Building Characteristics Modeling for Homeowners

Why It's Hard To Do and Why It's Worth Doing

Anand Khare, FCAS, MAAA, CPCU

Before We Begin - Prerequisites

- Actuaries who understand predictive modeling
- By-peril rating program planned or in place
- Large amount of detailed data
- Know what you're getting into

Why You Need Actuaries

- Building a model is hard, and statisticians are great at building models, so why not leave modeling projects like this to statisticians?

- Building a model *is* hard, but building the model is a relatively small part of the process!

- Much more time-consuming problems:
  - Managing and interpreting the data
  - Dealing with business considerations

Why You Need By-Peril

Why You Should *Want* By-Peril

- By-peril techniques yield more refined models
- By-peril plans are more understandable
- Creating a rudimentary by-peril model based on an existing plan is not all that difficult
- Most of the big writers are already doing it
- They have expanding market share and lower combined ratios to show for their efforts

Let's Discuss Data

- Chances are that you have some lying around

- Chances are that it's not nearly good enough

- There are some remedies for this

- The remedies are not cheap

- (Please recall the title of this presentation)

Reasons That Data May Be An Issue

- Homeowners is a relatively low-frequency line
- By-peril approaches require even more data
- Many building characteristics variables are categorical with many sparse levels
- There are remarkably high levels of correlation between many pairs of building characteristics
- Building characteristics information may not have been collected accurately or for long

Data Sources

- Tomorrow
  - Policyholders
  - Producers
  - Underwriters
- Today
  - Tax Assessors
  - Title Records
  - Third Parties

Policyholders

- Don't want to deal with the inconvenience of providing a lot of data just to get a quote
- Characteristic definitions may be unclear
- May feel it is in their best interest to report incorrect data or to simply make numbers up
- Need to worry about new vs. renewal process
- It will take a lot of time to gather enough data
- Primary data source – hard to get any better

Producers

- Don't want to deal with the inconvenience of providing a lot of data just to get a quote
- Characteristic definitions may be unclear
- May feel it is in their best interest to report incorrect data or to simply make numbers up
- Need to worry about new vs. renewal process
- It will take a lot of time to gather enough data
- Can take advantage of third-party programs

Underwriters

- Or maybe elves?

Self-Sourced Title/Tax Assessor Data

- Availability and detail varies by jurisdiction
- Field availability will be inconsistent
- Characteristic definitions will be inconsistent
- Enormously expensive to acquire
- Could potentially acquire data soon-ish

Third-Party Data

- Availability and detail varies by jurisdiction
- Field availability will be inconsistent
- Characteristic definitions will be inconsistent
- *Moderately* expensive to acquire
- Could potentially acquire data *today*

Bonus Data Source: Satellites

Examples of Tough Questions

- How old is your roof?
  - . . . which part of it?
- How many square feet is your house?
  - . . . are we counting the garage?
- How many rooms are in your house?
  - . . . are we counting bathrooms?
- Do you have a pool?
  - . . . does it matter that it's inflatable?

Verisk Insurance Solutions | ISO   AIR Worldwide   Xactware

What Style of Home is This?

Ambiguity and Inconsistency

- Major consideration for every data source
- Larger issue when combining data sources
- Geographic correlation or interaction with "missing-ness" of characteristic variables may force wholesale abandonment of certain fields
- May need to collapse tougher variables
- Might be better to stick to fields that are common across many or all data sources
- Might be better to avoid difficult questions

Modeling Considerations

- How many perils do you want to use?

- Which perils do you want to use?

- How do you want to handle catastrophes?

- Frequency/Severity or Pure Premium?

- How deep do you want to go?

Number and Choice of Perils

- What level of detail is your current program?
- How high is the quality of your claim data?
- How high is the quality of your *old* claim data?
- Can you use text mining to clean things up?
- **Data volume is a serious concern**

Handling of Catastrophes

- The three dominant catastrophe models handle building characteristics differently

- You have two reasonable options:

  - Accept that a portion of your rate cannot reflect whichever building characteristics are not included in your preferred catastrophe model

  - Impose perturbed peril-specific building characteristic relativities on the portion of the rate associated with a catastrophic cause of loss that closely resembles the peril

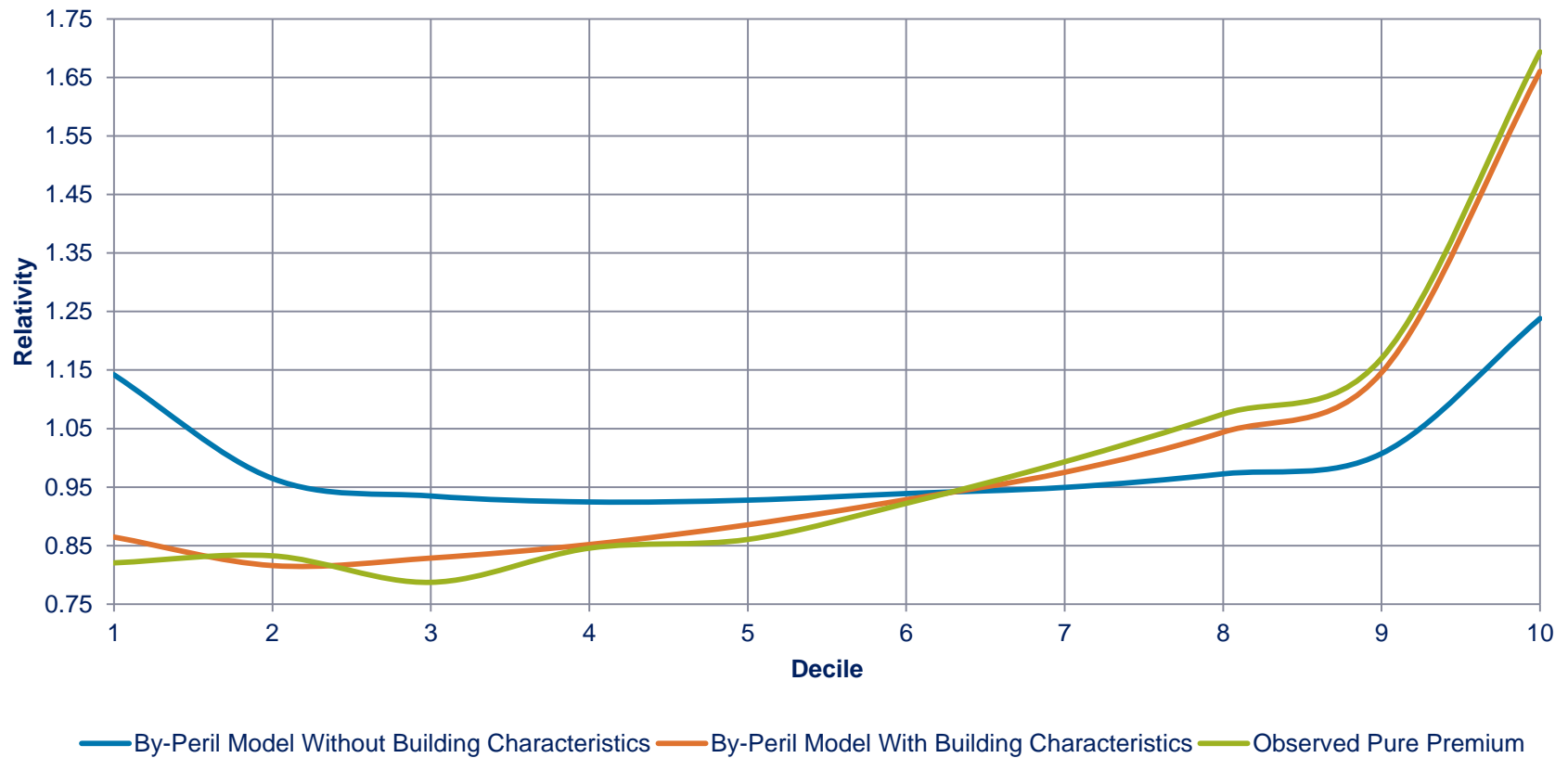# Frequency/Severity or Pure Premium

Time and Capital to Invest

- ROI is there but difficult to quantify
- **Have a plan before you get started**

There is Good News in All of This

Building Characteristics Lift

Legend: By-Peril Model Without Building Characteristics, By-Peril Model With Building Characteristics, Observed Pure Premium

There is Good News in All of This

Questions?