Exam 4: Introduction to Catastrophe Risk Management
Learning Objective No. 35
Study Note

Learning Objective 35: Understand the importance of data consistency to modeled results changes

Describe the reasons that modelled results may change over time in relation to exposure data consistency.

Quality of exposure data has significant influence on the catastrophe modeling outputs. The models require high-quality data not only for policy conditions, but location data (location conditions, risk characteristics and hazard profiles) as well. While companies keep enhancing exposure data quality, a good practice would be tracking data consistency using data audit. This study note will discuss key attributes to modeled output changes overtime, in relation to exposure data consistency.

1. Policy conditions
   Policy conditions usually contain the financial terms and conditions of the (re)insurance contracts. The key elements are but not limited to the following:
   a. Deductibles:
      - Percentage vs. absolute amount:
        - Whether percentage based deductible and absolute amount ones are captured in the system, completely, correctly and consistently.
        - Whether there is batch-coding to covert the two catalogue when system changes.
      - All perils vs. Sub-peril deductible:
        - Are these catalogues coded completely, correctly and consistently from contract language? For example, 'Fire' policies are coded with wildfire and quake fire following, along with applicable sub-peril deductibles.
   b. Limits and attachments:
      - All perils vs. Sub-peril limits
        - Similar to when tracking deductibles, are these limits coded completely, correctly and consistently from contract language? For example, 'Flood' limit are coded for inland flood, precipitation flood and coastal flood.
      - Blanket vs. coverage specific
        - Whether blanket limit and coverage specific ones are tagged with correct underlying coverage.
        - Where overall limit is not available, is there bulk-coding with consistent definition of policy maximum liability? For example, for warehouse endorsement where sum insured may fluctuate by seasonality, is there a clear coded definition of maximum liability per endorsement and per location?
   c. Shares and participation:
      - Are these fields consistently coded or fed from internal systems?
   d. Rate on Line and premium:
      - Are these fields consistently coded or fed from internal systems?
   e. Reinstatements:
      - For reinsurance treaties, are number of reinstatements coded completely, correctly and consistently from contract language? Especially for complex reinsurance programs, when there are multiple sections or perils sharing reinstatements.
   f. Inception and expiry date:
      - Are these fields consistently coded or fed from internal systems?

2. Location data

Modeled output, especially of wildfire and inland flood, are highly sensitive to the accuracy of geographic data, hazard profile and risk characteristics, on detailed location level. While companies continue to collecting more granular data, there are cases where aggregate data with reduced geographic resolution is the most readily available data for modelling. This section will discuss various aspects to consider when evaluating data consistency for both detailed and aggregate data.

- Detailed
  a. Geocode resolution
     - Is the overall geocoding resolution consistent for the underlying portfolio over the years? E.g. similar percentage of locations are coded with exact address.
     - Is the trend consistent with any operation or IT system enhancement?
     - Are there any changes to the geocoding engine used? For example, if the company switches from Pitney Bowel to Trillium in geocoding locations outside of US, the output could see differences in both hit ratio and high resolution percentage.
  b. Risk characteristics and hazard profiles
     - Is the overall risk profile for the underlying portfolio over the years? E.g. similar percentage of locations are coded as single family house and this is consistent with company underwriting strategies. If data shows a spike in locations coded as mobile home, this could be a flag to check if companies change portfolio or any bulk-coding errors.
     - Is there bulk-coding for the underlying risks? Is the data coding assumptions consistent over the years?
- Aggregate
     - Are aggregated data treated with similar level of scrutiny as detailed data? Data coding assumptions are documented properly, including coverage/line of business mapping (residential, commercial, engineering etc.), currency conversion and consistent data source.
     - Does the geocode change over the year? For example, for CRESTA-level aggregates in Turkey, are there changes of CRESTA definition or mapping?

3. Key aspects of data consistency:
   To summarize the above two sections, when evaluating data consistency of underlying portfolio or policies, the following aspects should be included:
   - Completeness
     Examples of queries could be:
     - What percentage of locations contains an exact street address?
     - How many buildings have occupancy and construction type given?
     - How many locations have unknown roof type?
     - What is the as-of date of the underlying data?
   - Augment
     When data is not incomplete or incorrect, the company should have guidelines and assumptions on how to handle that data, with well documented around controls. Examples of tests could be:
     - Is the missing data included as a loading factor to modeled output? Or uses industry information as proxy?
     - For incorrect data, does the process replace it with third party data or average value?
   - Accuracy
     In practice, data accuracy is difficult to check. Examples of techniques used in internal risk review and data audit may include:
     - Comparison to company data inventory or industry database as benchmarks.

- Geo-browsers to check high value properties via Zillow, Google Earth or satellite images. For example, if the average property in specific states is significantly lower than real estate property value, sum insured could be outdated and underestimated.
- Sense check based on logical interdependency. For examples, negative property value, deductible higher than sum insured.