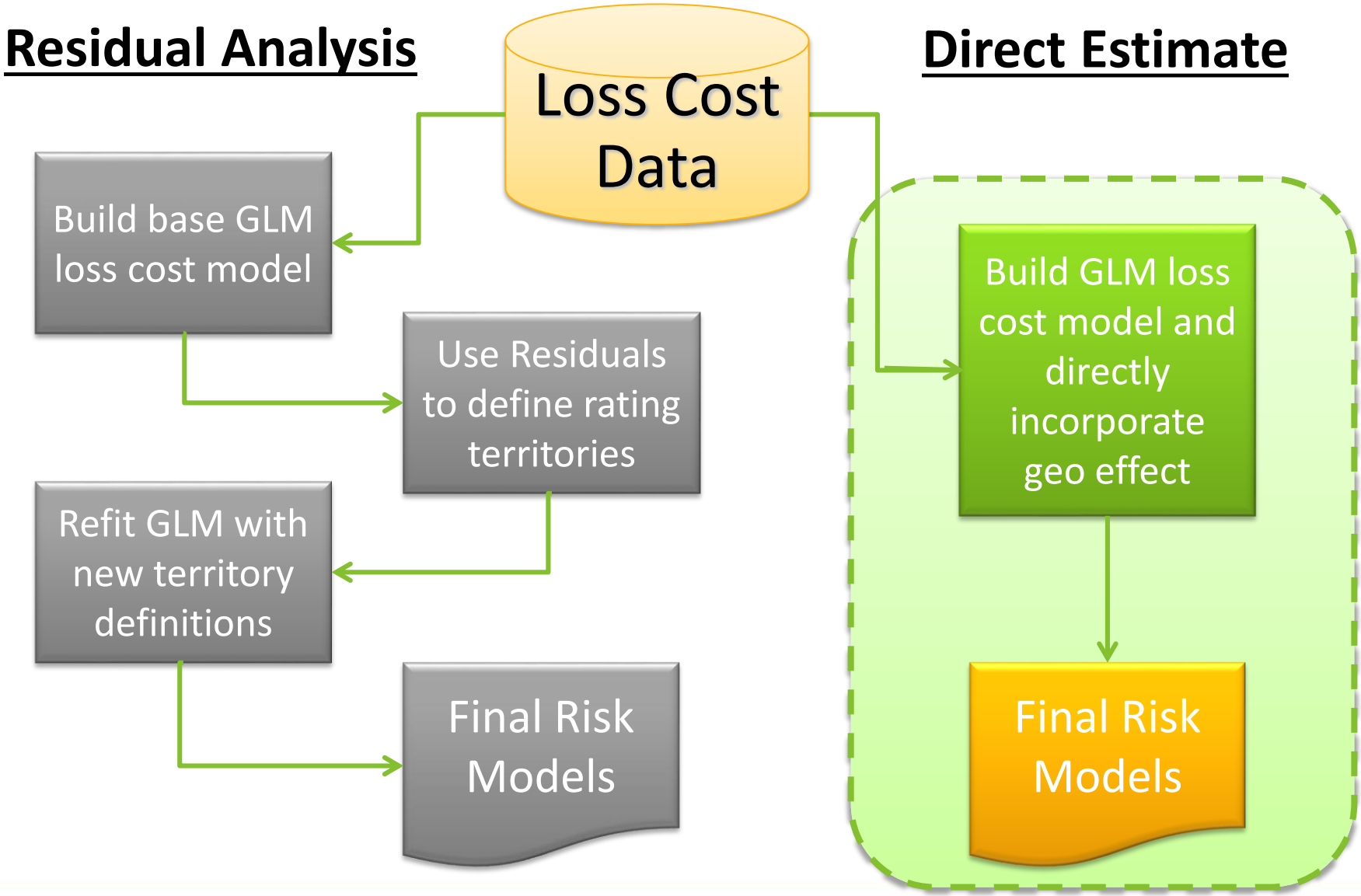
Agenda

EARNIX

- 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

EARNIX



Direct Estimate of Geo Effect

EARNIX

- 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)

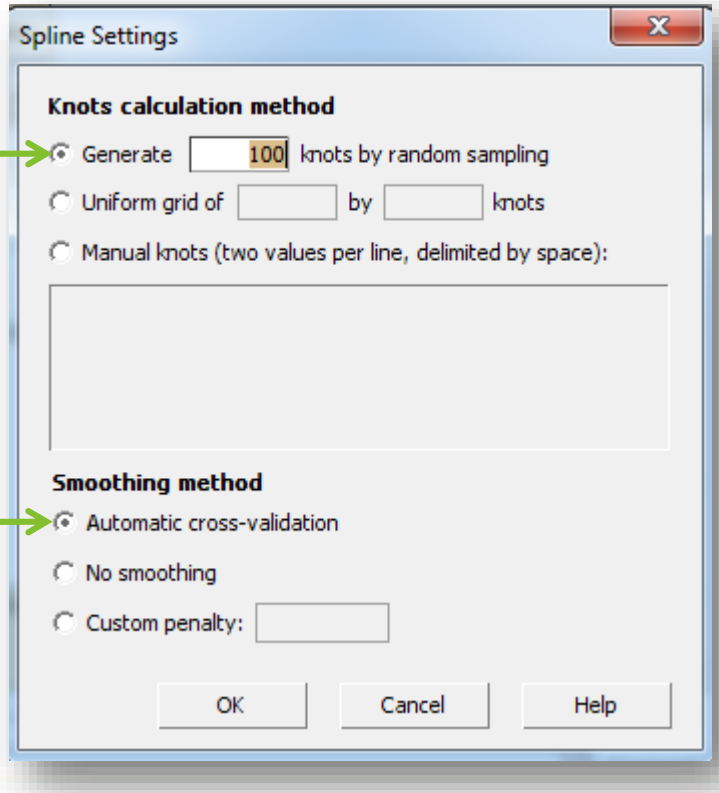
Direct Estimate of Geo Effect

EARNIX

- 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

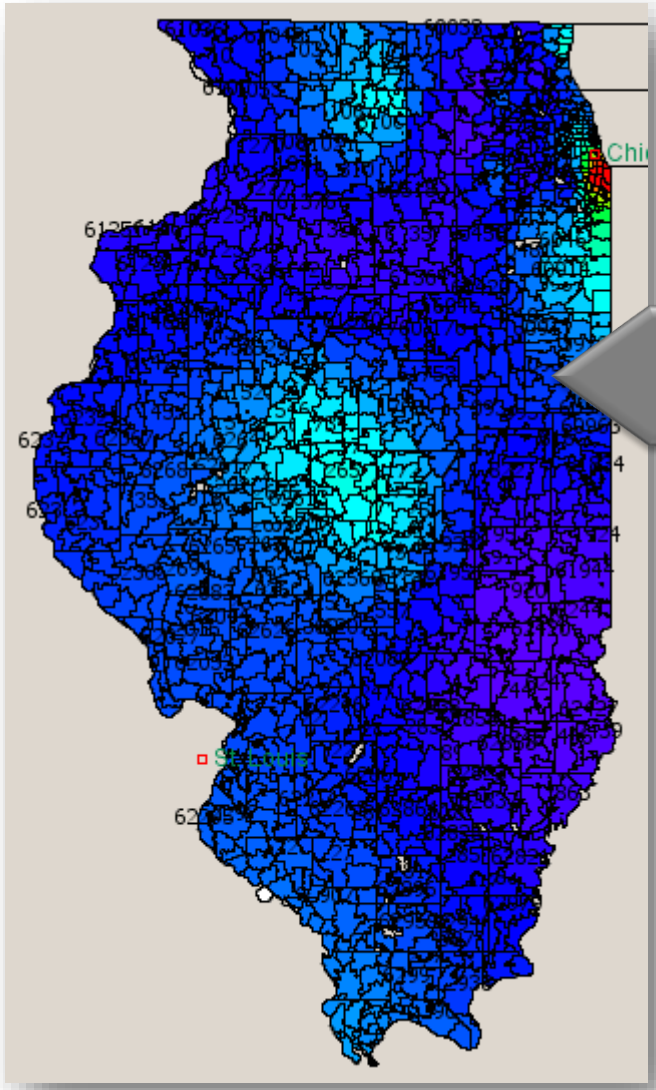
Generate knots by random sampling will add knots randomly proportional to observation density

Cross-validation ensures that the geo effect does not overfit the data



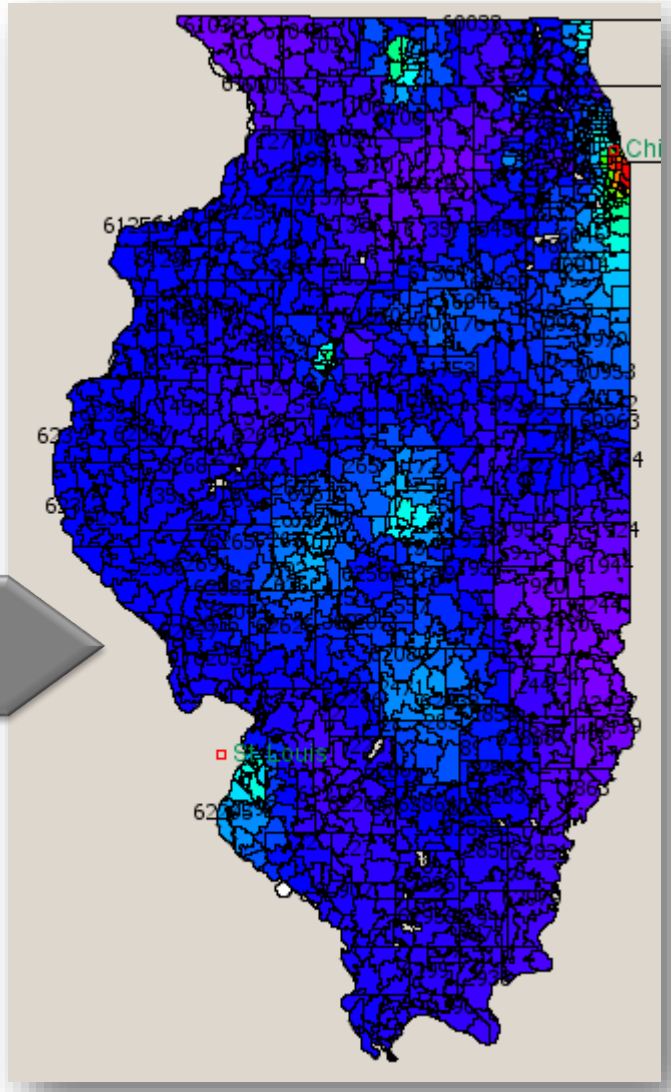
Direct Estimate of Geo Effect

EARNIX



50 Knots

100 Knots



Direct Estimate of Geo Effect

EARNIX

- 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

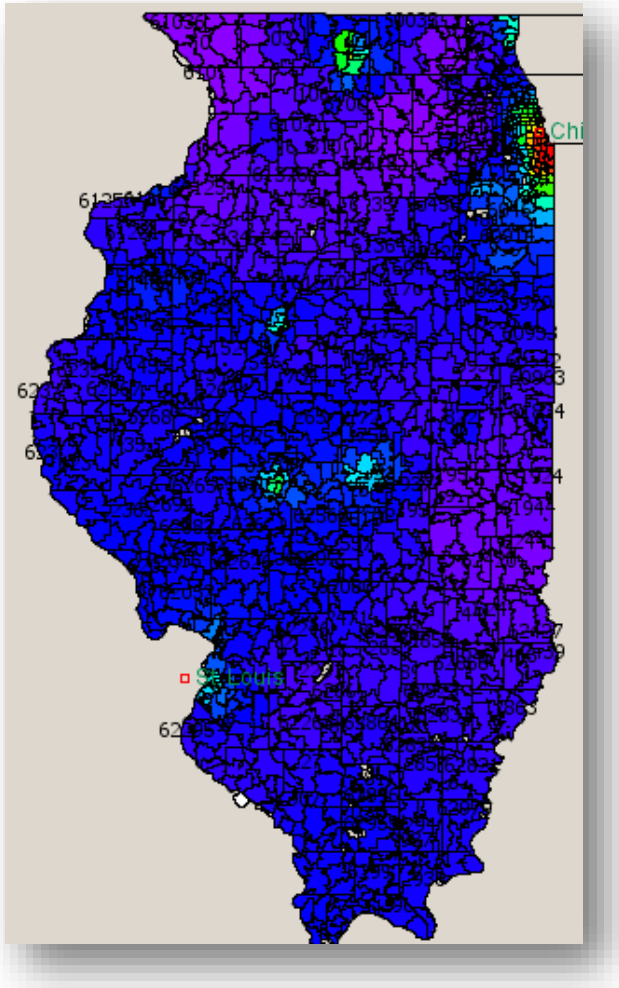
EARNIX

- 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

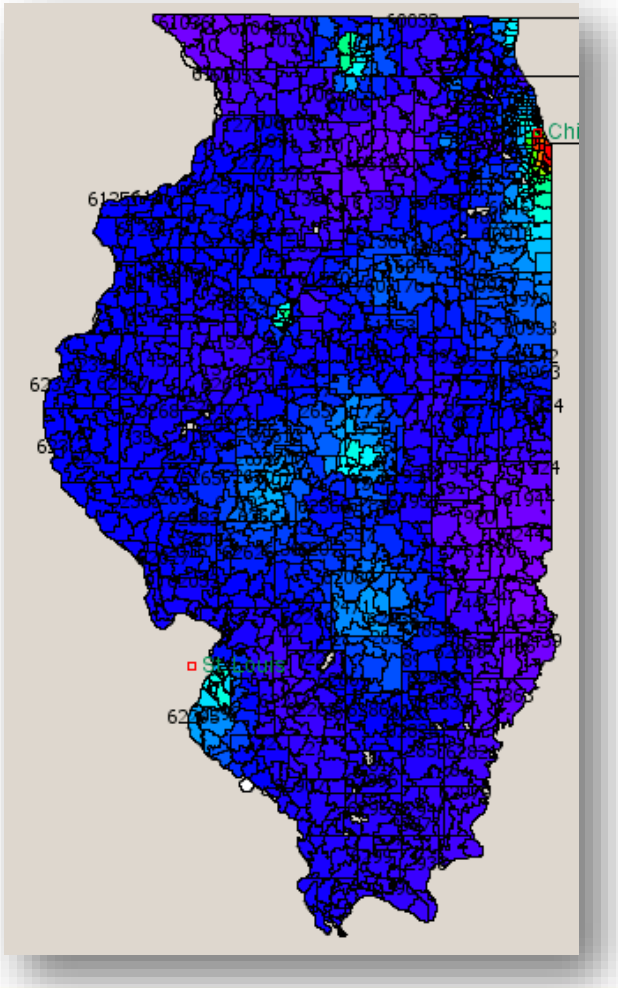
EARNIX

Residual Territory Modeling



≈

Direct Estimate



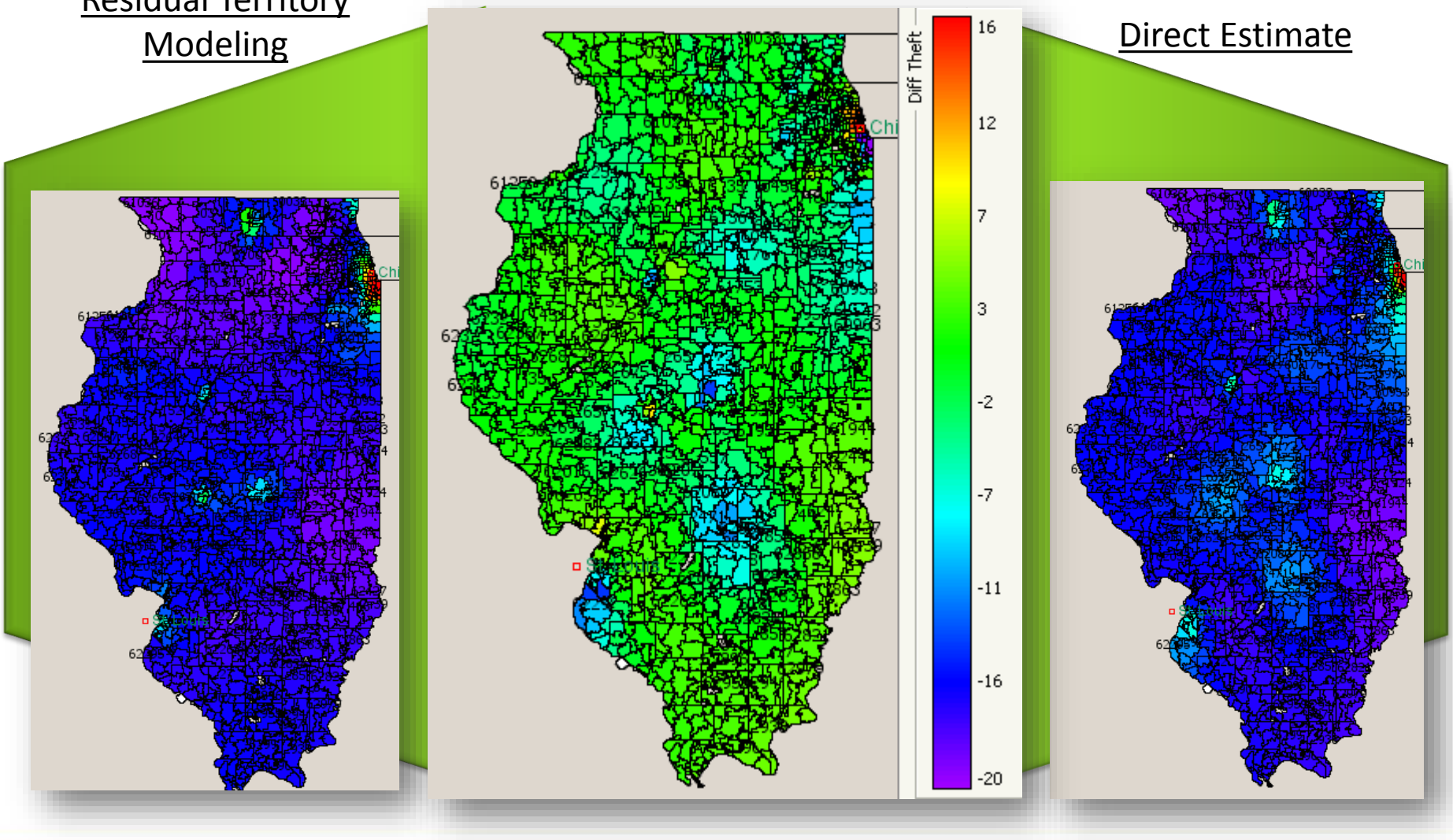
Comparison of Results – Theft Peril

EARNIX

Residual Territory Modeling

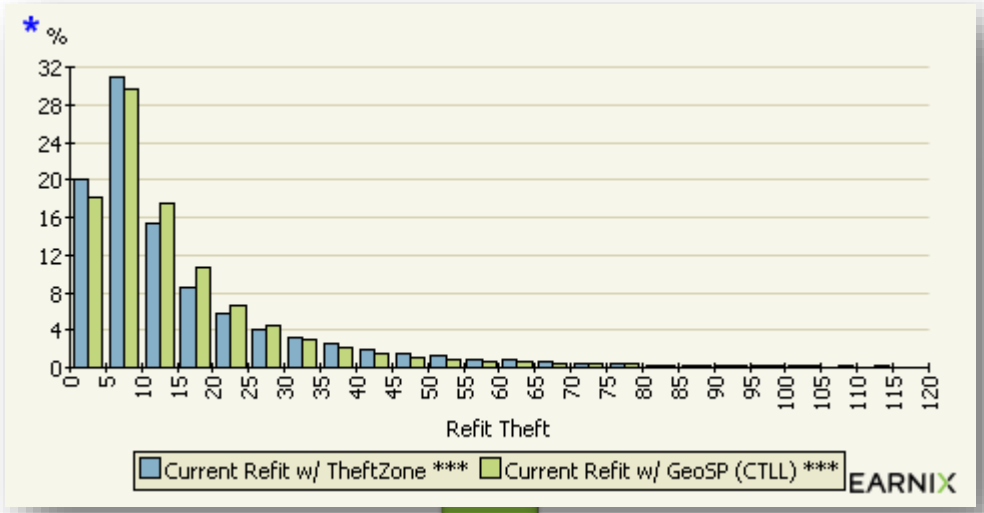
Difference in Theft Premium

Direct Estimate

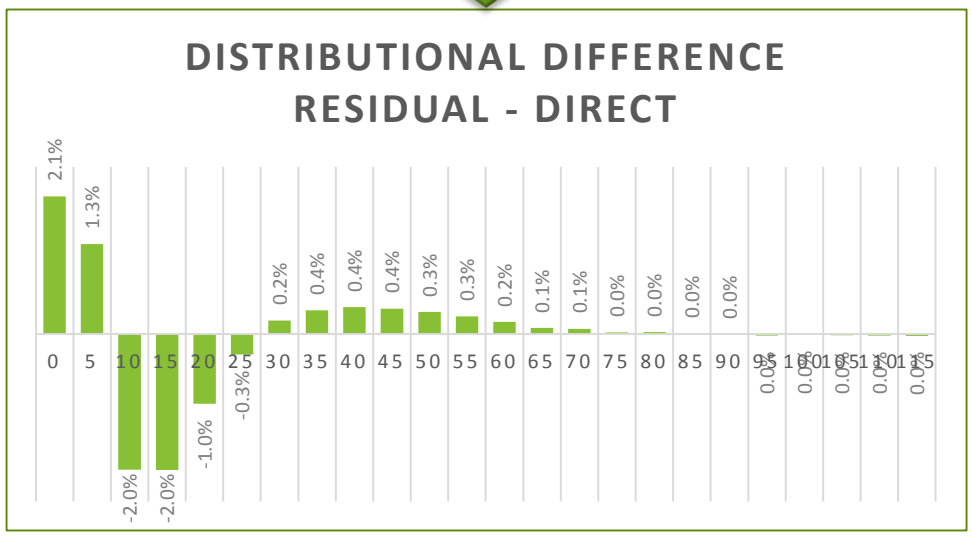


Comparison of Results – Theft Peril

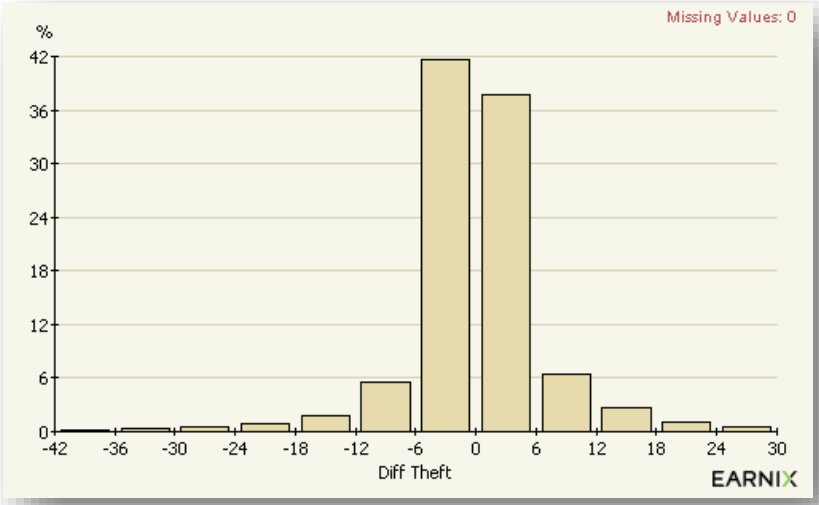
EARNIX



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



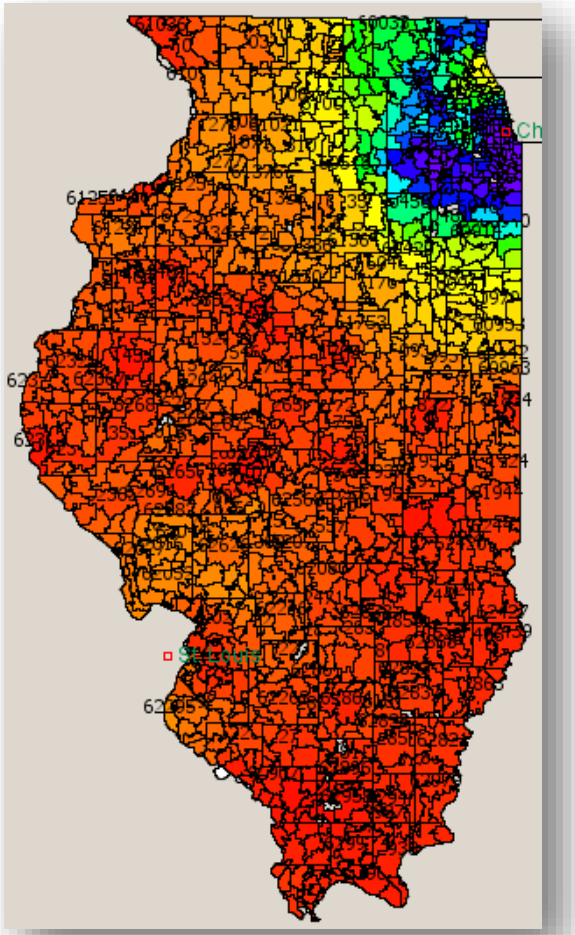
50% exposures within +-$2



Comparison of Results – Water Peril

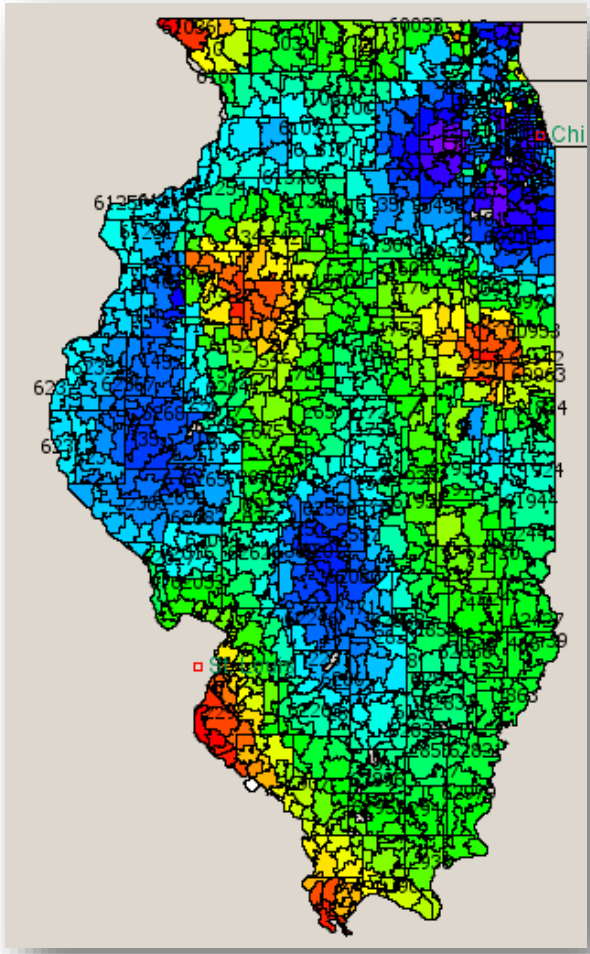
EARNIX

Residual Territory Modeling



≠

Direct Estimate



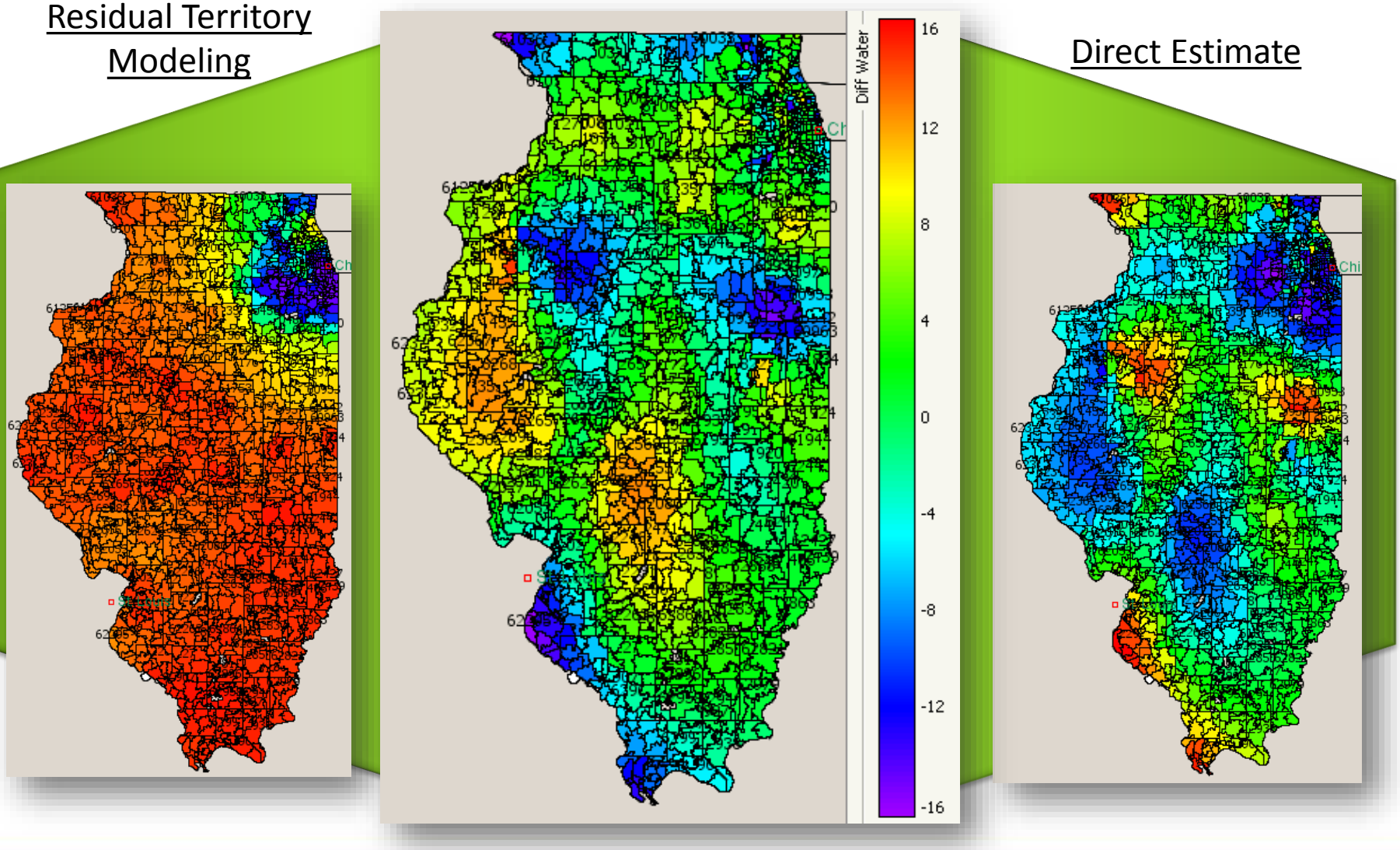
Comparison of Results – Water Peril

EARNIX

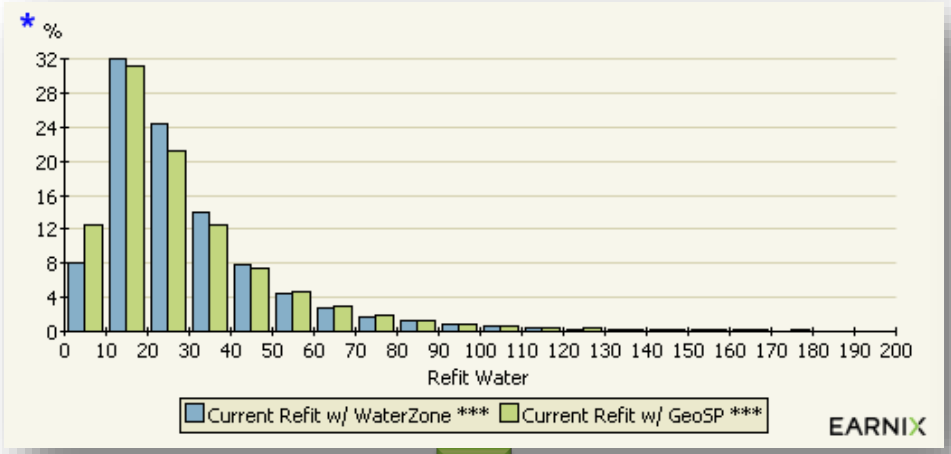
Residual Territory Modeling

Difference in Water Premium

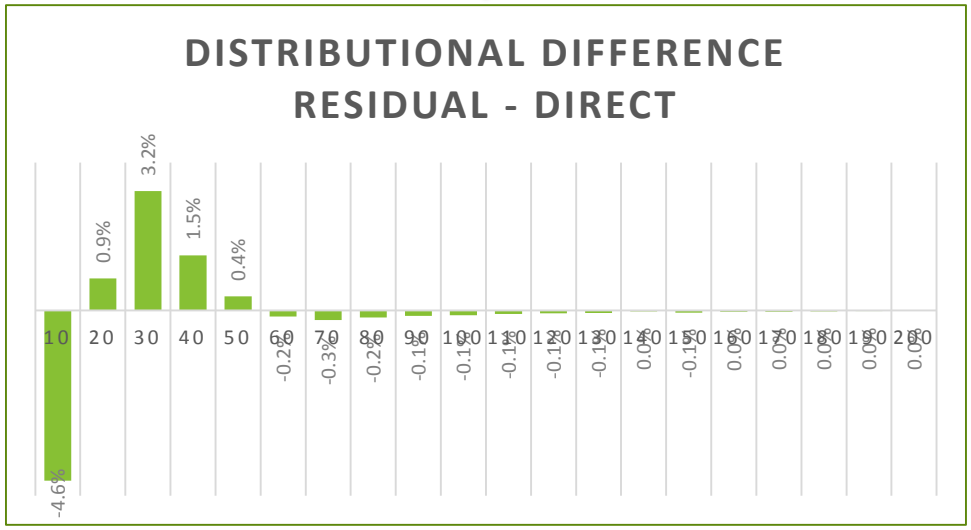
Direct Estimate



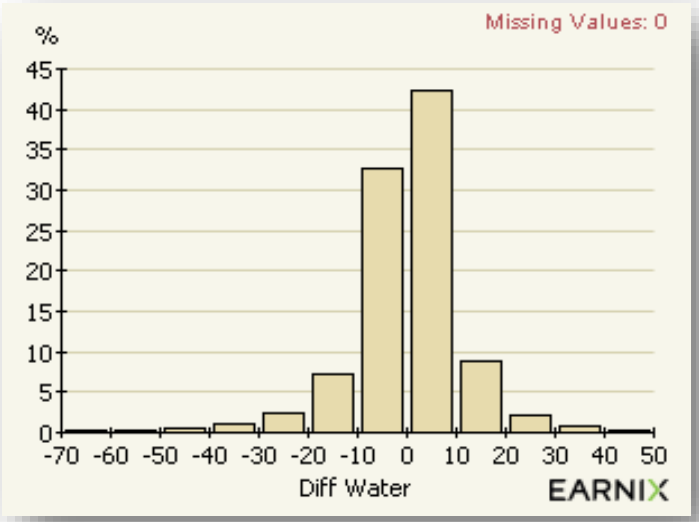
Comparison of Results – Water Peril



Range Analysis	Residual Modeling	Direct Estimation
1st Percentile	$6	$5
99th Percentile	$158	$168
Range (inner 98%)	$152	$163



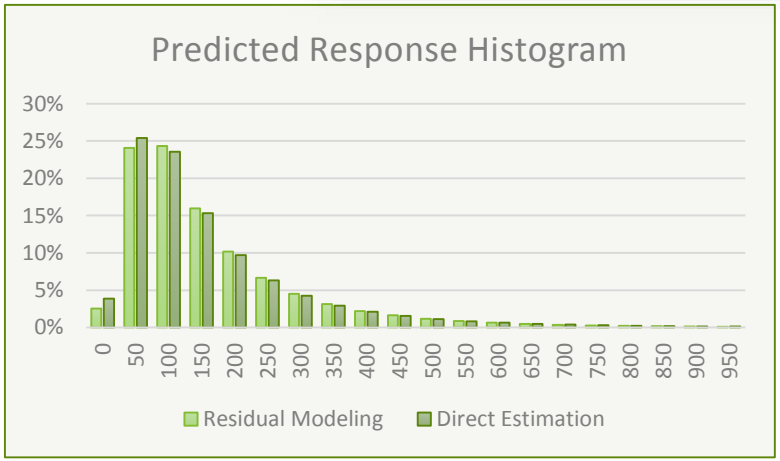
50% exposures within +-$5

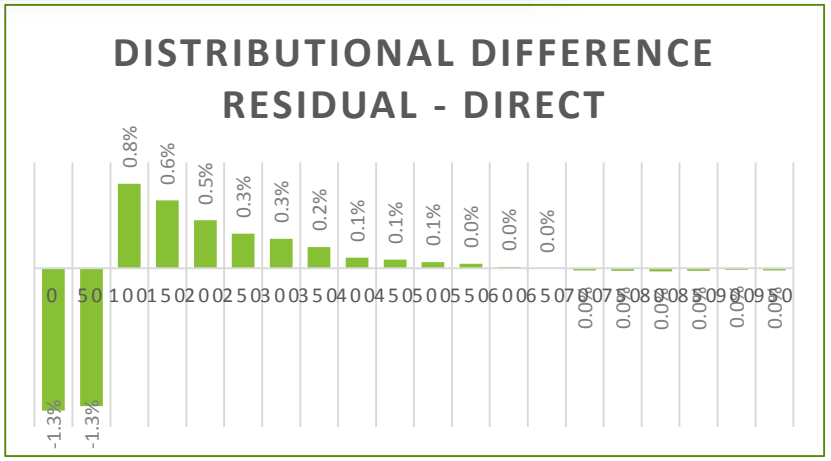


Comparison of Results – Combined Peril

EARNIX

Statistics (from holdout)	Residual Modeling	Direct Estimation
1st Percentile	$42	$38
99th Percentile	$808	$872
Range (inner 98%)	$766	$834 (9%)



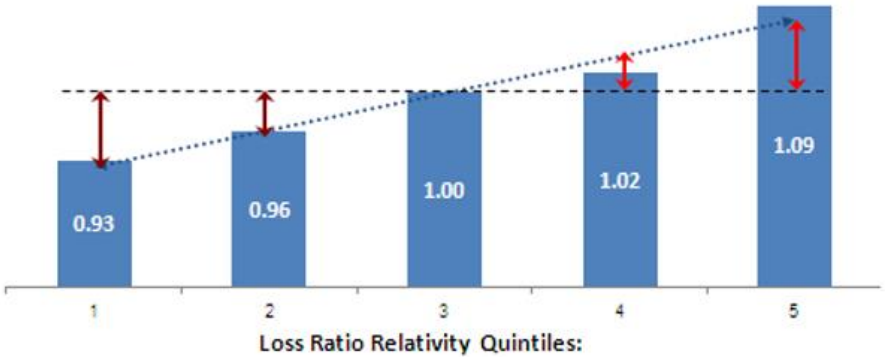


Additional segmentation is useless if segments do not result in <u>better risk classification</u>

Comparison of Results

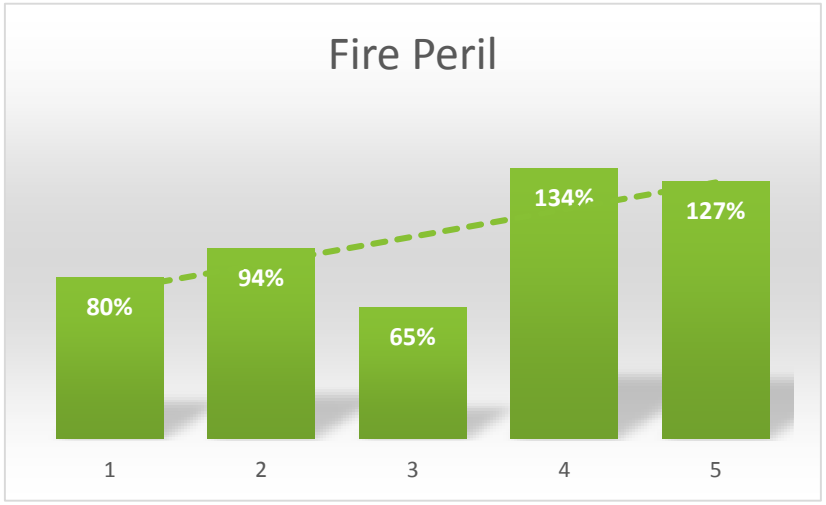
EARNIX

- Lift charts used to compare the results
- **Out-of-time** (2013) premiums were compared ~ Direct / Residual

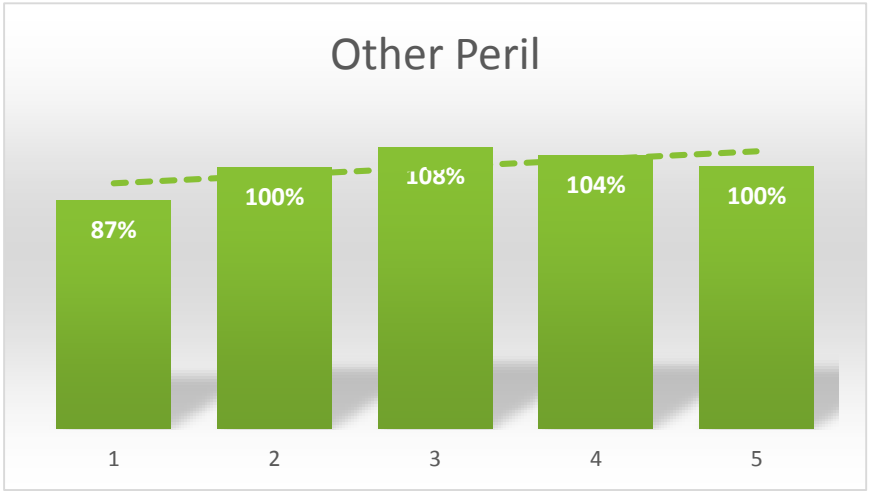


- The ordered values are bucketed into 5 equal exposure quintiles
- The loss ratio was then observed by comparing the observed losses to the current average premium within the group – Residual Premium
- Bars to the left depict where Direct Estimation approach predicts lower than Residual Estimation
- 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)

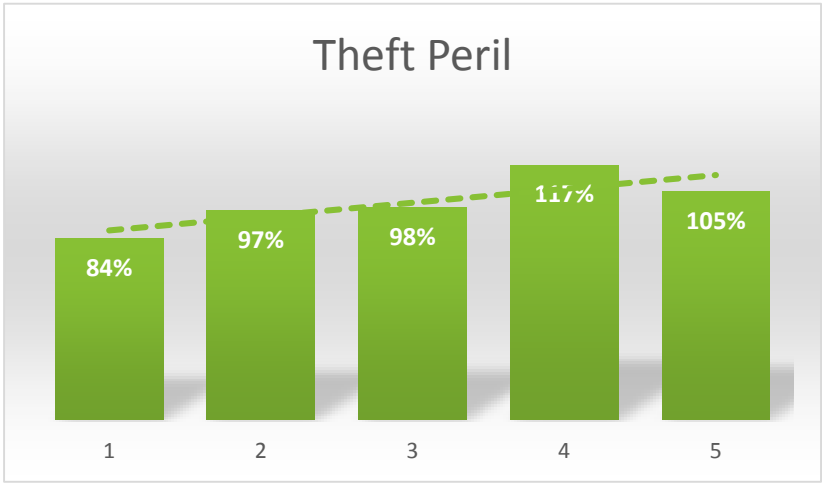
Lift Chart Analysis



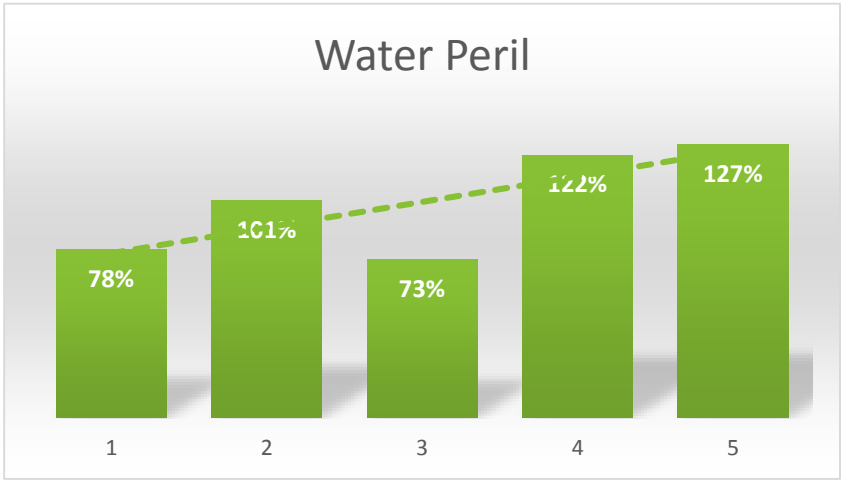
Fire Lift = (127% / 80%) – 1 = **80%**



Other Lift = (100% / 87%) – 1 = **15%**

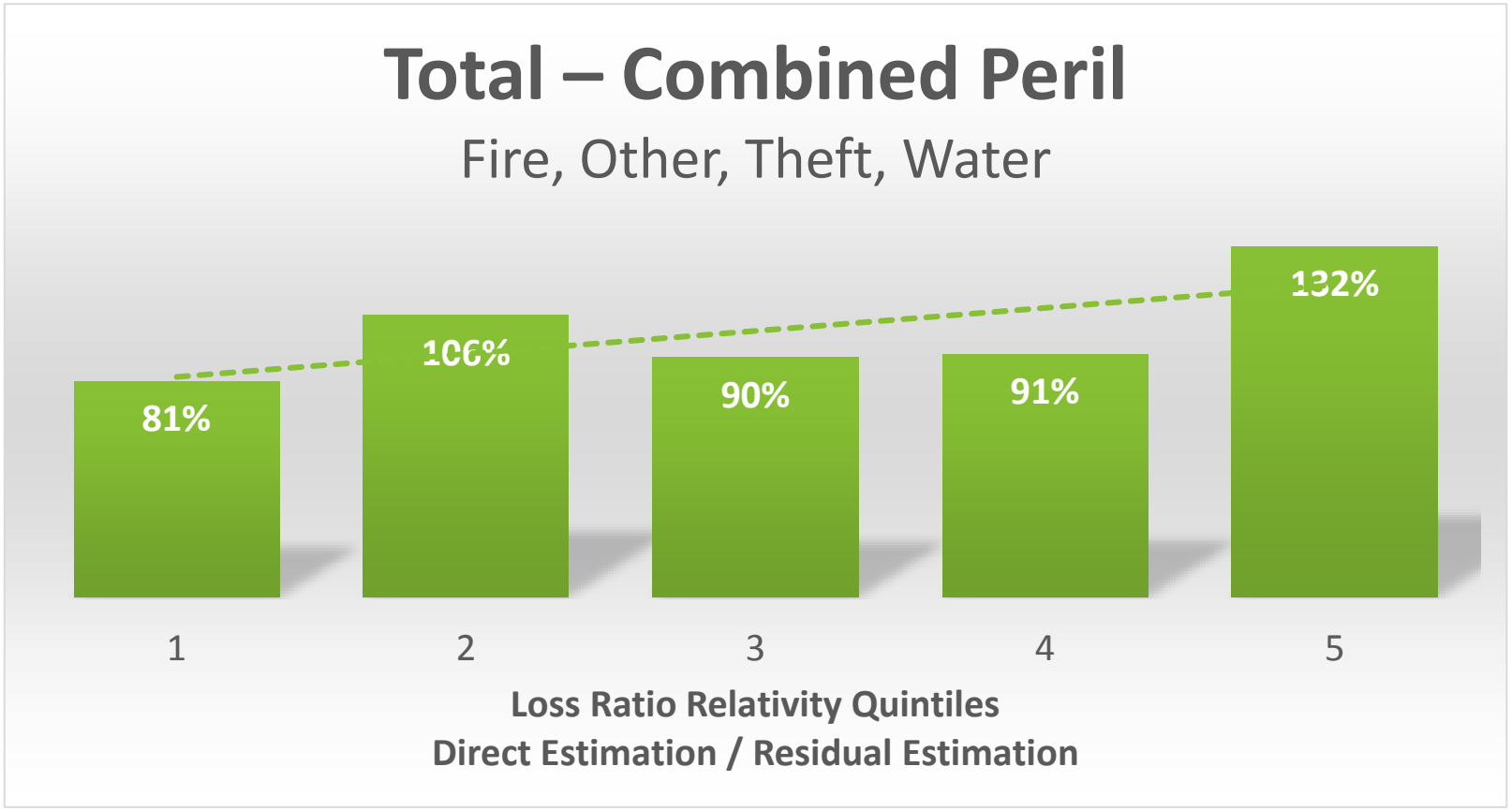


Theft Lift = (105% / 84%) – 1 = **25%**



Water Lift = (127% / 78%) – 1 = **62%**

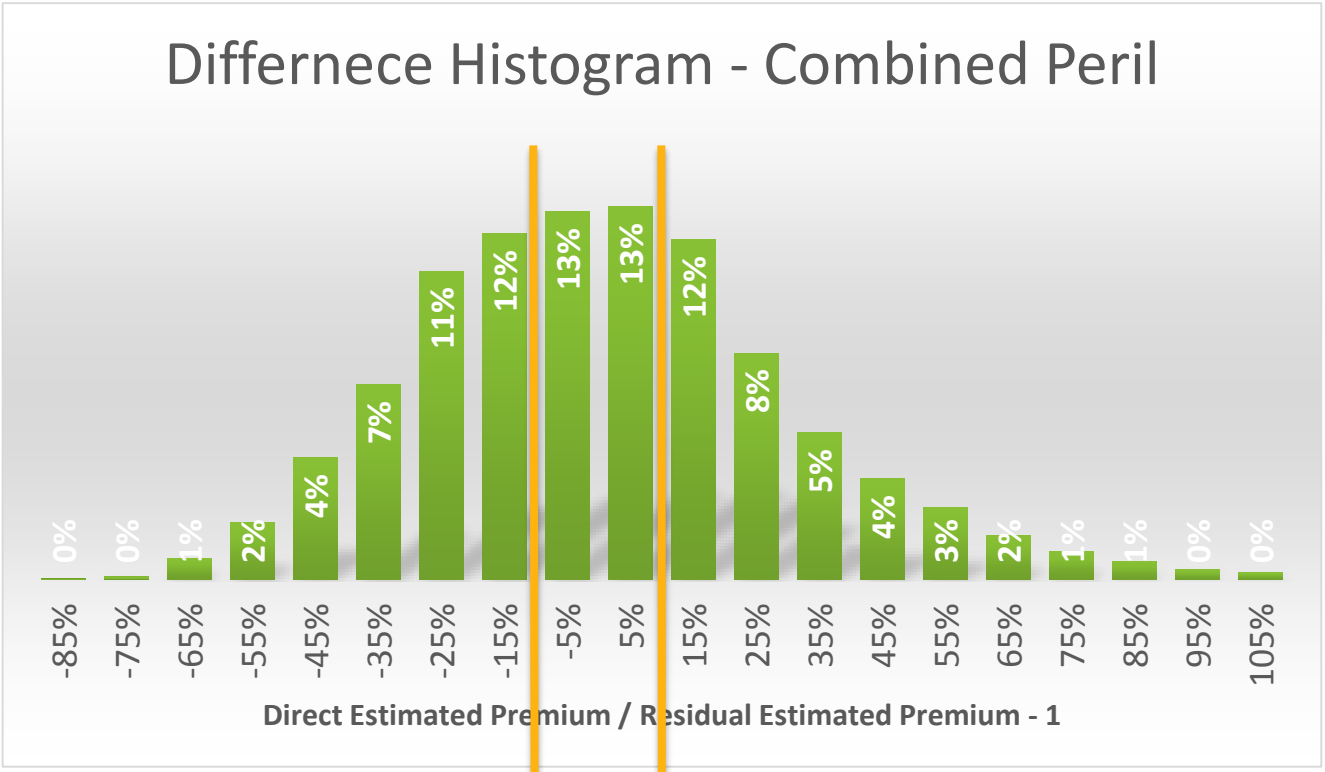
Lift Chart Analysis



Lift = (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

EARNIX

- 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

Comparison of Results

EARNIX

- Conclusion:
  - Both modeling techniques preform similarly on out-of-time sample

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

*once initial process is defined

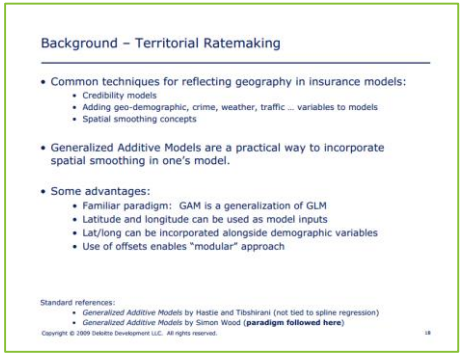
Thank You

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

EARNIX


Background – Territorial Ratemaking

- Common techniques for reflecting geography in insurance models:
  - Credibility models
  - Adding geo-demographic, 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:
  - Familiar paradigm: GAM is a generalization of GLM
  - Latitude and longitude can be used as model inputs
  - Lat/long can be incorporated alongside demographic variables
  - Use of offsets enables "modular" approach

Standard references:
- *Generalized Additive Models* by Hastie and Tibshirani (not tied to spline regression)
- *Generalized Additive Models* by Simon Wood (**paradigm followed here**)

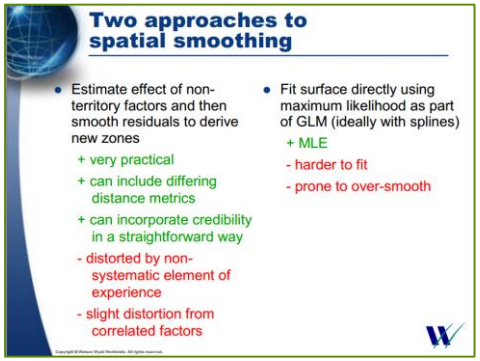
Deloitte.

WISCONSIN SCHOOL OF BUSINESS

Geo-spatial Analysis with Generalized Additive Models

CAS Annual Meeting
Chicago
November, 2011

Jim Guszcza
Deloitte Consulting LLP
The University of Wisconsin-Madison


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