Finite Mixture Models and WC Large Loss Regression Analysis

Luyang Fu, Ph.D., FCAS

Xianfang (Frank) Liu, Ph.D., FCAS

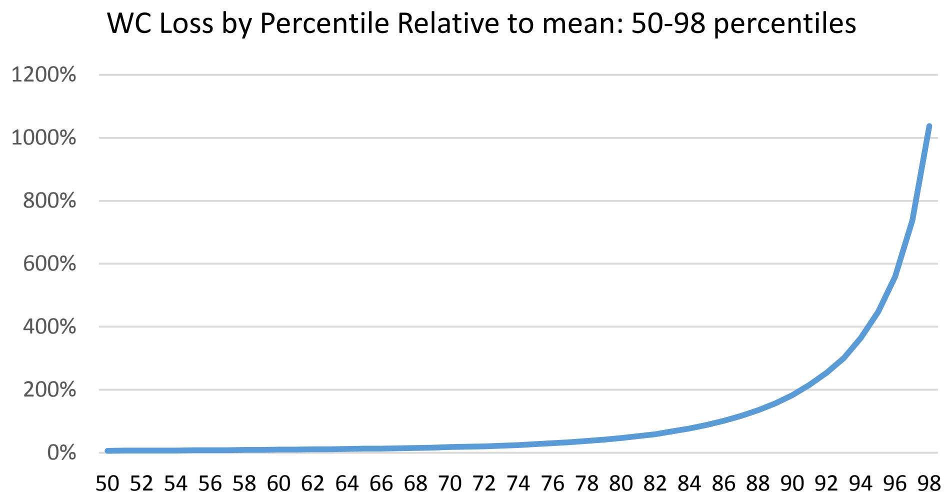
Agenda

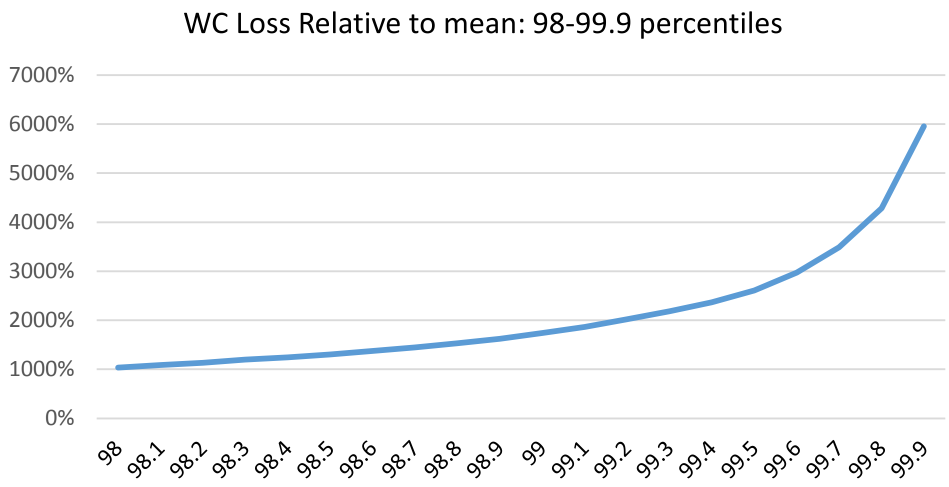
- Introduction
- Whole Book Distribution Analysis
- GLM, DGLM, FMM
- Regression Analysis Results
- Case Study - Claims Triage Models
- Case Study - High Deductible Pricing

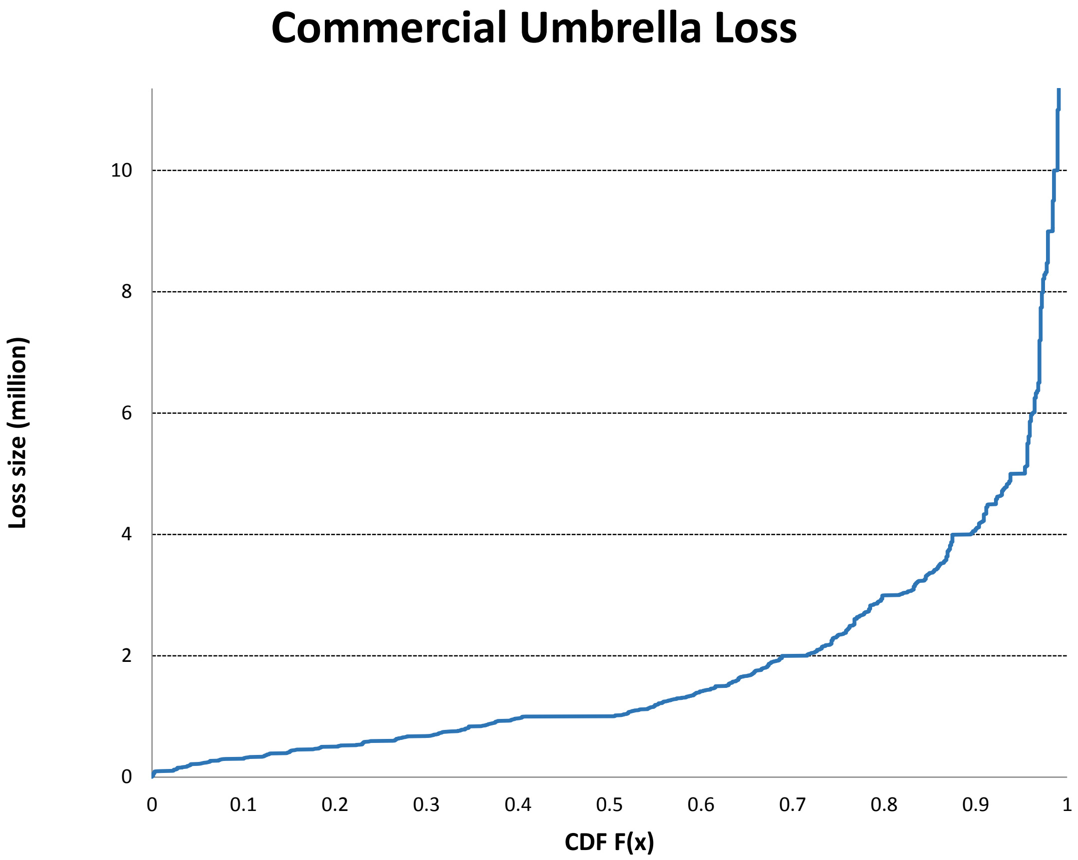
Introduction - Loss Distribution

- Conventional Large Loss Distribution Analysis
  a. Applications – ILFs, Deductible, XOL pricing, Reinsurance, ERM, etc.
  b. Single Distribution - Lognormal, Gamma, Pareto, Weibull, Inverse Gaussian, etc.
  c. Issues with the right tails or extreme events.
  d. Not considering the impact of covariates at individual risk level.

Loss Distribution: Fat-tail and Skewedness



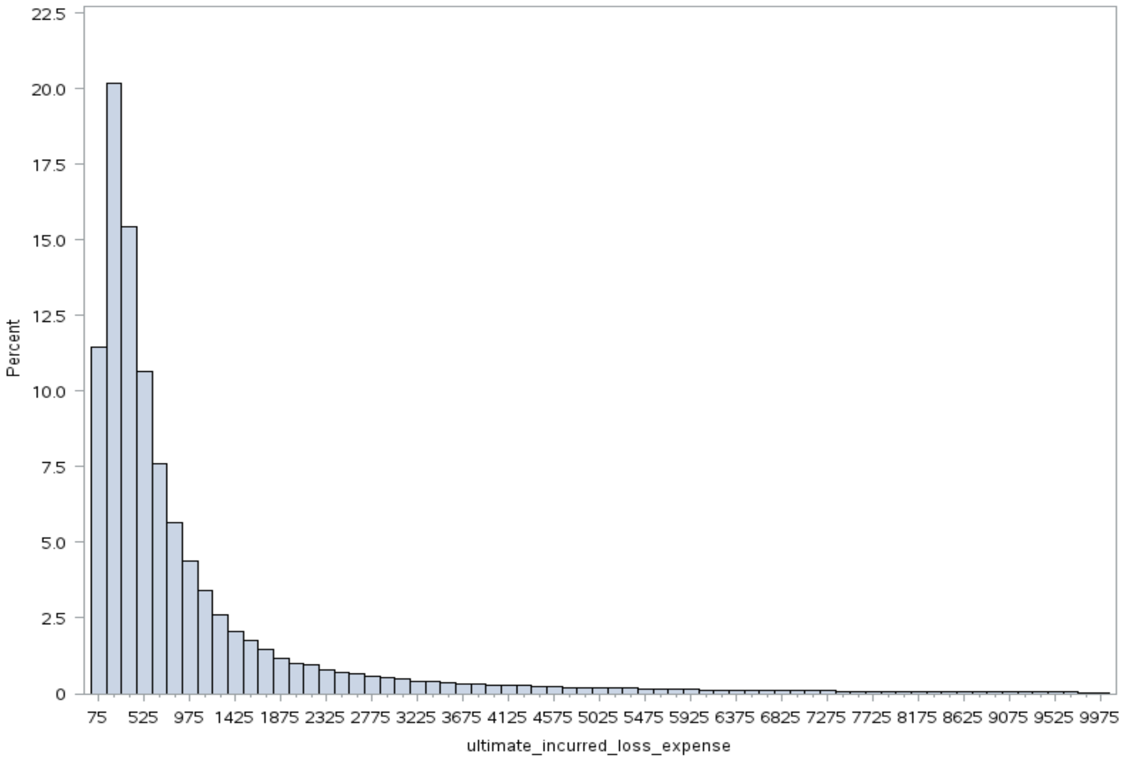




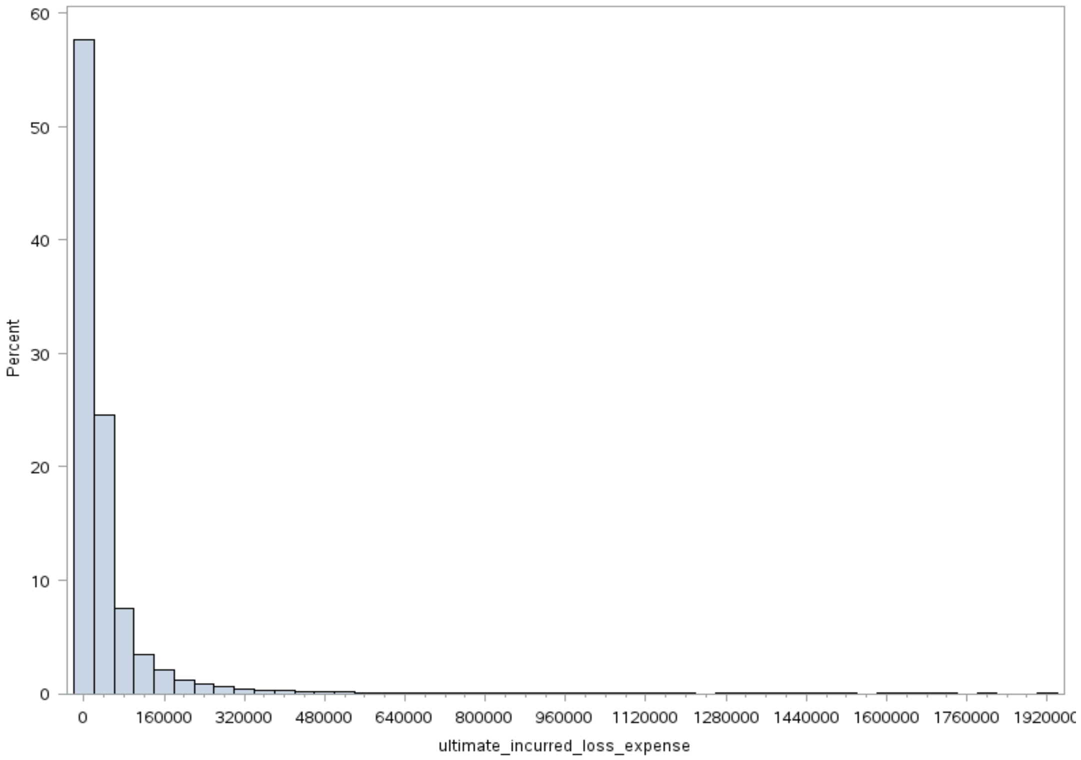
Loss Distribution: Heterogeneous and Mixed

- WC loss: small medical-only claims vs. large long-tail claims with indemnity

Medical only: Mean < 2k

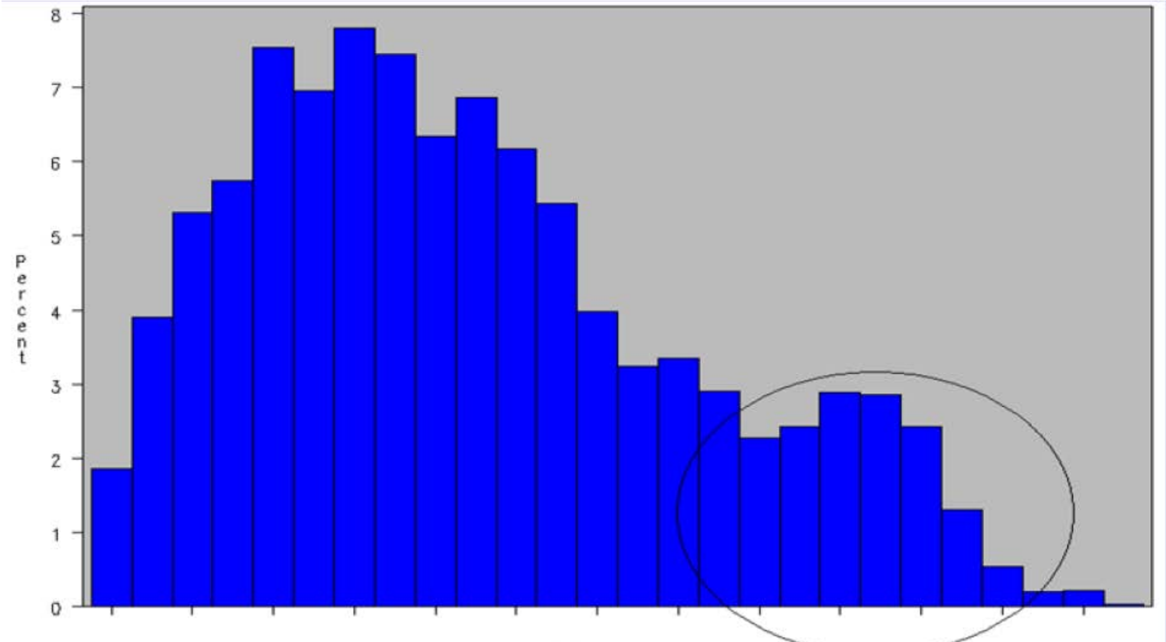


With Indemnity: mean >30K

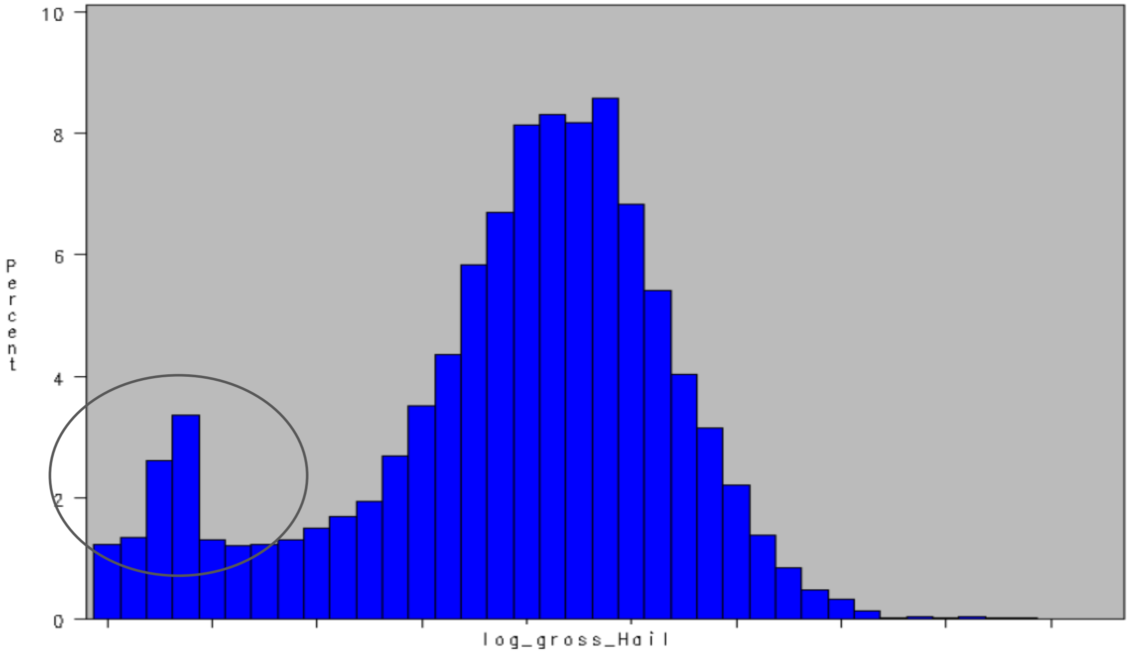


Loss Distribution: Heterogeneous and Mixed

- Home fire loss: a small percentage of total loss
- Home hail Loss: the smaller mode is on the left side?



Histogram plot of fire loss in log scale



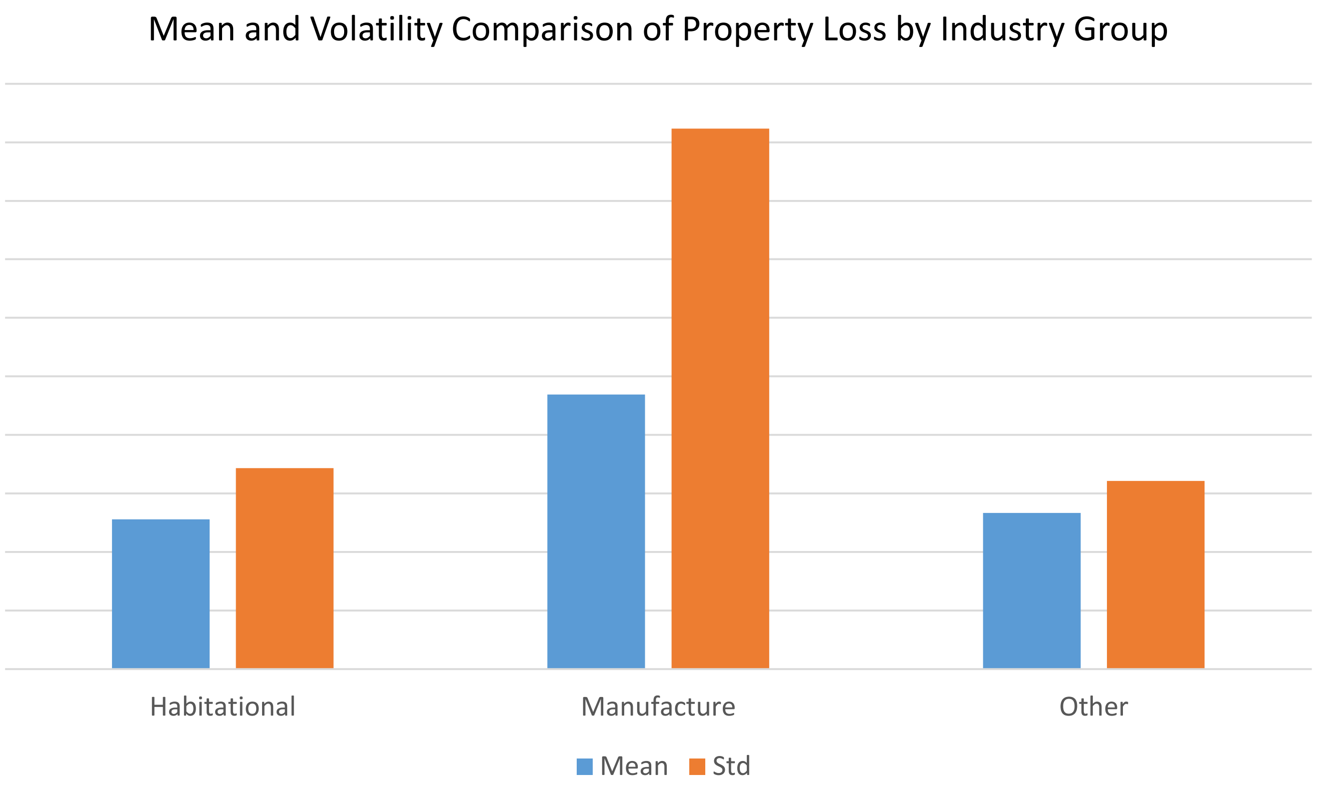
Histogram plot of hail loss in log scale

Introduction - Loss Regression

- Conventional Regression Analysis on Severity
  a. Take the specific information for each claim into consideration
  b. The whole book loss distribution analysis does not provide the insights of individual risk classifications
  c. Business Shift
  d. GLM and its various applications

Heterogeneous Underwriting Risks

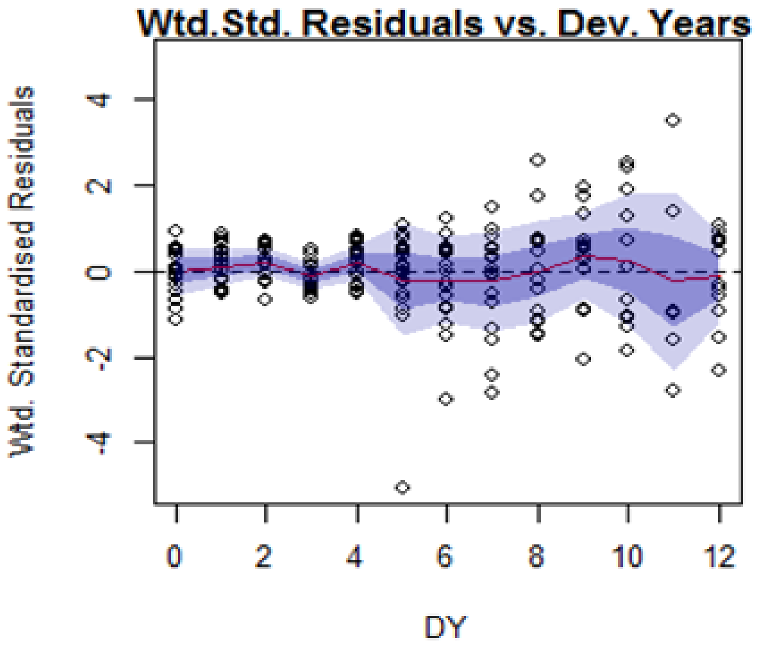
- Loss distributions are not homogenous

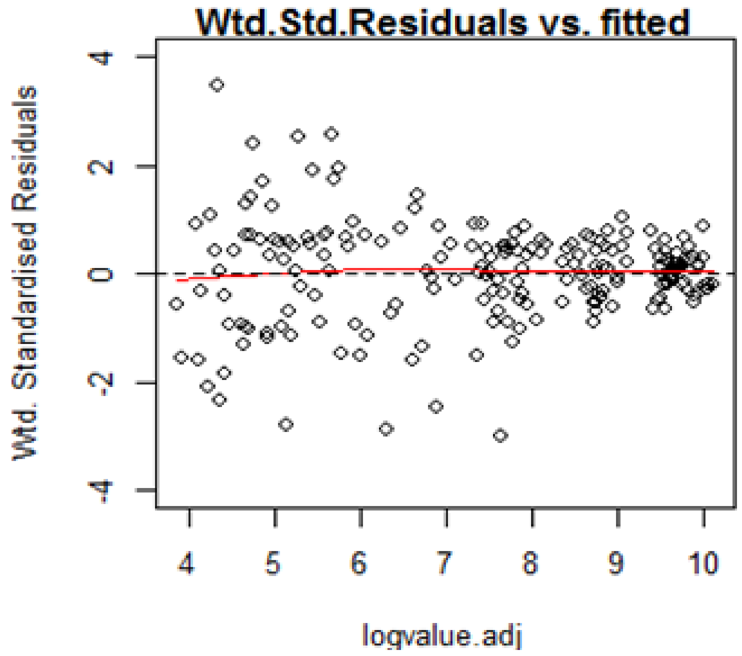


Heterogeneous Reserve Variability

- Loss distributions are not homogenous

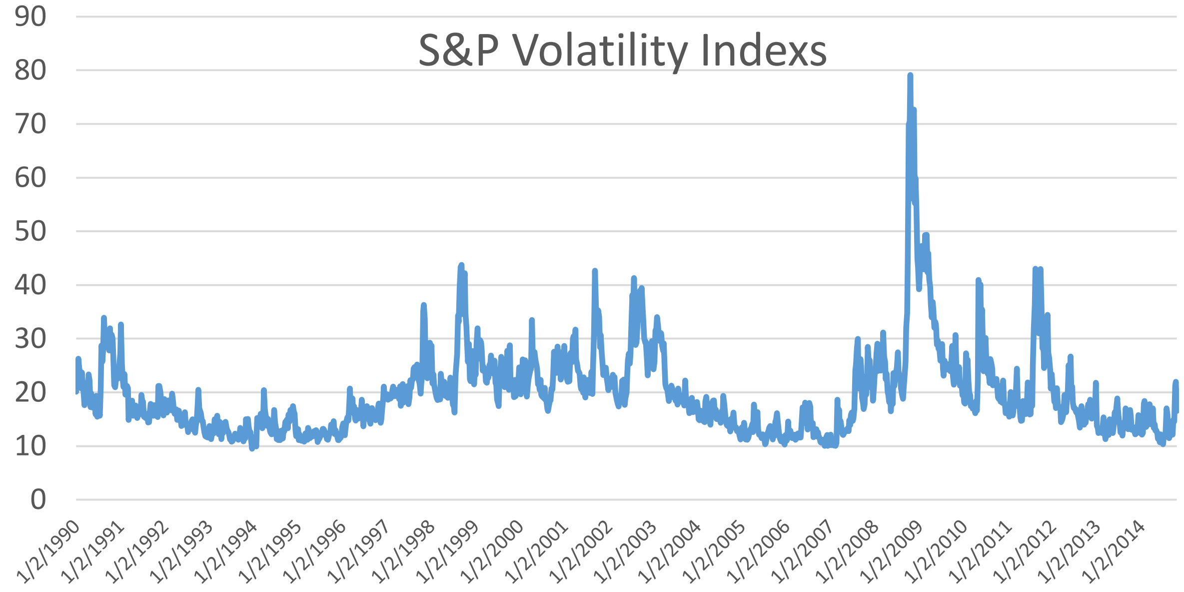
Umbrella Reserve Heteroskedasticity (log-linear model on incremental paid loss)

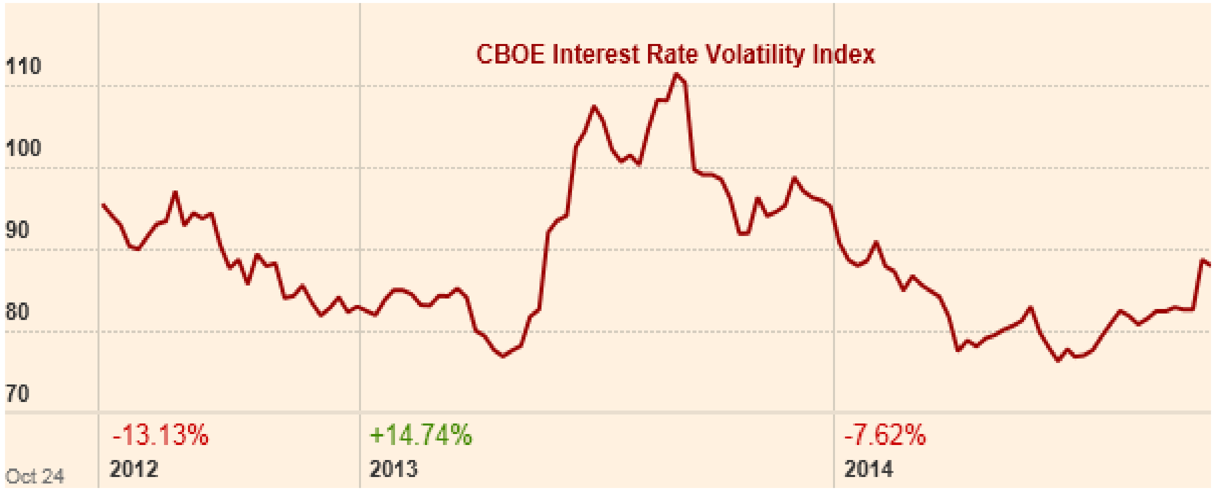




Heterogeneous Investment Risks

- Time-varying equity and interest rate risks





Introduction - Regression Analysis with Fat-Tail Distributions

- Combination of Regression and Fat-Tail Analysis
  a. Generalized Minimal Bias Method
  b. Generalized Linear Model with non-exponential family distribution
  c. Generalized Beta of 2^{nd} kind
  d. Copula Regression
- Issues
  a. No easy to use package
  b. Single distribution not good enough
  c. Observable Heterogeneity – Split the data
  d. Unobservable Heterogeneity

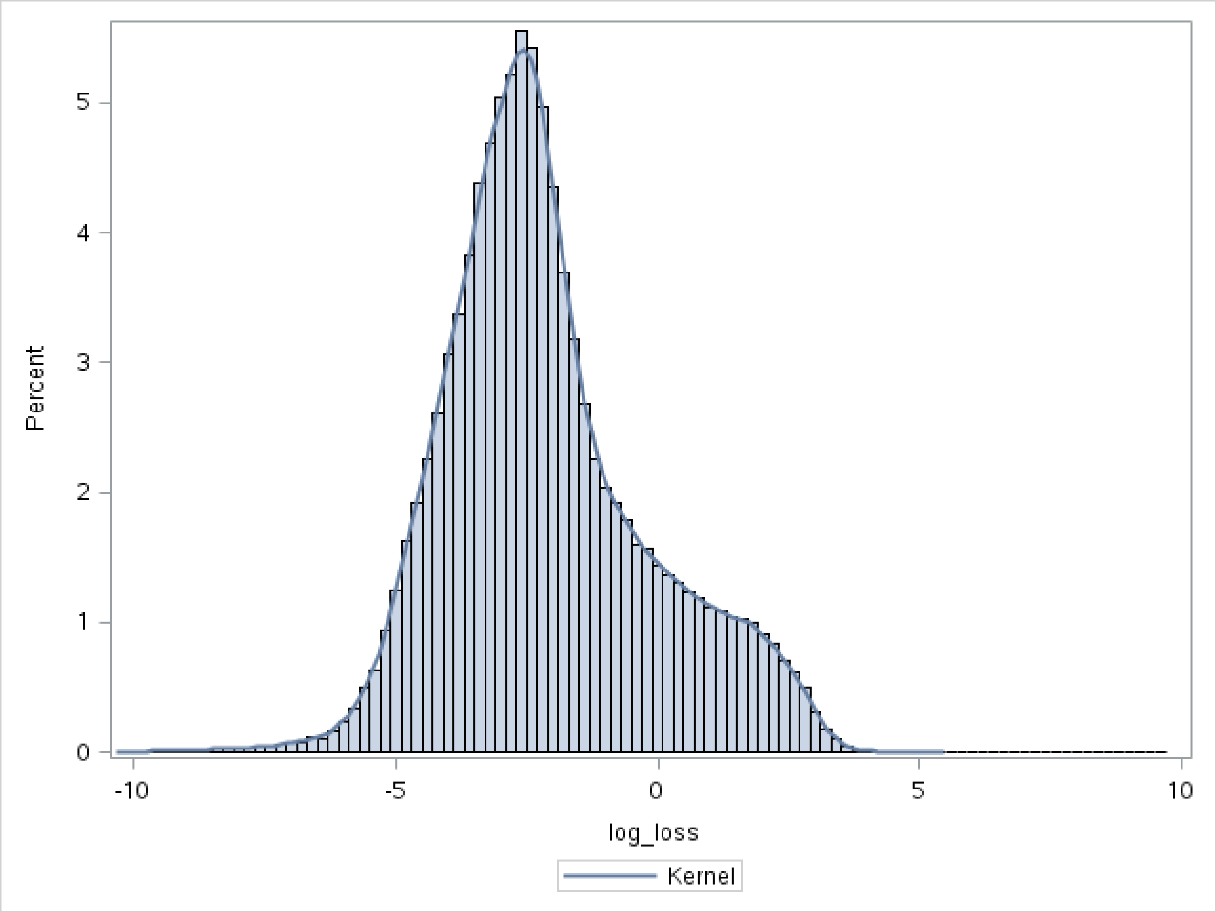
Introduction - Mixture Distribution and Regression

- Mixture Distribution and Regression
    a. Take the specific information for each claim into consideration
    b. Flexible
    c. Fat Tail
    d. Heteroskedasticity
    e. Over dispersion, under dispersion
    f. Observed heterogeneity, unobserved heterogeneity
- Parameter risks

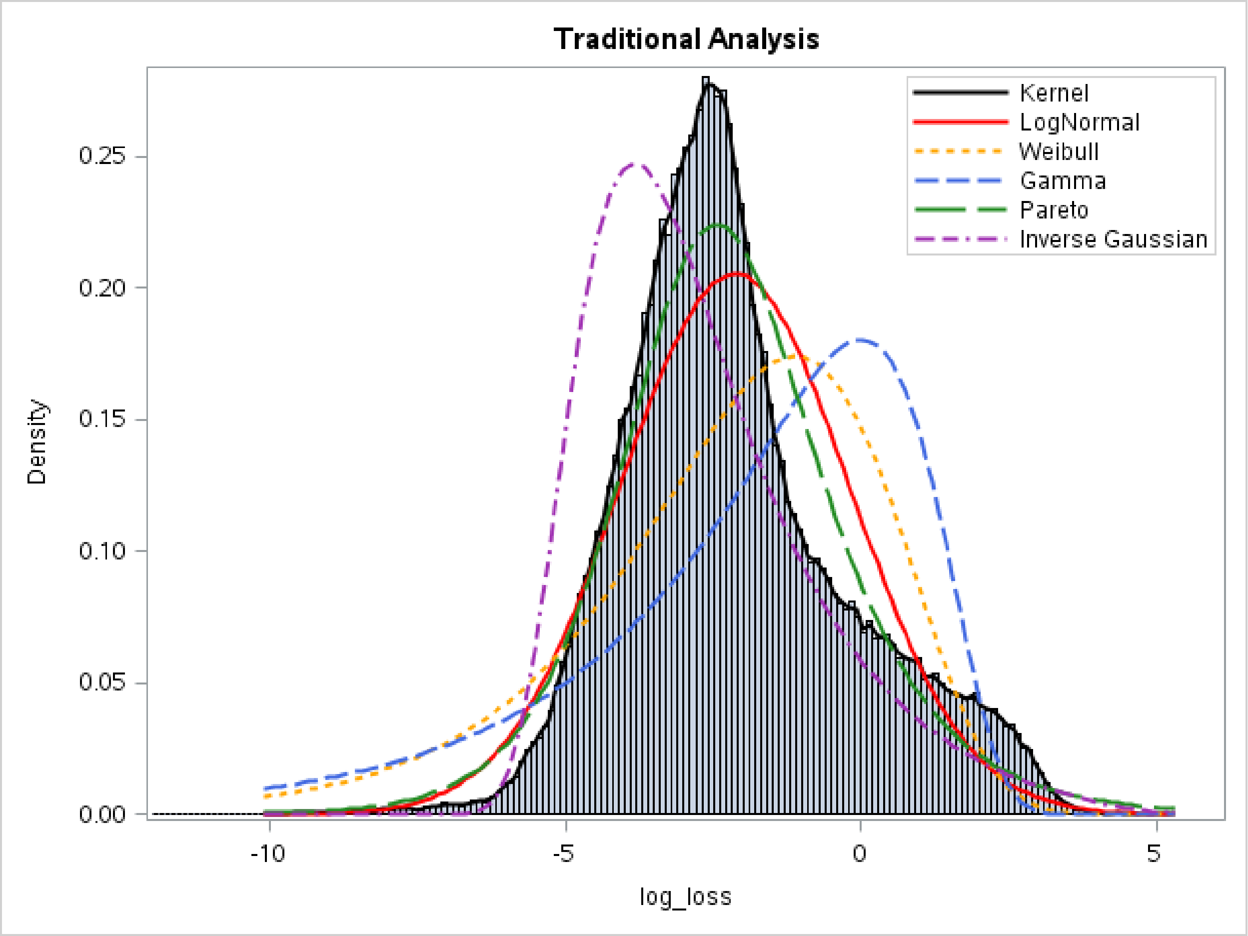
Data - Workers Compensation Claims

- Simulate claim final costs
- Simulate five variables known at the first notice of loss:
  - Two groups of injury codes: rank from low to severe
    - ✓ Both Injury Codes are from injury related information
    - ✓ Correlated, but not always point to the same direction
  - Age
  - Class Group
  - Fatal claim indicator
- Total 110,461 claims with mean set to be 1.

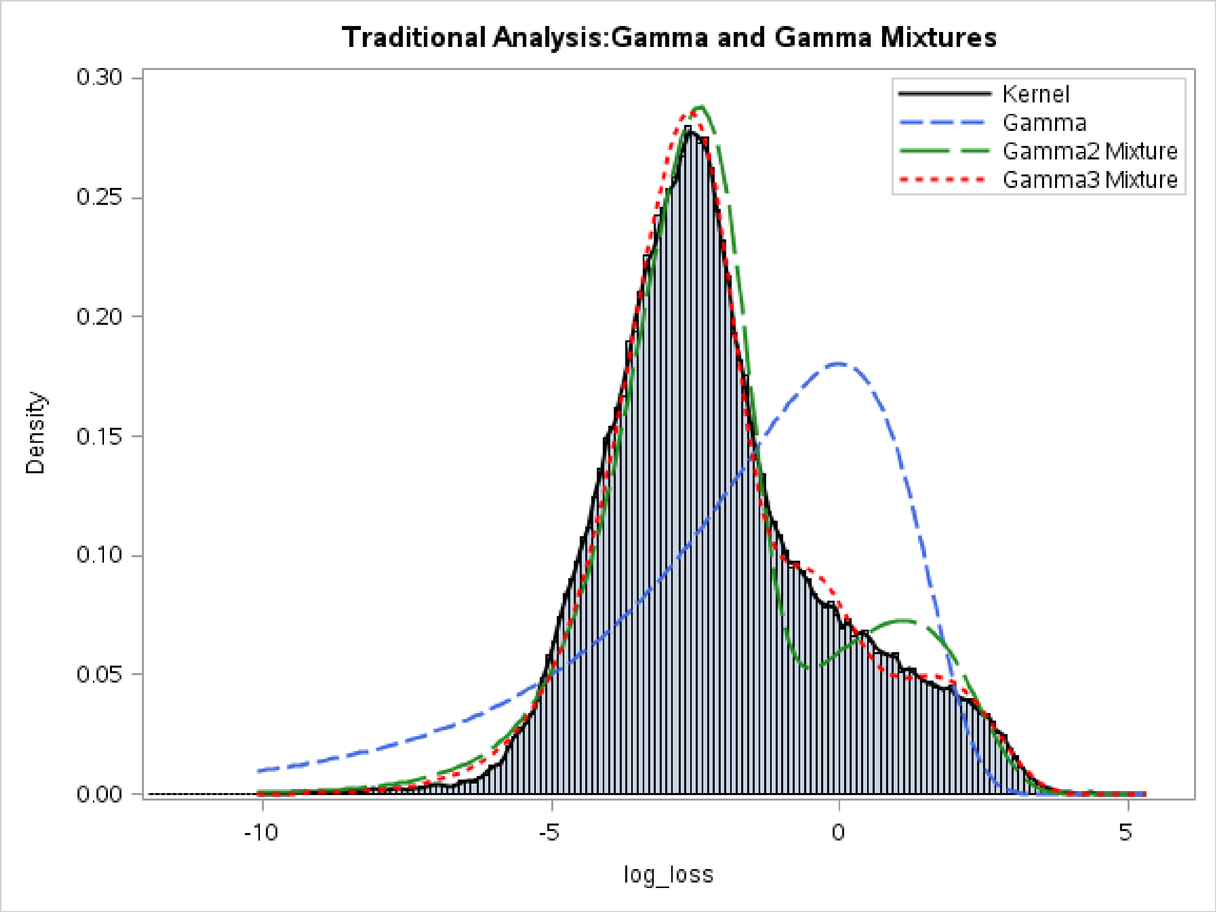
Histogram of the Simulated Claims



Traditional Whole Book Analysis



Traditional Whole Book Analysis - Mixture



Claim Counts in Ranges

Lower	Upper	Observed	% of Counts	Lognormal	Weibull	Gamma	Pareto	Inverse Gaussian	Gamma2 Mixture	Gamma3 Mixture
0	1	93,854	84.97%	95,456	90,222	79,674	96,283	99,574	91,373	94,027
1	5	10,490	94.46%	12,025	17,321	26,349	9,597	7,094	12,282	10,005
5	10	3,256	97.41%	1,744	2,205	3,814	1,794	1,603	4,362	3,548
10	25	2,541	99.71%	912	666	622	1,347	1,296	2,281	2,556
25	50	281	99.96%	221	44	3	567	531	157	315
50	∞	39	100.00%	103	2	0	873	362	2	11

Goodness of Fit: Chi-Square

Lower	Upper	Lognormal	Weibull	Gamma	Pareto	Inverse Gaussian	Gamma2 Mixture	Gamma3 Mixture
0	1	27	146	2,524	61	329	67	0
1	5	196	2,694	9,545	83	1,626	261	24
5	10	1,311	500	82	1,191	1,705	280	24
10	25	2,910	5,279	5,922	1,058	1,196	30	0
25	50	16	1,282	25,520	144	118	98	4
50	∞	40	600	1,444	797	288	685	71
χ^2		4,499	10,501	45,037	3,335	5,261	1,421	123

GLM

$$f(y_i;\theta_i,\phi,\omega_i) = \exp\left\{\frac{y_i\theta_i - b(\theta_i)}{\phi/\omega_i} + c(y_i,\phi/\omega_i)\right\}$$

$$E[y_i] \stackrel{\text{def}}{=} \mu_i = b'(\,\theta_i)$$

$$Var[y_i] = \frac{\phi}{\omega_i} b''(\theta_i) = \frac{\phi}{\omega_i} V(\mu_i)$$

$$g(\mu_i) = X_i^T \beta$$

- $V(\cdot)$ is the variance function
- ϕ is the dispersion parameter, constant
- g is the link function

DGLM

$$f(y_i;\theta_i,\phi_i,\omega_i) = \exp\left\{\frac{y_i\theta_i - b(\theta_i)}{\phi_i/\omega_i} + c(y_i,\phi_i/\omega_i)\right\}$$

$$E[y_i] \stackrel{\text{def}}{=} \mu_i = b'(\theta_i)$$

$$Var[y_i] = \frac{\phi_i}{\omega_i} b''(\theta_i) = \frac{\phi_i}{\omega_i} V(\mu_i)$$

$$g(\mu_i) = X_i^T \beta$$

- $V(\cdot)$ is the variance function
- ϕ_i is the dispersion parameter
- g is the link function

FMM

$$f(y) = \sum_{j=1}^{k} \pi_j(z, \alpha_j) p_j(y;\ x_j^T \beta_j, \phi_j)$$

$$\pi_j \ \geq 0, \text{for all j}$$

$$\textstyle\sum_{j=1}^{k} \pi_j(z, \alpha_j) = 1$$

The regression analysis of case study will be discussed in Spring meeting