Large Loss Distribution and Regression Analyses

Andrew Kwon, FCAS

Luyang Fu, Ph.D., FCAS

Agenda

- **Section 1:** a least-squares based substitute for maximum likelihood, and extensions therein
- **Section 2:** incorporation of covariate terms and real-world case studies

Large losses matter

- **Ratemaking**
  - A required consideration for rate adequacy concerns
- **Reserving**
  - May not share properties with remainder of runoff
- **Reinsurance**
  - General foundation of excess of loss treaties
- **Risk management**
  - Disproportionately volatile compared to remainder of book; can impact capital decisions and risk mitigation strategies

Approaches

- **Risk loading**
  - "Skip" problem by inserting a calculated load into applications where large loss analysis would be relevant
- **Extreme value theory**
  - Fisher-Tippett-Gnedenko and Pickands-Balkema-de Haan
- **General distribution fitting**
  - Methods (moments, percentiles, least squares, maximum likelihood)
  - Distribution selection

Insurance complications

- **Losses** are often not recorded on a ground-up basis (ie. total loss to insured)
- **Policy terms** and conditions, including loss limits and deductibles, obscure the ground-up distribution of losses
- **"Undo"ing** these conditions can be valuable in understanding loss patterns not directly observable
- **Regardless** of data modifications, extremely unlikely to get tight fits

Why do we use MLE?

- **Favorable properties**
  - *Convergent:* solutions generally exist
  - *Consistent*: estimators converge to actual values as size of data set increases
  - *Unbiased:* only asymptotically (see above)
  - *Efficient:* leads to minimum variance unbiased estimates
  - *Normal:* asymptotic, but allows us to make statements about the volatility of estimates
  - *Explainable:* we maximize the probability of the data set occurring by adjusting the parameters

Why do we use MLE?

- **From the exam 4 syllabus:**

H. Construction and Selection of Parametric Models (25-30%)

1. Estimate the parameters of failure time and loss distributions using:
   a) Maximum likelihood
   b) Method of moments
   c) Percentile matching
   d) Bayesian procedures
2. Estimate the parameters of failure time and loss distributions with censored and/or truncated data using maximum likelihood.

Why try least squares?

- **Different distance functions, different estimates**
  - Not a given that likelihood or squared error is a superior measure of fit for any given purpose
  - Least squares implicitly weights larger observations more strongly by virtue of its distance function – good?

- **It's not just plain old linear regression**
  - Non-linear least squares fits are generally recognized and are available in statistical packages
  - Similar in execution to linear regression
  - Not without its own shortcomings, though

Setup and example

Refer to spreadsheet #1.

Curve fitting demonstration on idealized data

Selected fits, exhibits, and model selection thoughts on modified property data

Censoring adjustment

- **Value changes from point to interval**

- **Maximum likelihood treatment**
  - Numerator changes from PDF to some cumulative function
  - More or less accepted as given methodology

- **Least squares treatment?**
  - Initial focus of research
  - Several proposals found in literature, all reflecting a decrease in information indicated by a censored value

The EP’ algorithm

- Credit to Leo Breiman, Yacov Tsur, Amos Zemel
  - What if we “fill in” the missing information?
  - (E) step is a data transformation; only censored values change, value = last value + expected conditional error
  - (P) step is a numerical optimization; determine new parameters for distribution based on modified data
  - Repeat (E) step with change in expected errors, then (P), then repeat until least-squares estimator satisfactorily converges
- Simplifying assumption
  - What can we do to sidestep defining the error distribution?

Setup and example

Refer to spreadsheet #2.

Demonstration of maximum likelihood and least squares adjustments for censored observations on idealized data

Selected fits, exhibits, and model selection thoughts on modified liability data

Truncation adjustment

- Unable to access portion of the distribution
- Maximum likelihood treatment
  - Denominator changes from 1 to survival function
  - More or less accepted as given methodology

- Least squares treatment
  - Direct corollary – recognize we are not fitting on the entire [0,1] domain of possibilities
  - What if we allow the lower bound of the domain to vary with the change in fitted parameters?

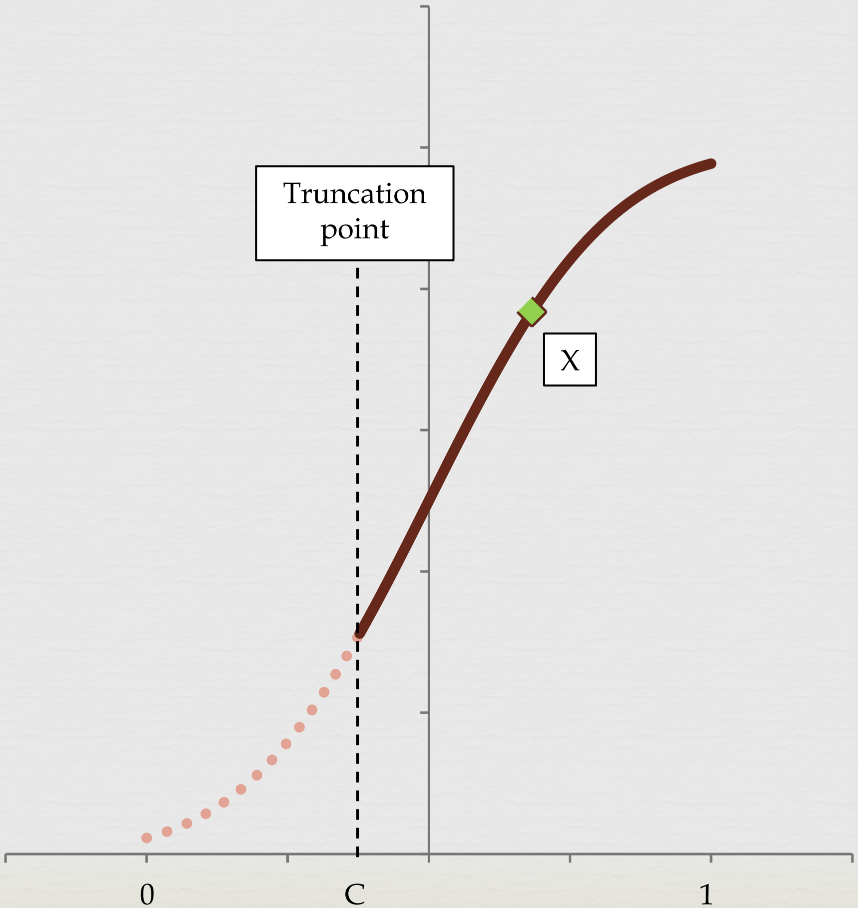
Transformation

The observed data point x falls in the probability interval [c,1], but in order to pull the correct inverse CDF, we need its probability over [0,1] instead.

If you envision a mixed distribution with weight F(c) on zero and S(c) on observed data, this provides a direct solution to the problem.

Because a CDF must be monotonically increasing, the F(c) weight generally comes first, so:

$F(p) = F(c) + S(c)F(p \mid p>c)$



Setup and example

Refer to spreadsheet #3.

Demonstration of truncation adjustments for maximum likelihood and least squares on idealized data

Practitioner's notes

- Still possible for optimization to fail on least squares
- Truncation workarounds in case of divergence
  - Solution one: judgmentally or analytically select a truncation point
  - Solution two: left-shift data instead and correct afterwards
- Methods have variable degrees of success coping with different modifiers and different data
- Modifiers to loss data can sometimes be more significant to fit quality than data itself

Business Mixes and Large Losses

- We just checked the distribution of loss (Y), not the predictors behind loss (X).
- Business mixes impact Property large loss propensities
    - Coverage A
    - Territory / protection class
    - Industry group
    - Sprinkler system
    - Other

Business Mixes and Large Losses

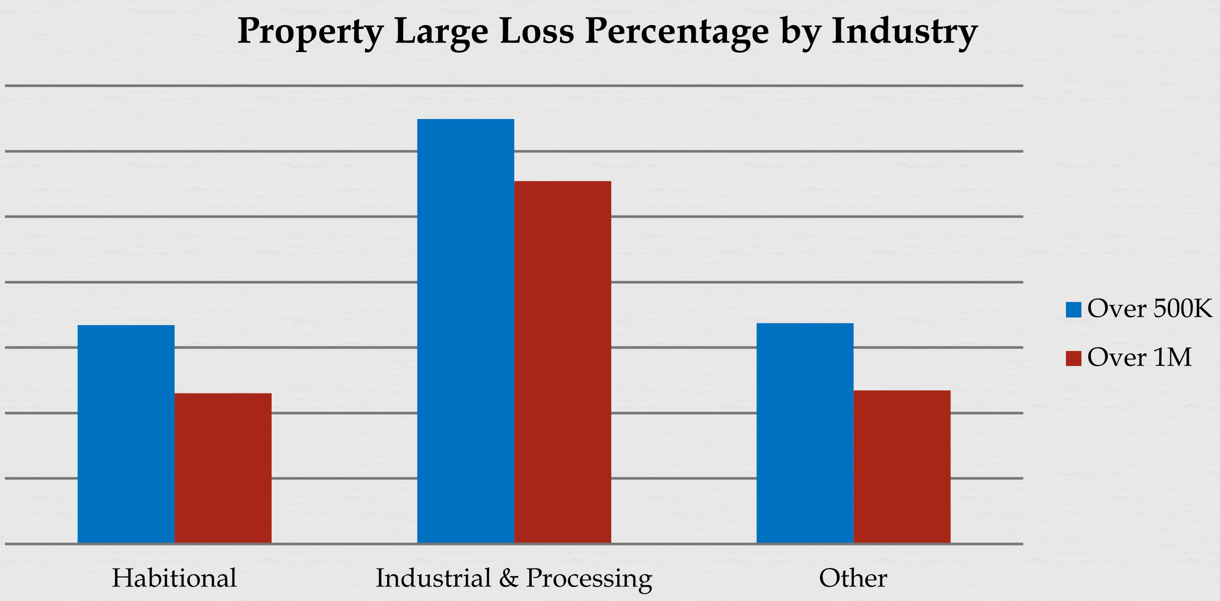
- Business mixes impact GL large loss propensities
  - Class or NAIC codes;
  - Sublines
  - Limit groups
    - Table 1,2,3 for PremOP;
    - Table A, B, C for Product and Complete Op
  - Contractors and subcontractors
  - Umbrella coverage
  - Other

Injury Mixes and Large Losses

- Injury mixes impact WC large loss propensities
  - AIA codes
  - ICD9
  - CPT
  - Class and NAICS
  - Drugs
  - Hazard Group
  - Etc.

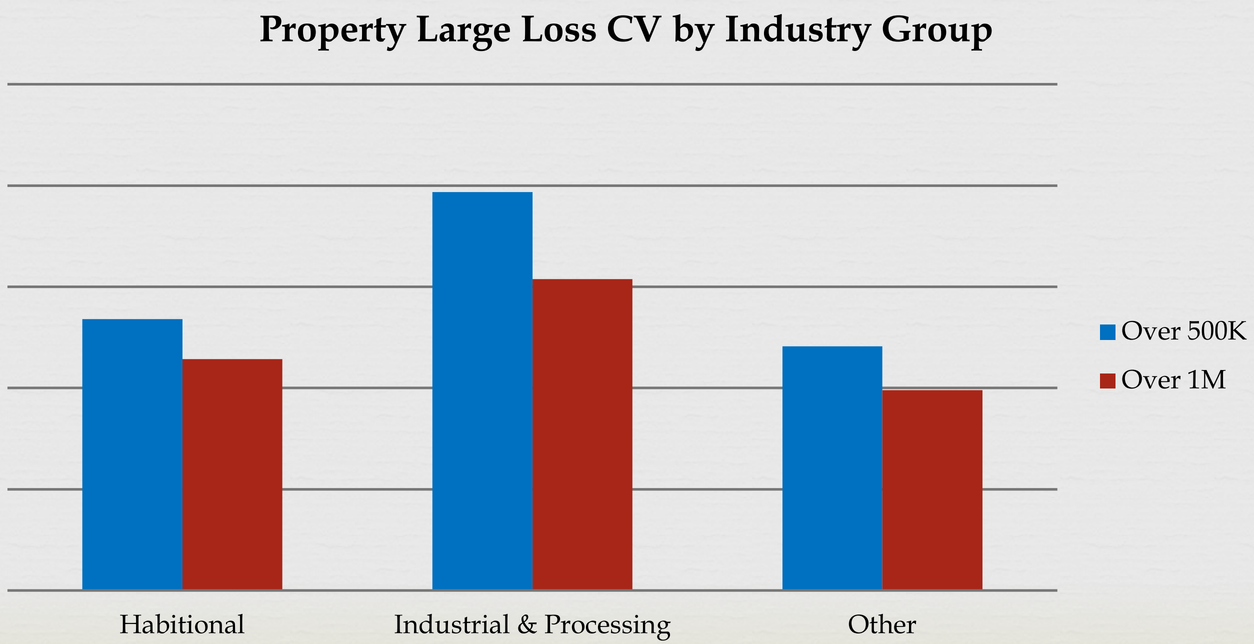
Commercial Property Large Loss

- Property large loss by industry group



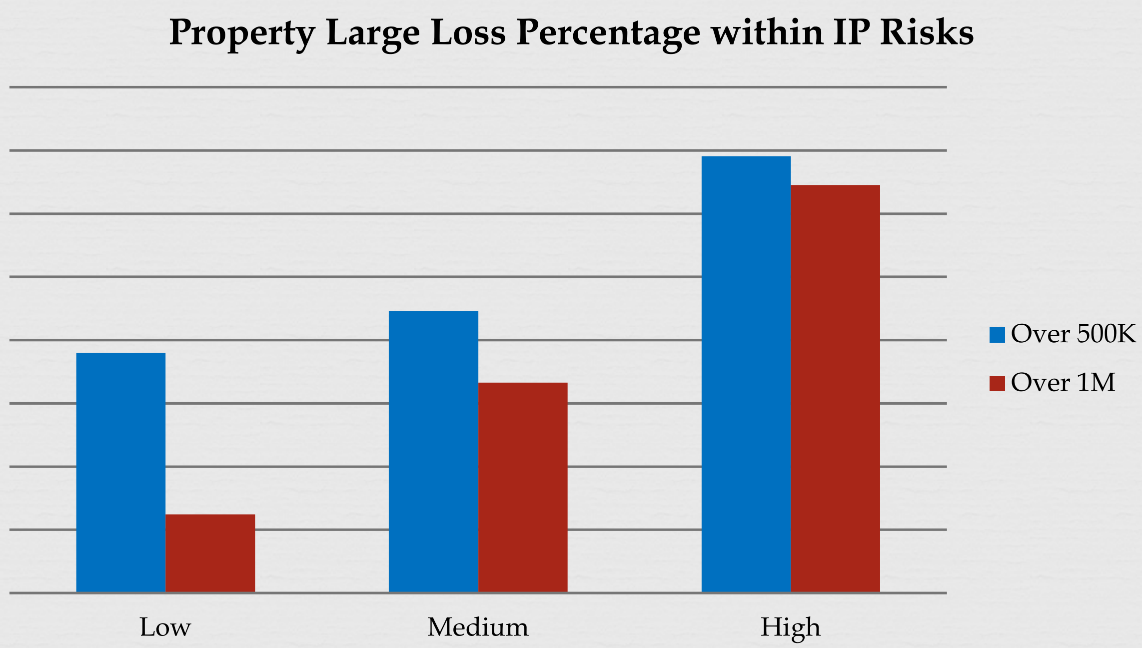
Commercial Property Large Loss

- Loss distribution is defined by both mean and volatility
- Volatility is very important for reinsurance pricing and ERM
- Volatility is often heterogeneous



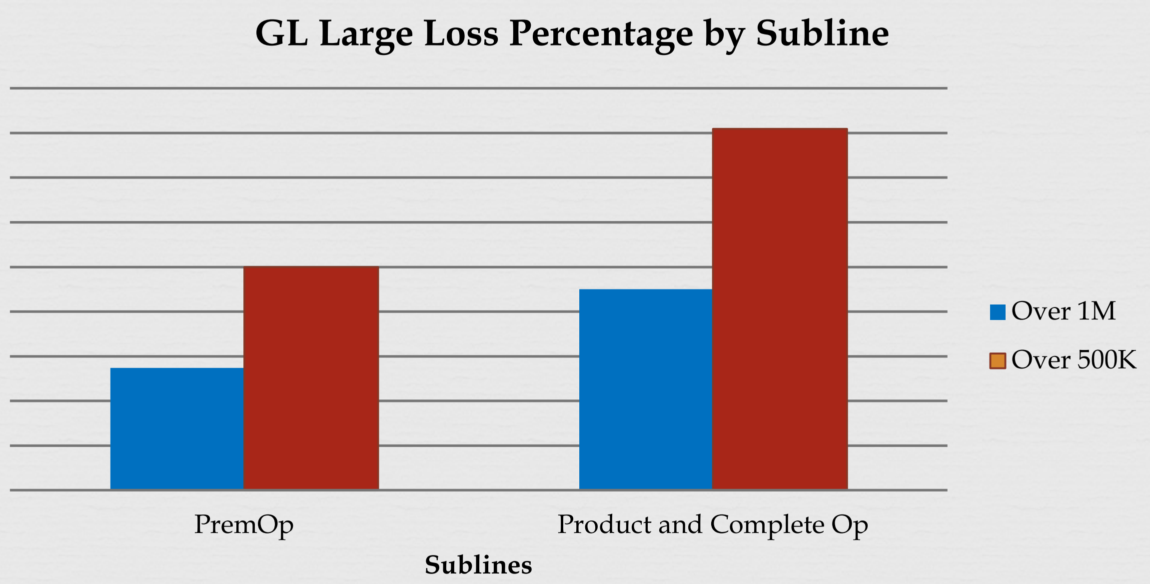
Commercial Property Large Loss

- Property large loss within IP by manufacture type



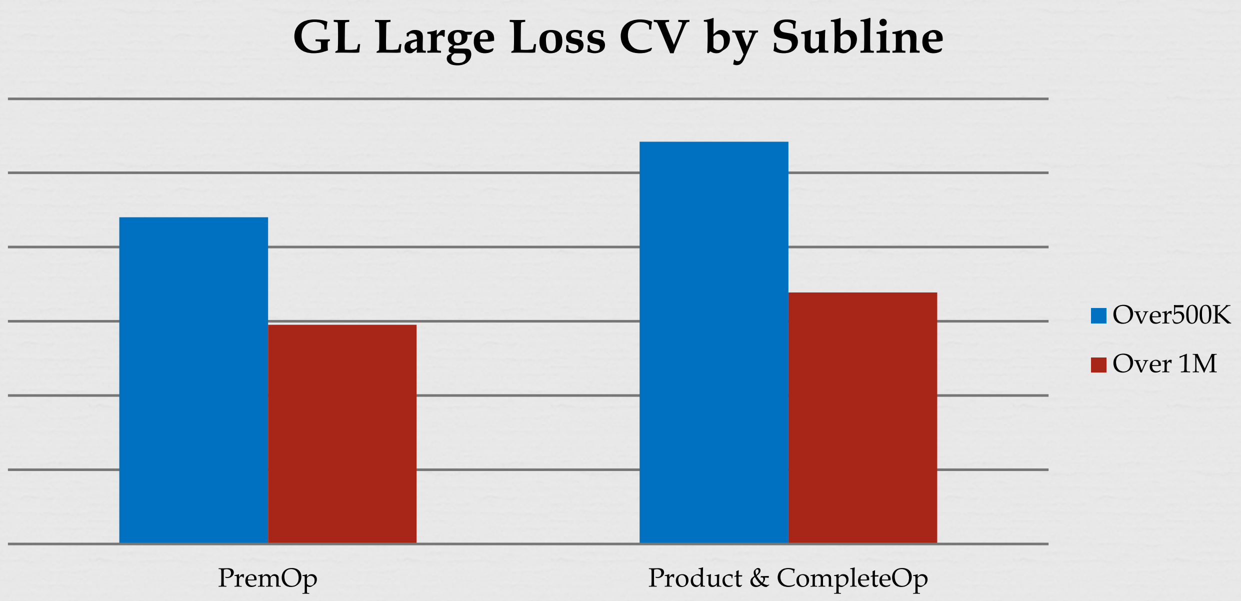
General Liability Large Loss

- GL large loss by subline



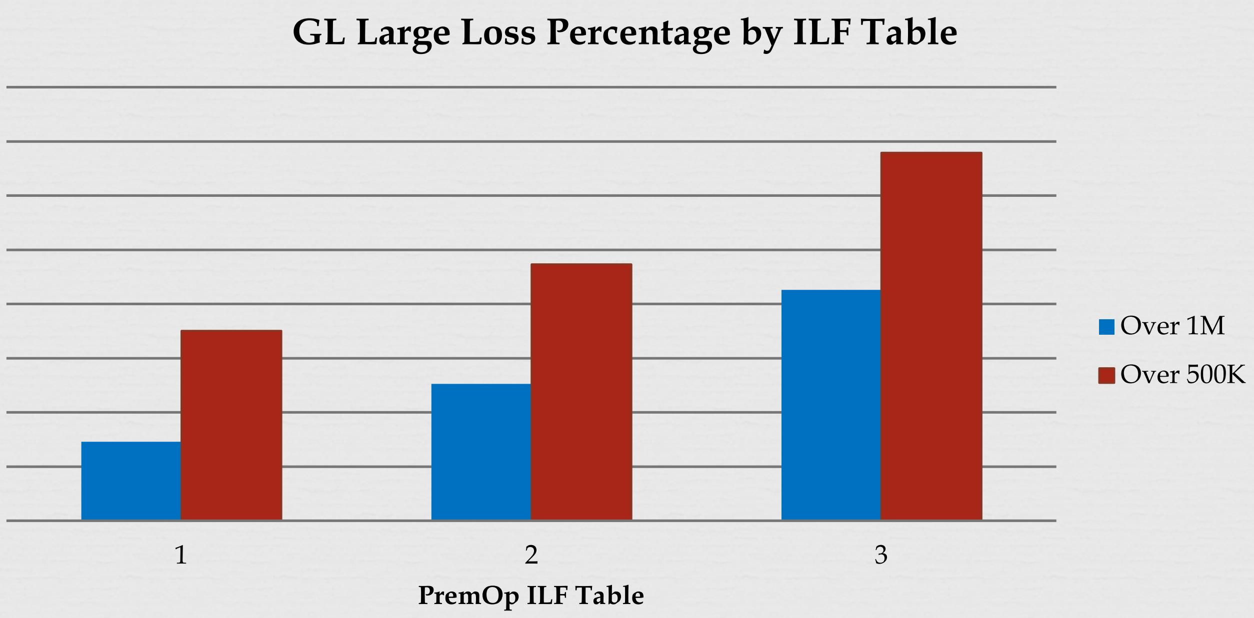
General Liability Large Loss

- GL large loss coefficient of variation by subline



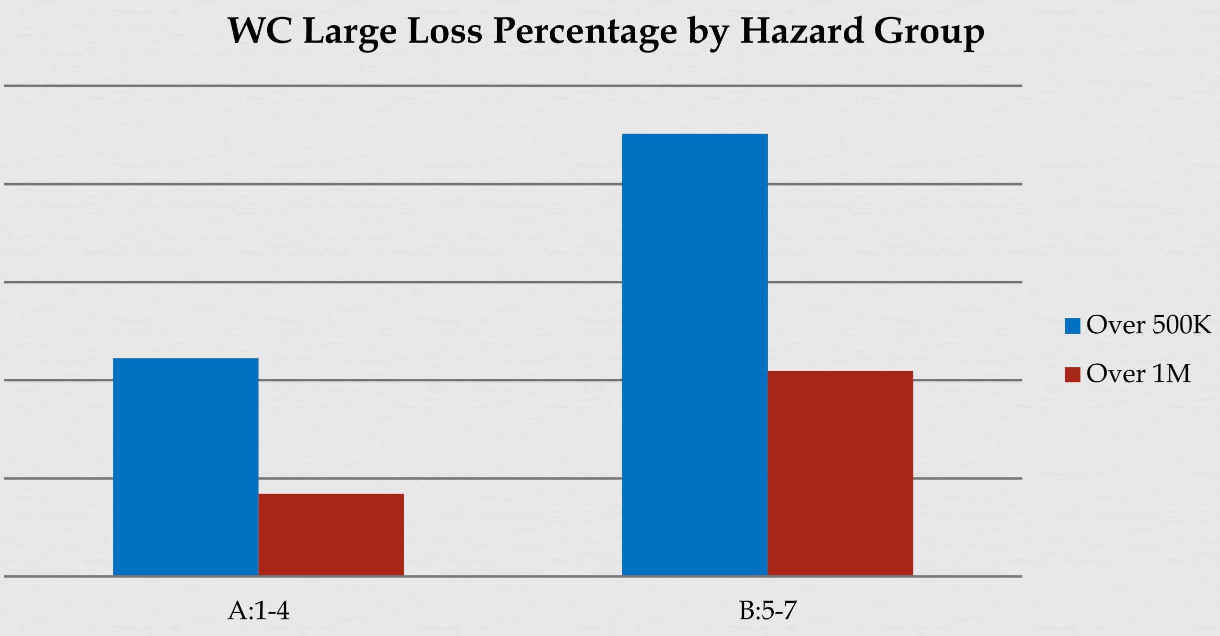
General Liability Large Loss

- GL large loss within PremOp by ILF table



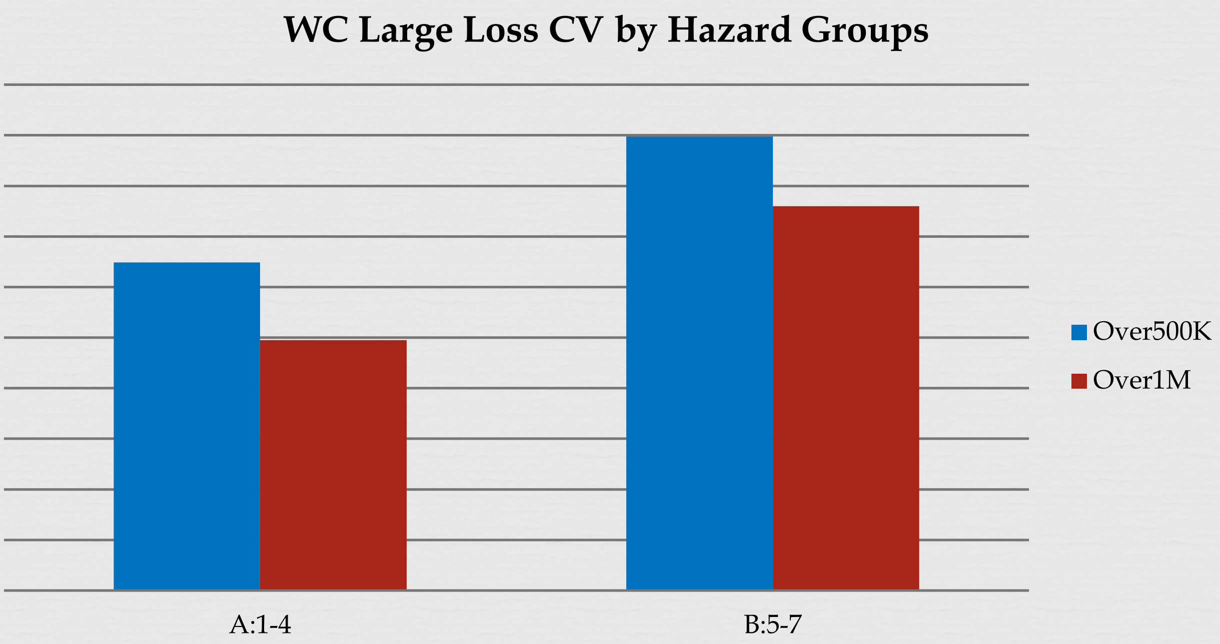
WC Large Loss: Claims Perspective

- WC large loss by hazard group



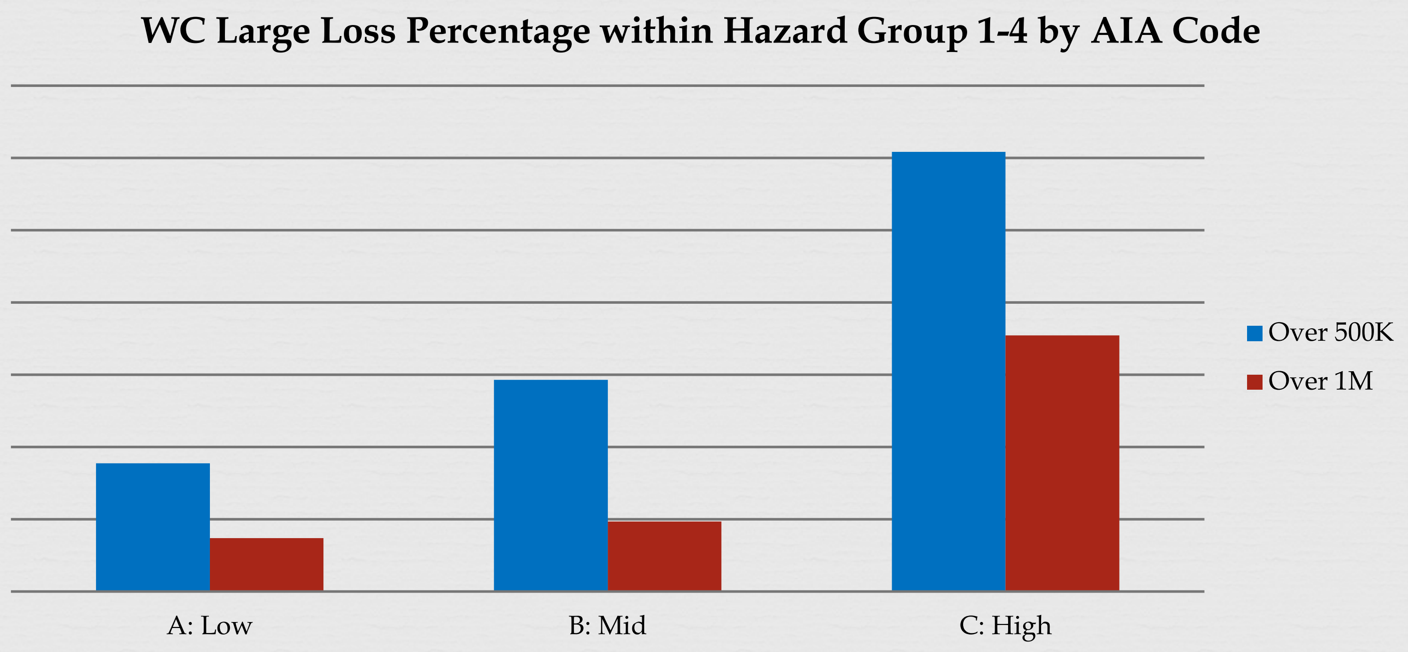
WC Large Loss: Claims Perspective

- WC large loss coefficient of variation by hazard group



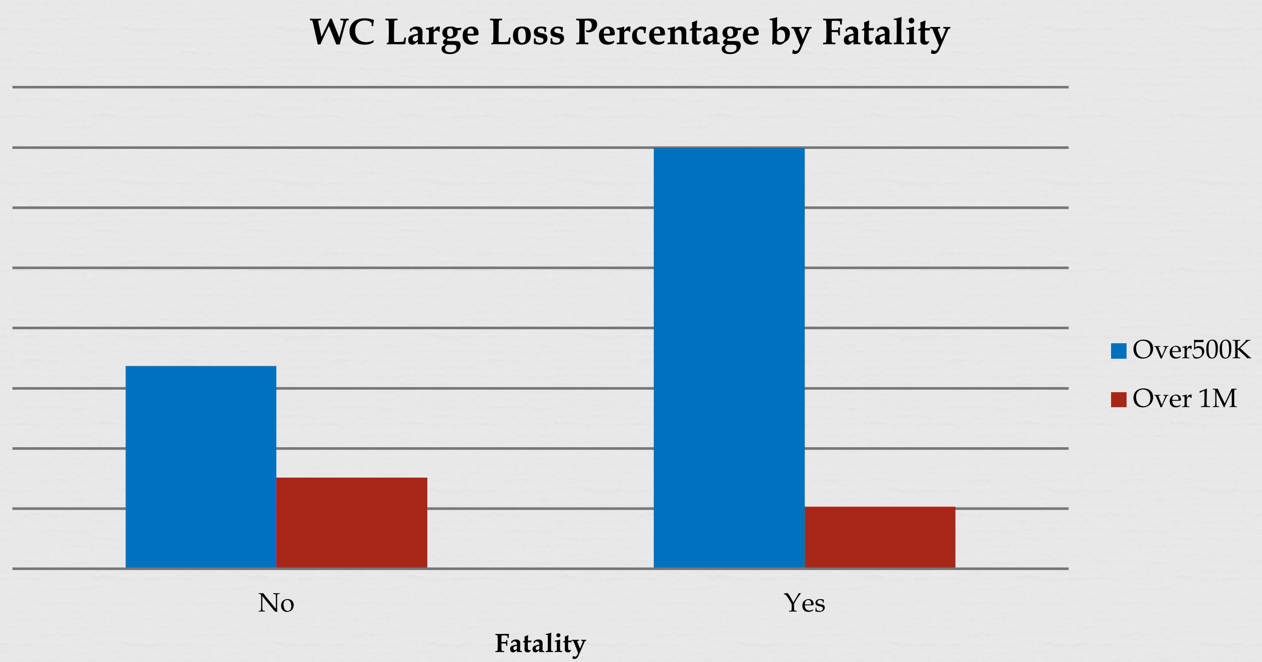
WC Large Loss: Claims Perspective

- WC large loss within hazard group 3 by AIA code



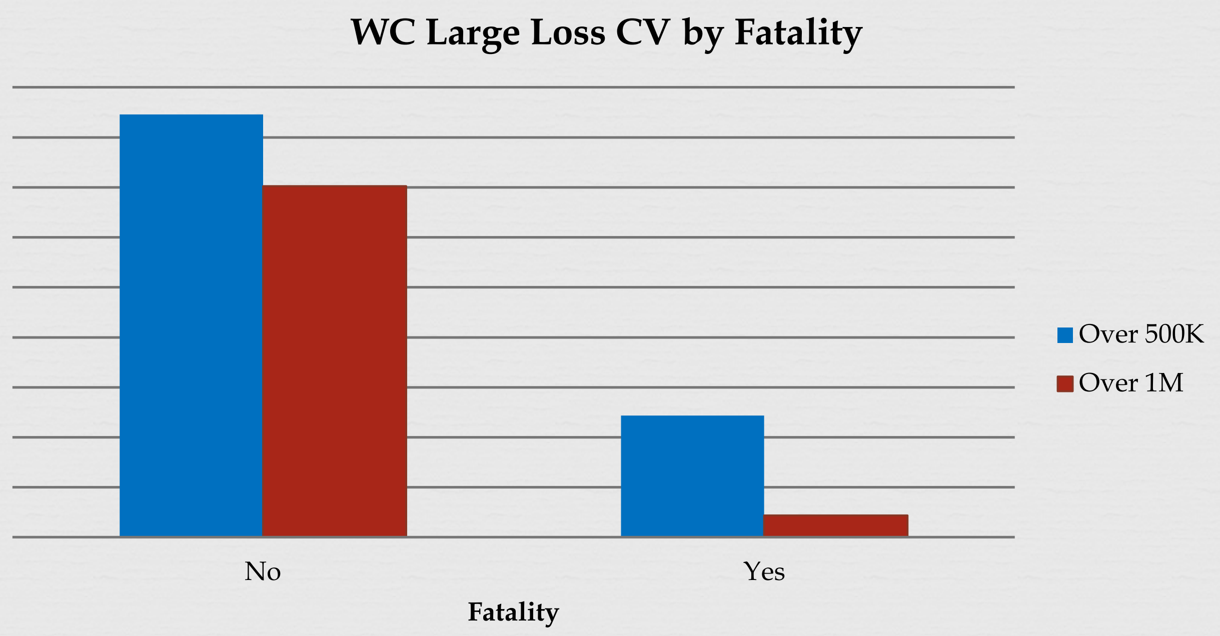
WC Large Loss: Claims Perspective

- WC large loss by Fatality



WC Large Loss: Claims Perspective

- WC large loss coefficient of variation by fatality



Regression Models

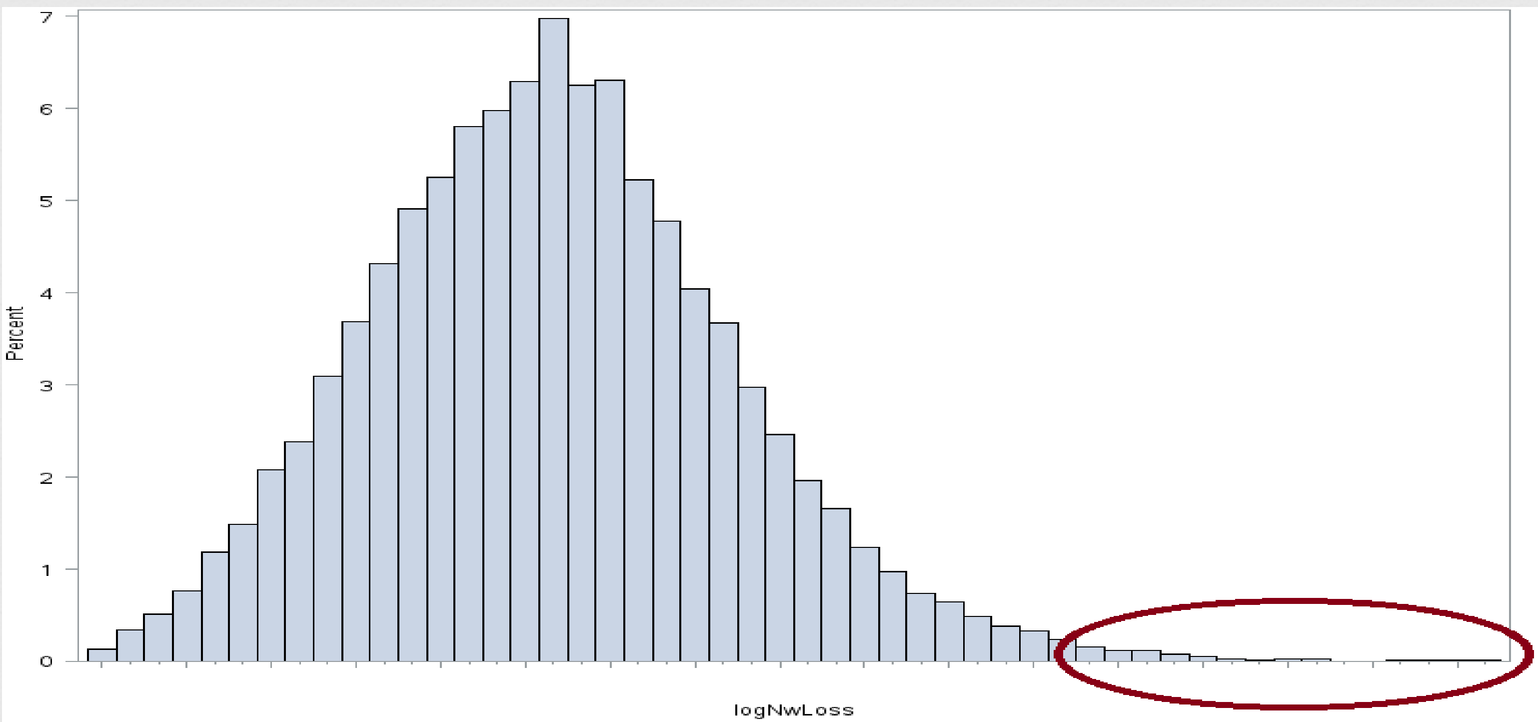
- Model individual policies instead of whole book
  - Contemplate underlying risk characters
  - Granular trending: by peril, by subline, etc.
  - More work than conventional distribution fitting
- Large loss frequency: logistics or GLM
- Severity: GLM, log-linear, and other more complicated models

Regression Models

- Severity Model: Double GLM
- Certain risks can be much more volatile, which implies that GLM dispersion factor may not be constant,
  - Traditional GLM:
    - $mean(loss) = \hat{y} = \exp(X\beta_1)$
    - $variance(loss) = constant * \hat{y}^{\alpha}$
    - X is the vector of predictive variables including the constant term
    - $\hat{y}^{\alpha}$ is the variance function.
  - Double GLM: heterogeneous variance
    - $mean(loss) = \hat{y} = \exp(X\beta_1)$
    - $variance(loss) = \exp(X\beta_2) * \hat{y}^{\alpha}$
    - $X\beta_1$ is the first GLM to fit mean; $X\beta_2$ is the second GLM to fit for heterogeneous dispersion factor

Regression Models

- Severity Model: Finite Mixture model
  - Single distribution usually does not model heavy tail densities well



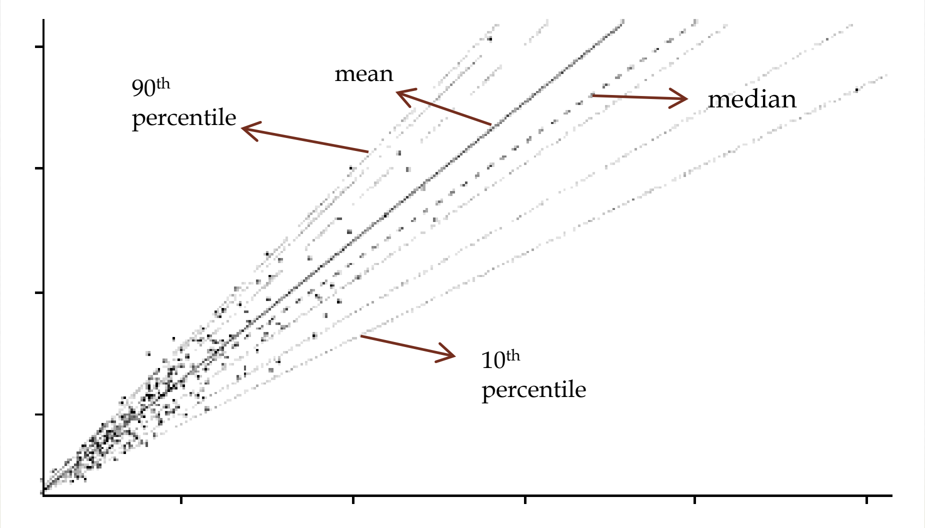
Commercial property loss has a longer tail than lognormal

Regression Models

- Severity Model: Finite Mixture model
  - $f(X, \beta_1, \beta_2) = \pi(X) * f_1(X, \beta_1) + \left(1 - \pi(X)\right) * f_2(X, \beta_2)$
  - f1 is normal loss distribution; f2 is severe loss distribution; X is the vector of predictive variables; $\pi(x)$ is the probability of being in a severe distribution
  - $\pi(x)$ varies by business mix. The probability of plastic manufacturers to be in severe distribution is much larger than average book

Regression Models

- Severity Model: Quantile Regression
  - Tail performance can be very different from the mean
  - Predict percentiles of potential loss other than just mean or variance
  - Robust and less sensitive to extreme values



Regression Models

- How to deal with data censorship
  - Tobit
  - Double GLM and FMM with censoring data
    - Solve the maximum likelihood function directly
      - Numerical solutions through R or SAS (Proc nlmixed)
    - EP (Expectation and Projection) algorithm
      1. E: Run regression using censored data
      2. P: fill those censored losses with predicted value from the "E" step
      3. E: refit the model using the fitted values on censored records
      4. Redo P step and keep iterations

Case Studies

- Case studies will be presented in the seminar