Heterogeneous Risks and GLM Extensions

CAS Annual Meeting, New York, Nov. 2014

Luyang Fu, FCAS, Ph.D.

The author's affiliation with The Cincinnati Insurance Company is provided for identification purposes only and is not intended to convey or imply The Cincinnati Insurance Company's concurrence with or support for the positions, opinions or viewpoints expressed.

Agenda

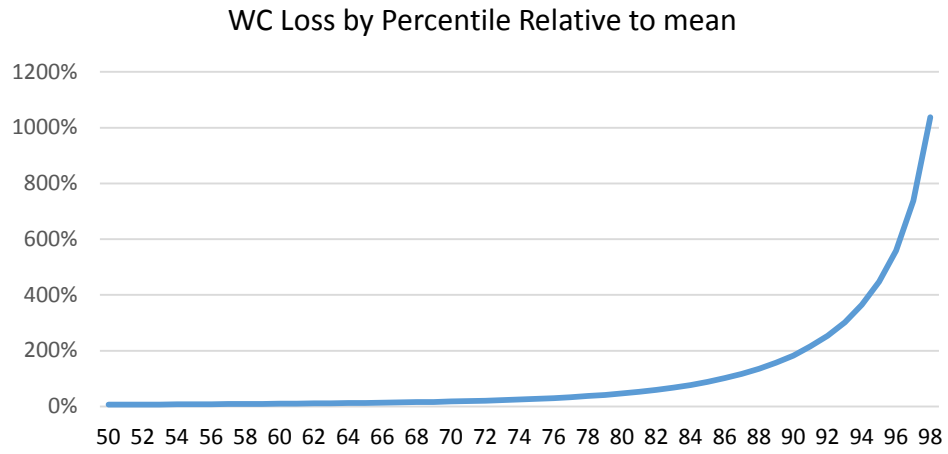
- Skewness and Fat-tail
- Heteroskedasticity
- Unobserved Heterogeneity
- Mixture Distribution
- GLM extensions
  - Generalized Linear Mixed Model
  - Double GLM
  - Finite Mixture Model

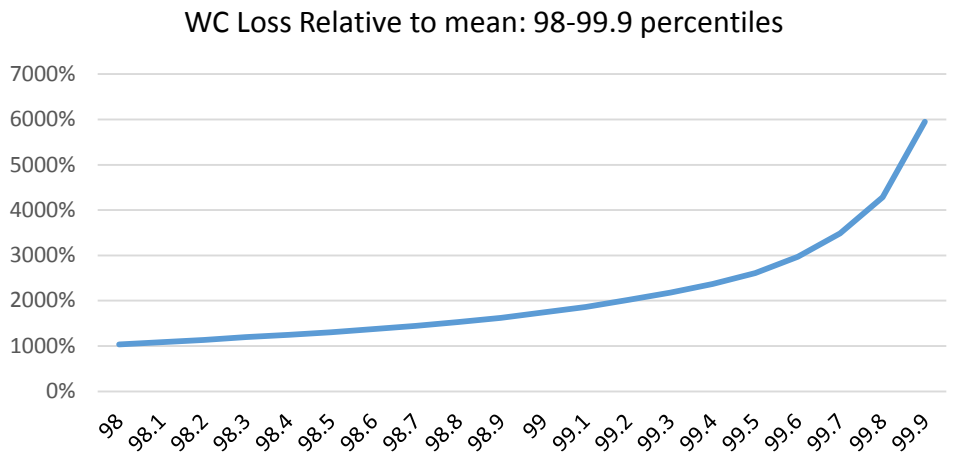
THE CINCINNATI INSURANCE COMPANIES

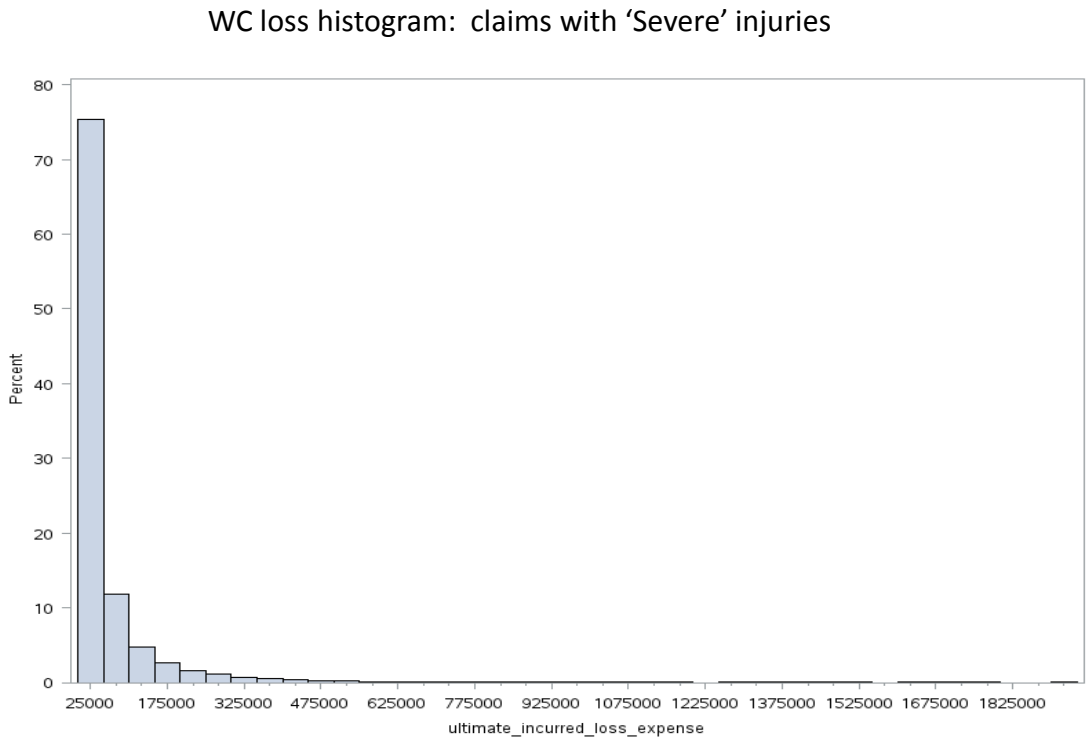
Yep, we are skewed!

Fleming G. K. (2008)

WC Loss: Skewness = 50.1; Median/mean=6.4%; Mean is at 86% percentile

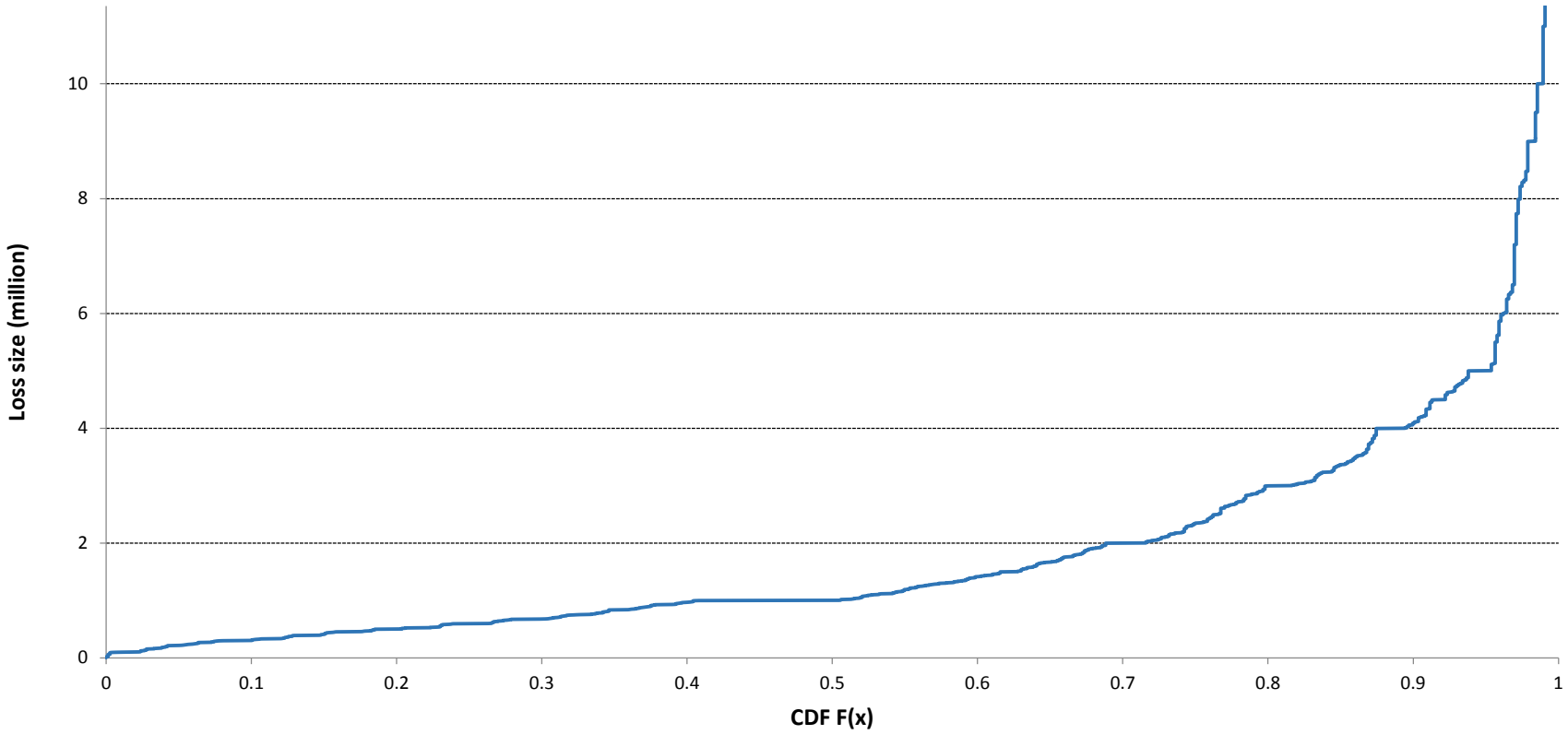






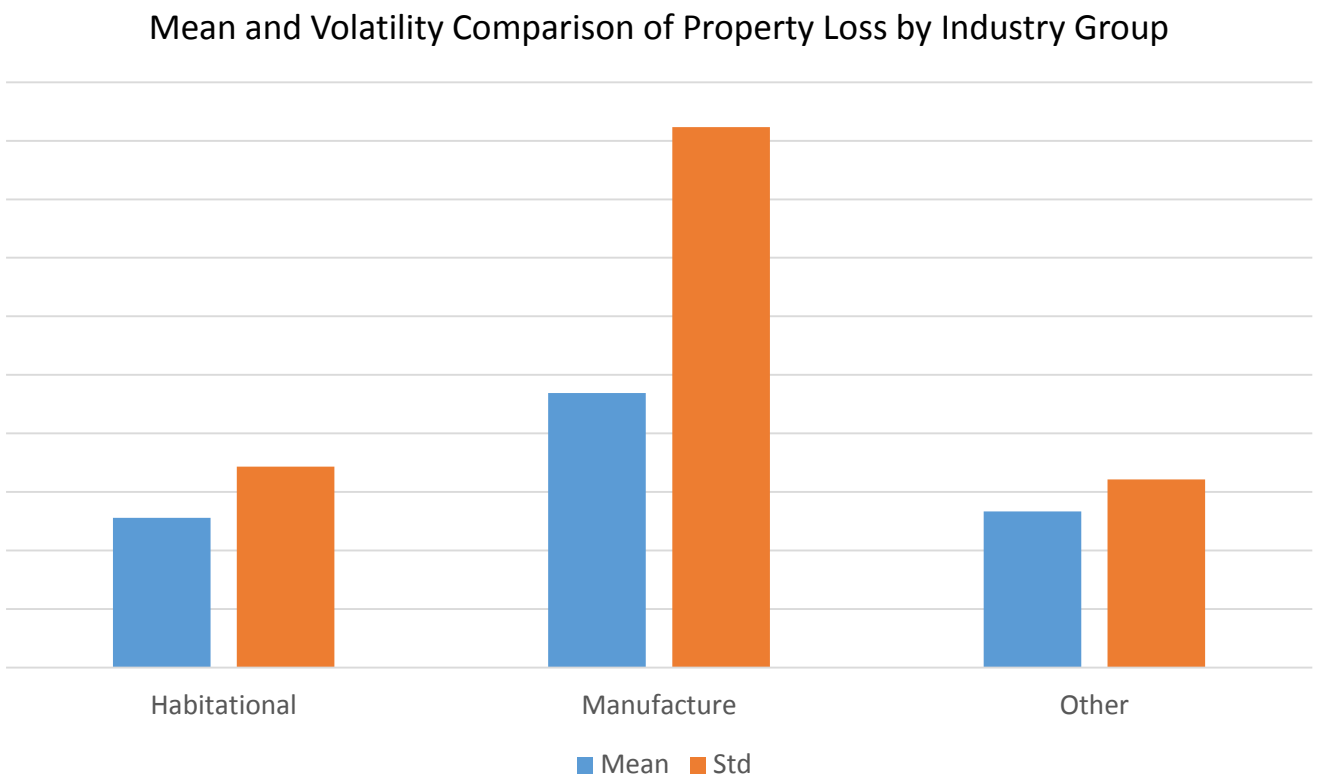
Yep, we are skewed!

Commercial Umbrella Loss



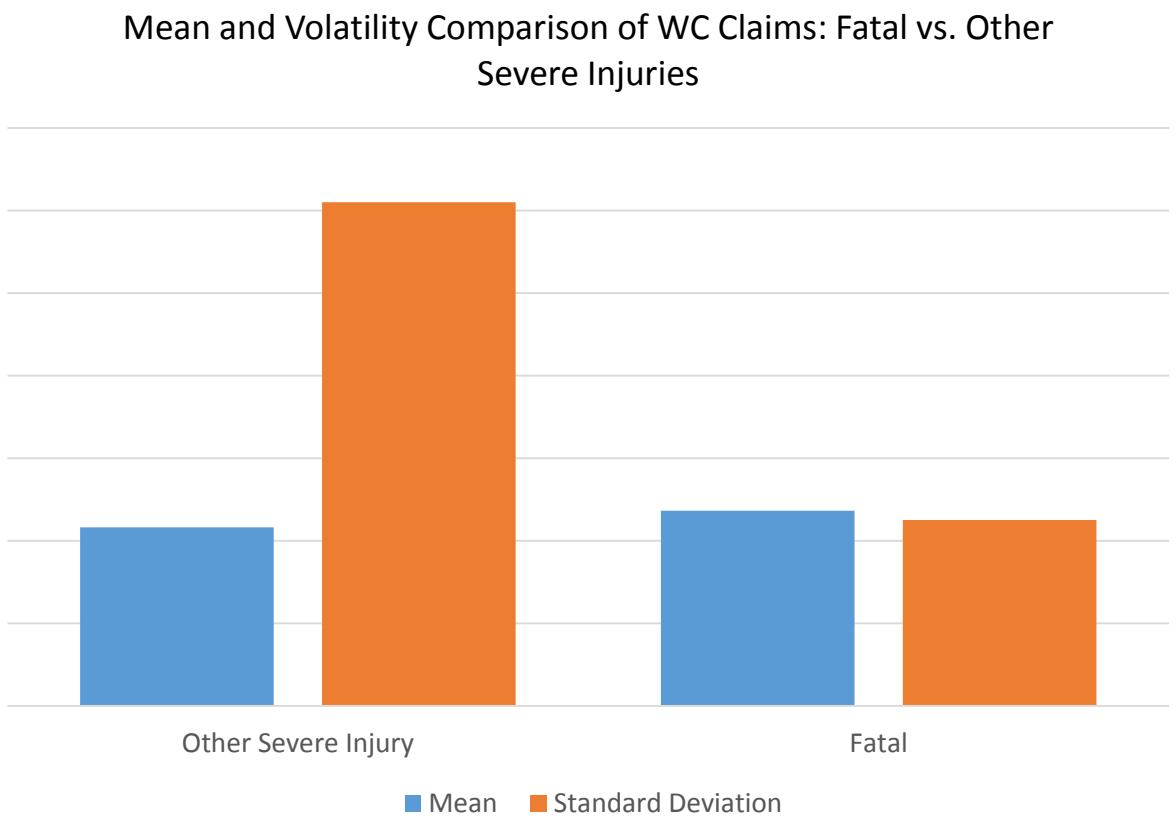
Heteroskedasticity

- GLM assumes homogenous variance
- Non-homogenous variance is a common insurance phenomena



Heteroskedasticity

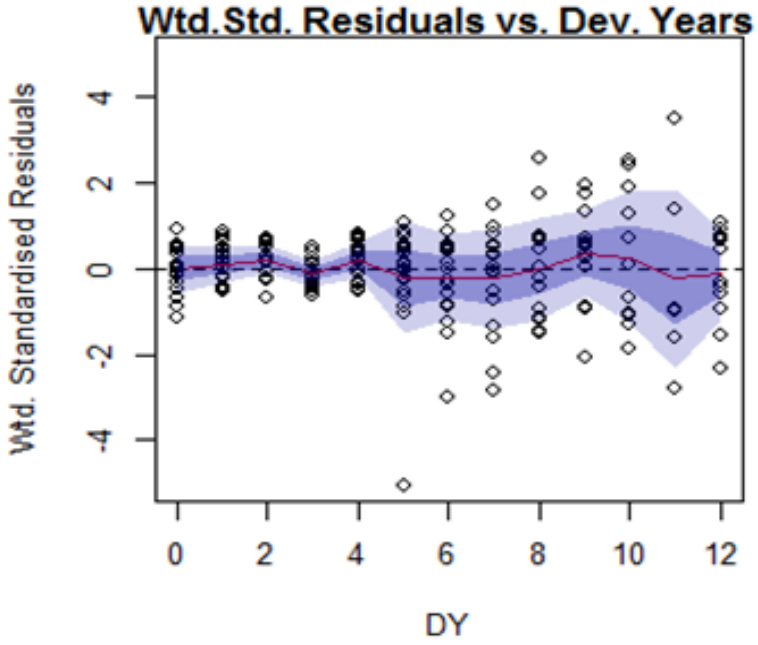
- Non-homogenous variance is a common insurance phenomena

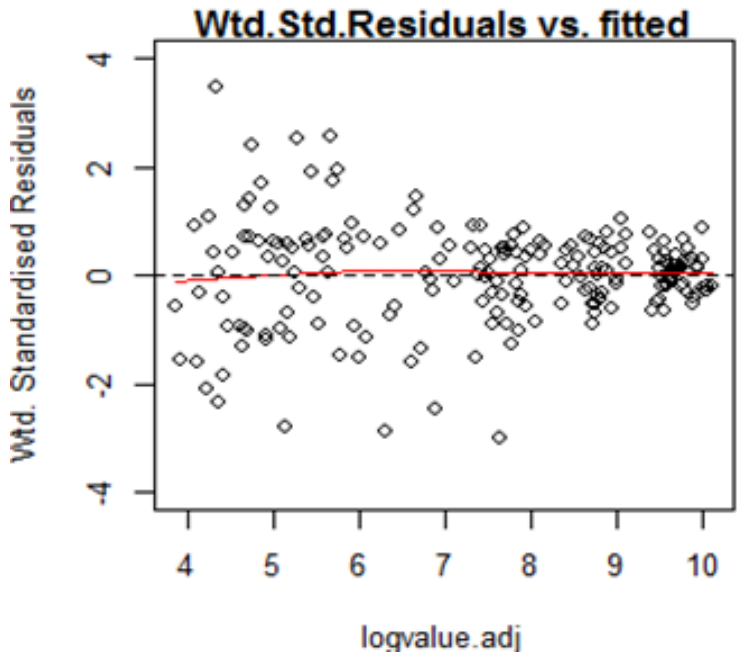


Heteroskedasticity

- Non-homogenous variance is a common insurance phenomena

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

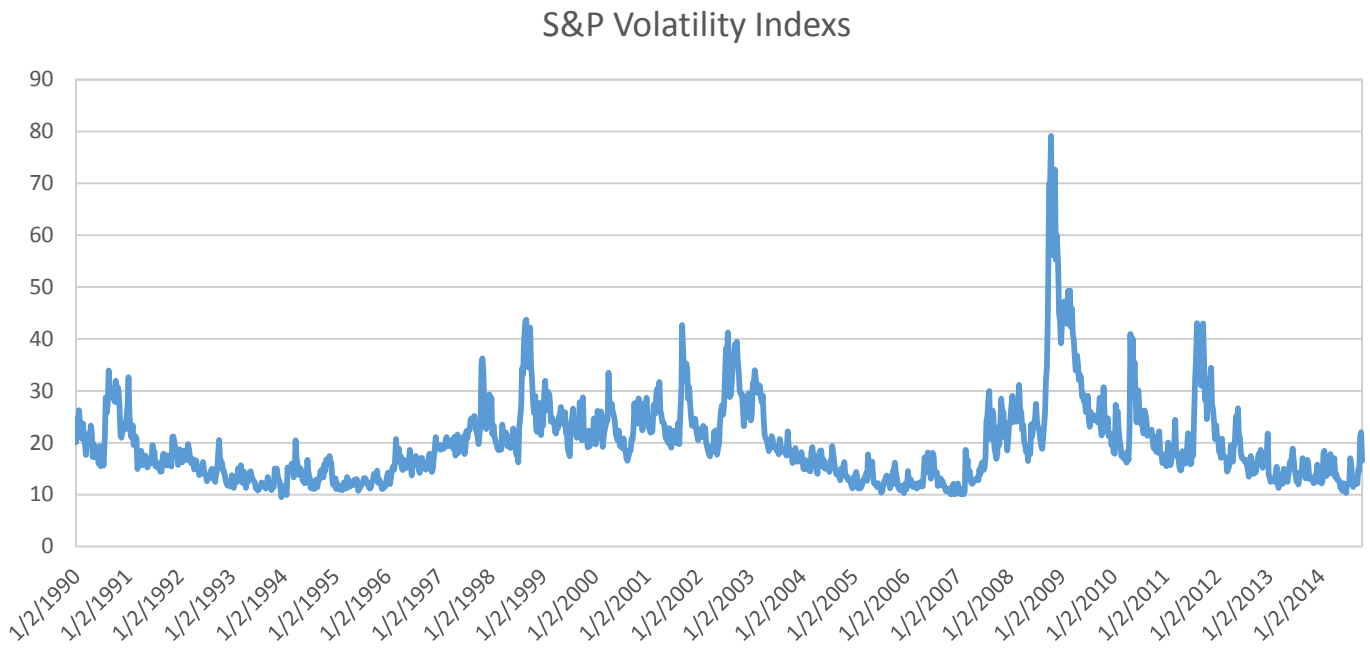




Heteroskedasticity

Non-homogenous variance is a common economic phenomena

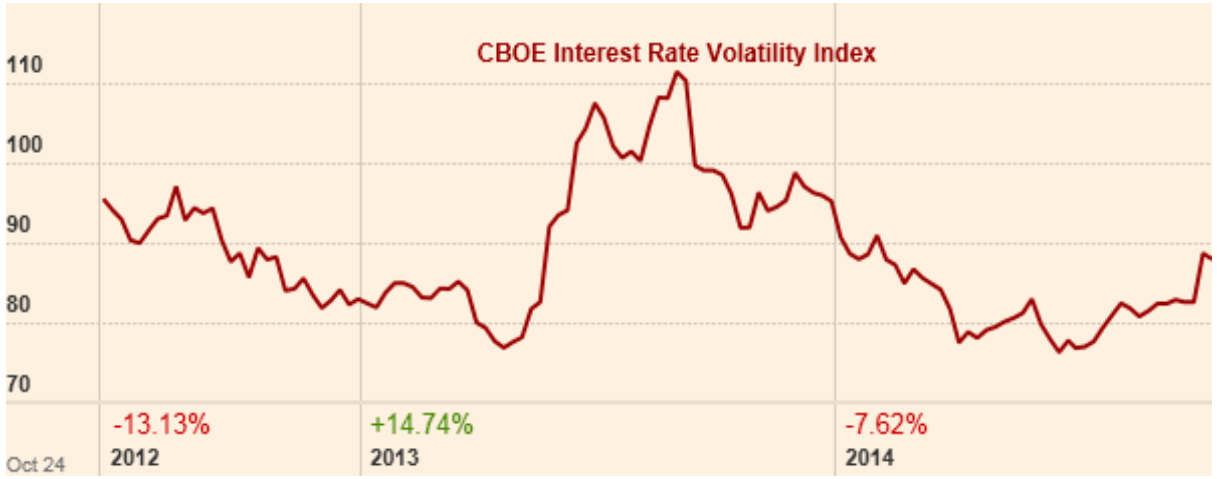
- Equity risk is not constant, but time-varying
- Autoregressive conditional heteroskedasticity (ARCH) and GARCH models treat variance as a time series.



Heteroskedasticity

Non-homogenous variance is a common economic phenomena

- Three-factor interest rate model: the third component is stochastic volatility



$$dr_t = (\theta_t - \alpha_t)\,dt + \sqrt{r_t}\,\sigma_t\,dW_t,$$

$$d\alpha_t = (\zeta_t - \alpha_t)\,dt + \sqrt{\alpha_t}\,\sigma_t\,dW_t,$$

$$d\sigma_t = (\beta_t - \sigma_t)\,dt + \sqrt{\sigma_t}\,\eta_t\,dW_t.$$

Unobserved Heterogeneity and Mixture Distribution

Many things are unobserved or unobservable

Auto pricing

- Frequent drinker vs. not
- Driving habit (careful drivers vs. not careful ones)
- Time of driving

Worker Comp claims at first notice of loss, little information on

- Health condition and comorbidity (with diabetes, obesity, etc., vs. not)
- Medical only vs. with indemnity
- Objective measure of injury severity (Johnson, Baldwin, and Bulter 1999)

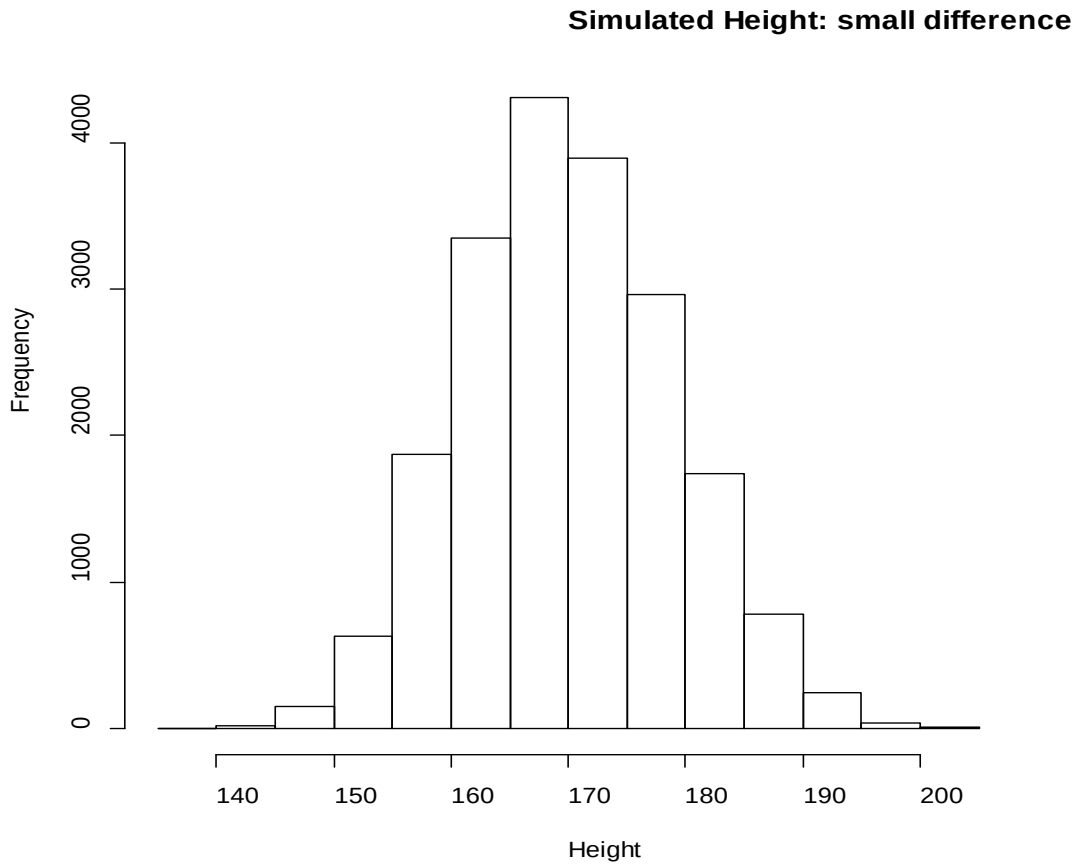
Unobserved Heterogeneity and Mixture Distribution

- If gender is unobserved, height follows a bi-modal distribution.
- When heterogeneity is weak, single distribution is OK

```
# Simulate man's Height
XM<-rnorm(10000, 175, 8)

# Simulate Women's height
XF<-rnorm(10000, 165, 7)

Height<-c(XM, XF)
hist(Height )
```

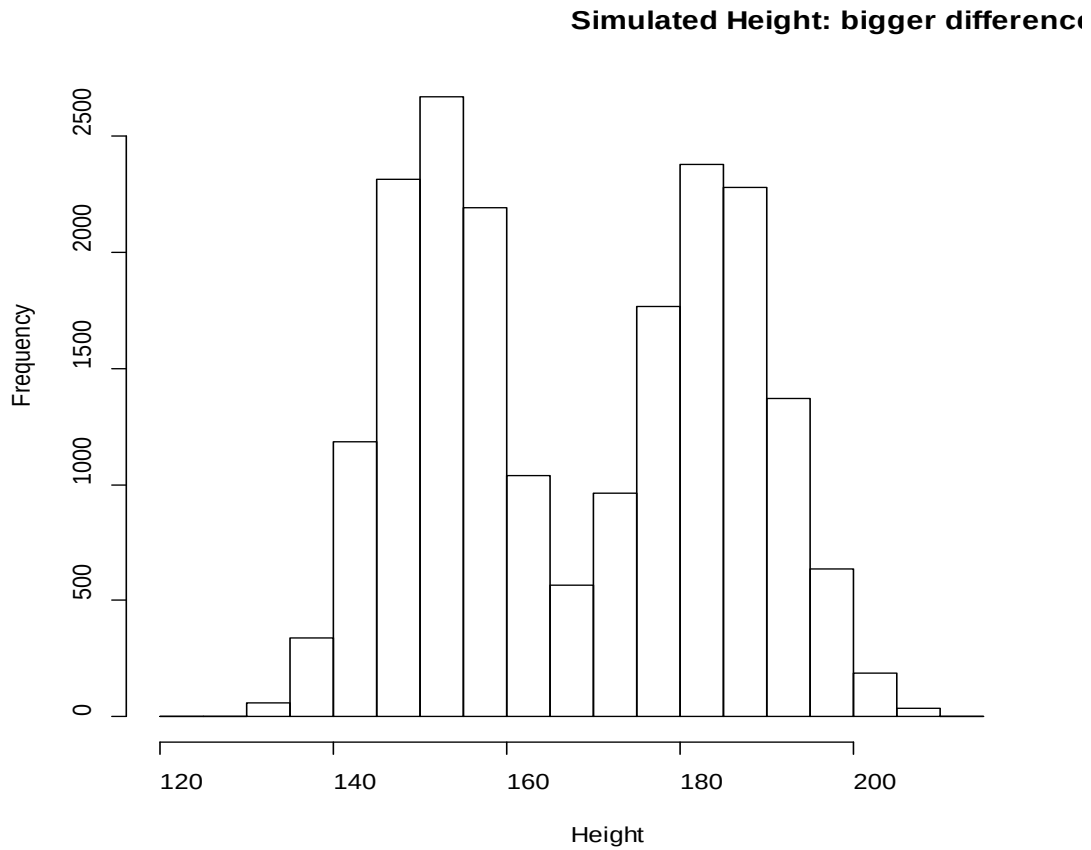


Unobserved Heterogeneity and Mixture Distribution

- If gender is unobserved, height follows a bi-modal distribution.
- When heterogeneity is strong, mixture distribution fits the data much better

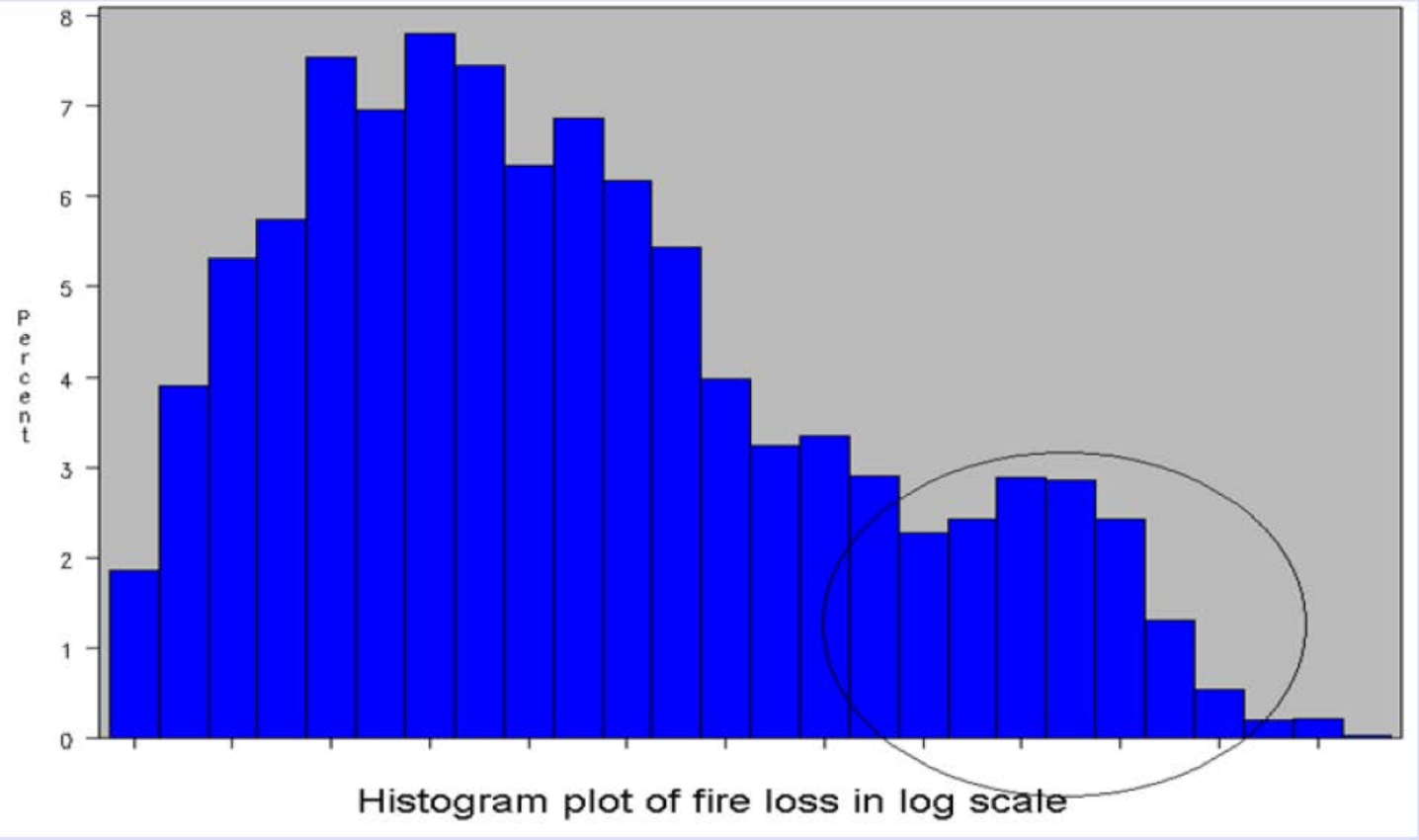
```
# Simulate man's Height (Netherland)
XM<-rnorm(10000, 184, 8)

# Simulate woman's height (Vietnam)
XF<-rnorm(10000, 152, 7)
Height<-c(XM, XF)
hist(Height )
```



Unobserved Heterogeneity and Mixture Distribution

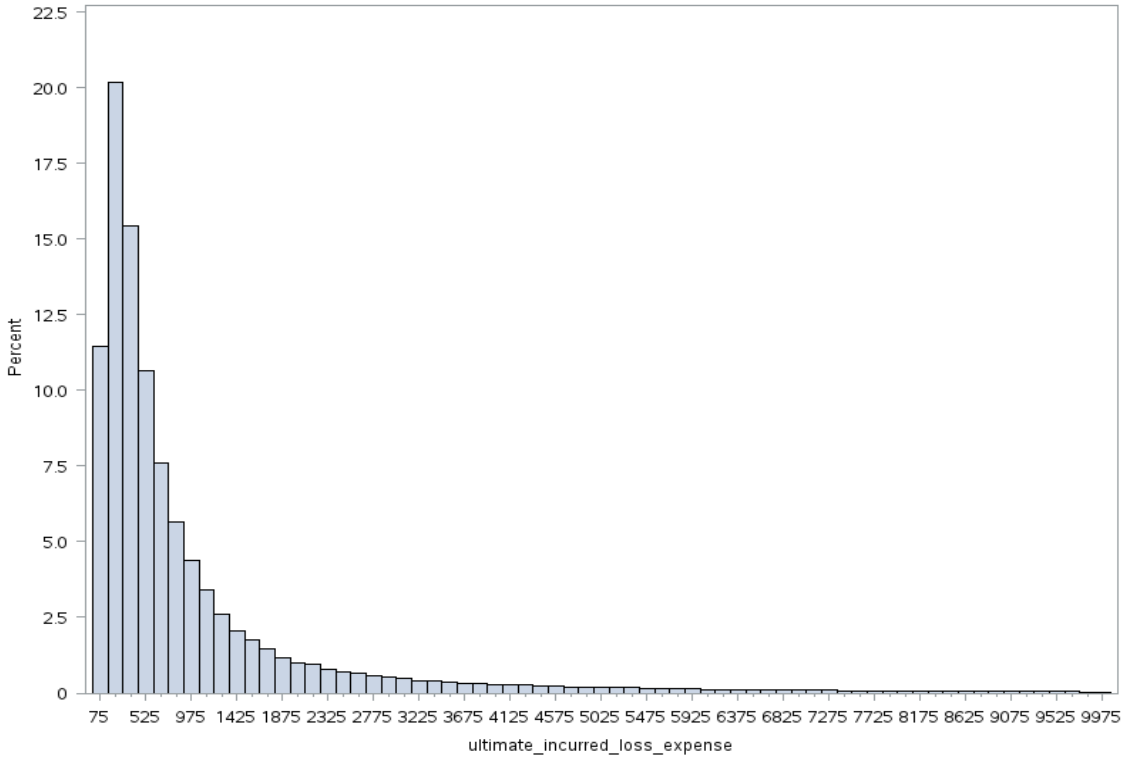
- Heterogeneity in P&C Insurance is strong
- Homeowner fire loss: partial loss + a small percentage of total loss



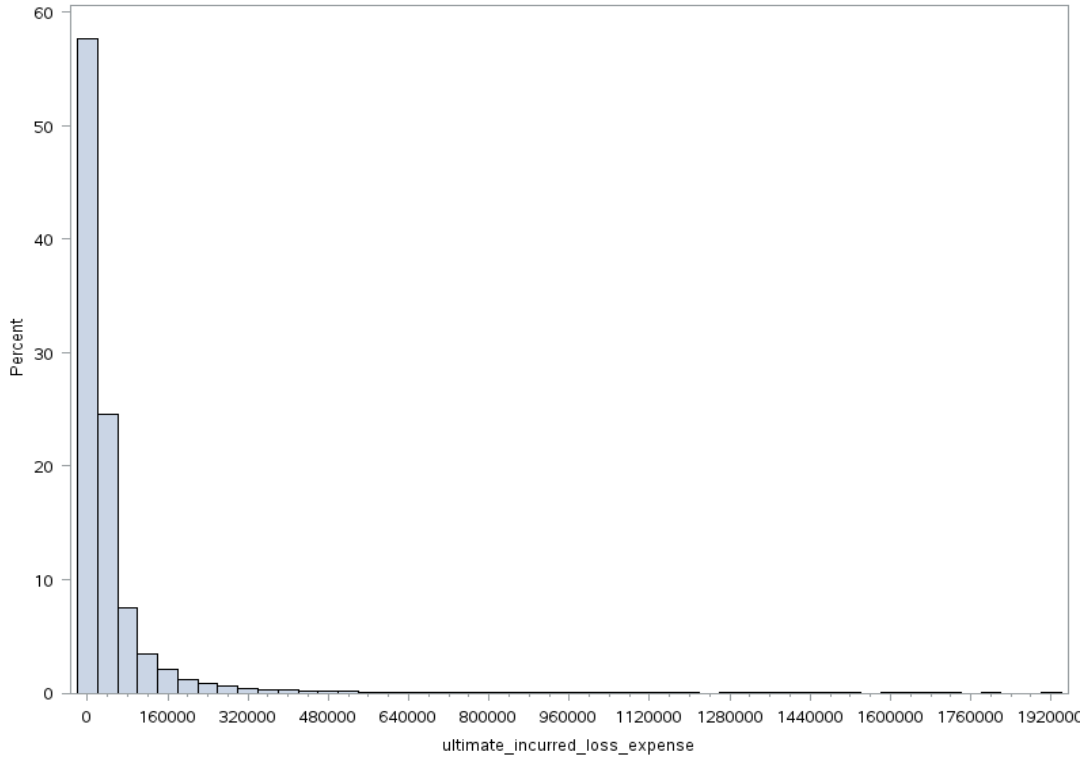
Unobserved Heterogeneity and Mixture Distribution

- Heterogeneity in P&C Insurance is strong
- When pricing WC, it is unknown that the future claims will be medical only or with indemnity

Medical only: Mean < 2k



With Indemnity: mean >30K



Unobserved Heterogeneity and Mixture Distribution

Arellano M. (2003), *Panel Data Econometrics*, Chapter 2, *Unobserved heterogeneity*

"*Statistical inferences may be erroneous if, in addition to the observed variables under study, there exist other relevant variables that are unobserved, but correlated with the observed variables*"

Unobserved Heterogeneity and Mixture Distribution

Assume we are studying the impact of diet and excising on weight; if gender is missing, the result can be very biased

```
# man's calories
ManCal<-rnorm(10000, 3000, 1000)
#average exercise hours
ManExe<- rnorm(10000, 1, 0.3)
# random term
ManNoise<-rnorm(10000, 0, 30)
# Man's weight
ManWeight<-180+0.02*(ManCal-3000)-20*(ManExe - 1)+ManNoise;
```

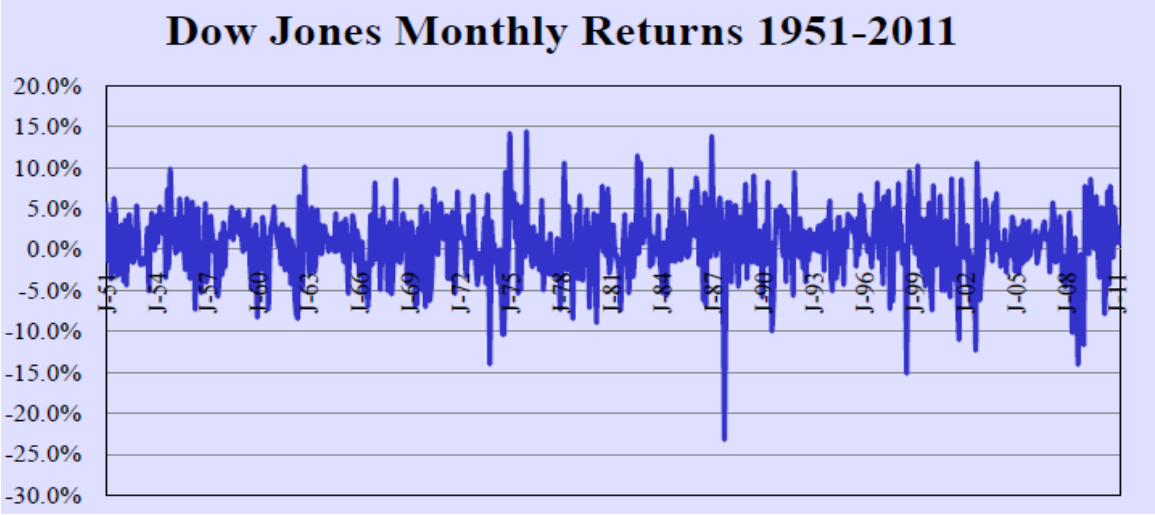
```
# women's calories
WomanCal<-rnorm(10000, 2300, 800)
#average exercise hours
WomanExe<-rnorm(10000, 0.8, 0.25)
# random term
WomanNoise<-rnorm(10000, 0, 25)
# WoMan's weight
WomanWeight<-130+0.02*(WomanCal-2300)-20*(WomanExe - 0.8)+WomanNoise;
```

Coefficients:	Estimate	Std. Error	t value
Intercept)	83.53	0.93	90.29
Calories	**0.03**	0.00	111.73
Exercise	**-0.16**	0.79	-0.20

Unobserved Heterogeneity and Mixture Distribution

Stock Return:

- Assuming normal distribution, the likelihood of monthly loss over 14.1% is 0.02%; actual observation is 0.55% (27 times than the single normal assumption)
- Mixture model: regime switching Hamilton (1990), D'Arcy and Govett (2004)
  - Investment return follows two distributions with low and high volatility



	Low Volatility	High Volatility
Mean	0.96%	-2.20%
Standard Deviation	3.59%	7.17%
Probability of Switching	3.37%	30.87%

GLM Extension: Case Studies

Case studies on P&C insurance will be presented in the meeting