



Data analysis

Indra Loljeeh

02-03/02/2022



Pricing steps

Pricing steps:

- ✓ Analysis of the reinsured portfolio
- ✓ Statistical data set revalorisation
- ✓ Selection of claims
- ✓ Estimation of the Severity distributions
- ✓ Estimation of Frequency distributions
- ✓ Computation of the pure premium
- ✓ Calculation of special clauses



Data composition

- One dataset claims amount (Last know situation = Ultimate)
- For each claim, the occurrence date, the sum insured
- For each year, the corresponding premium income
- The expected management expenses of your company + direct brokerage for insurers
- The request for quotation



Statistical revaluation

- Objective : set data in quotation year economic situation
- Revaluation:
 - Claims inflation
 - Superimposed claims inflation
 - Tariff correction (exposition measure!)



Exposure measure

Problematic related to change in exposure

- ✓ Representativity of past data
 - ✓ Change in Underwriting policy
 - ✓ Change in legislation
 - ✓ Inflation
 - ✓ Merge & Acquisition
- Correction of the past data in function of new exposure



Premium indexation

- Objective : set premium on a same tariff basis
- Ideally, based on average premium evolution (of the ceding company – or market value)
- Alternatively, it can be based on published indices if relevant (construction prices, consumer prices,...)
- All premiums are indexed to the quotation year (2022)



Premium indexation: Example

Year	Pr.	Avg. P	Index
	Income		
2012	10.000	48	100
2013	10.200	51	106
2014	10.500	51	106
2015	10.600	48	100
2016	10.800	47	98
2017	11.200	45	94



Claims indexation

- Total claims inflation =
inflation (construction prices, consumer prices, wages)
+ superimposed inflation (if relevant)
- Depends on the country
- Estimation based on:
 - Data themselves
 - External sources (INS,...)
 - Forecasts

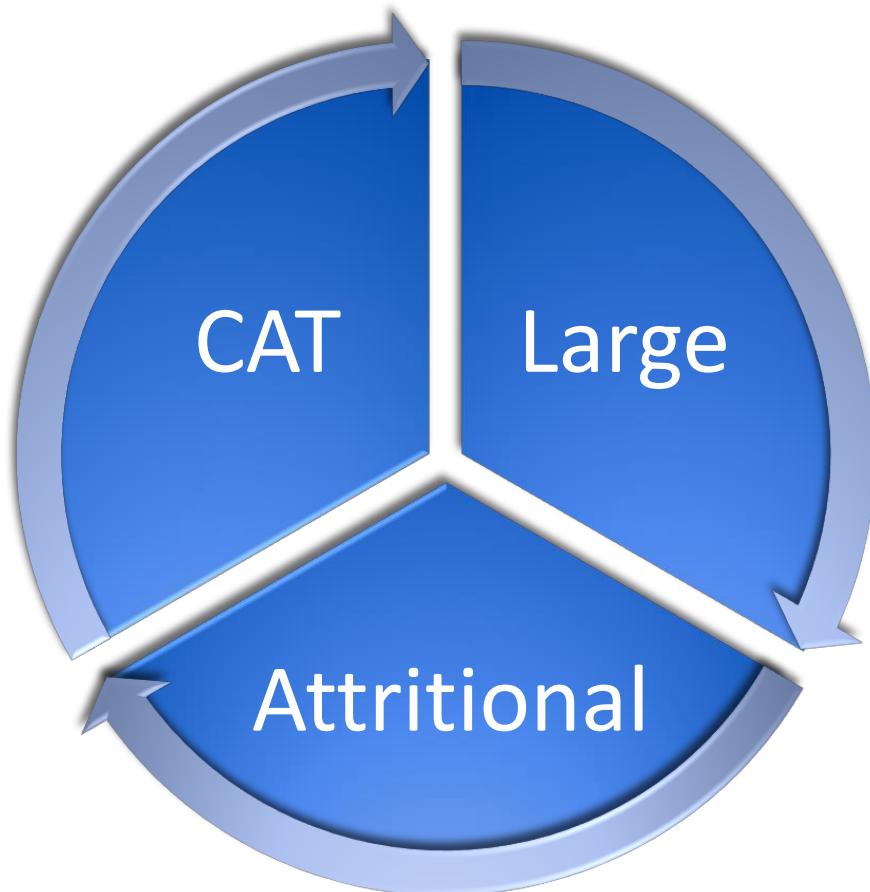


UW Cycle

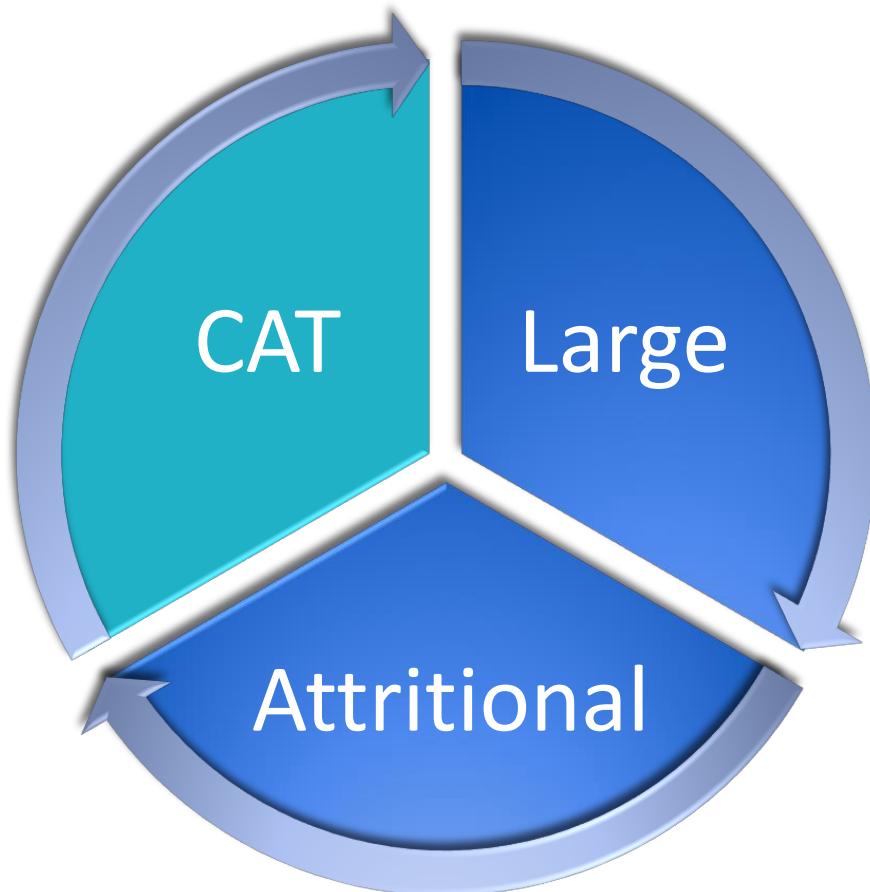
Be aware that not all parameters are taken into account:

- Expenses
- Renewal rate
- New business premium
- Exposition
- Trends
- Strategy
- etc

CLAims split



CLClaims split





CAT losses

Very extreme events: very low frequency/very high severity process

CAT events for all LoB?

Examples:

1. Property: Windstorm, Earthquake, Flood,...
2. Marine: Windstorm, Tsunami, Hurricane,...
3. Life & Accident: Terrorism, Pandemic, CATNAT?
4. Casualty: Terrorism?, Crashes?, Other?
5. ...



CAT losses models

Very rare events imply that there are few data available on the market

Solution:

- Softwares: RMS, Equecat,...
- Mathematical models (reliability?)
- Scenarios

Results: Event Loss Table

- No estimation required
- Only simulation process

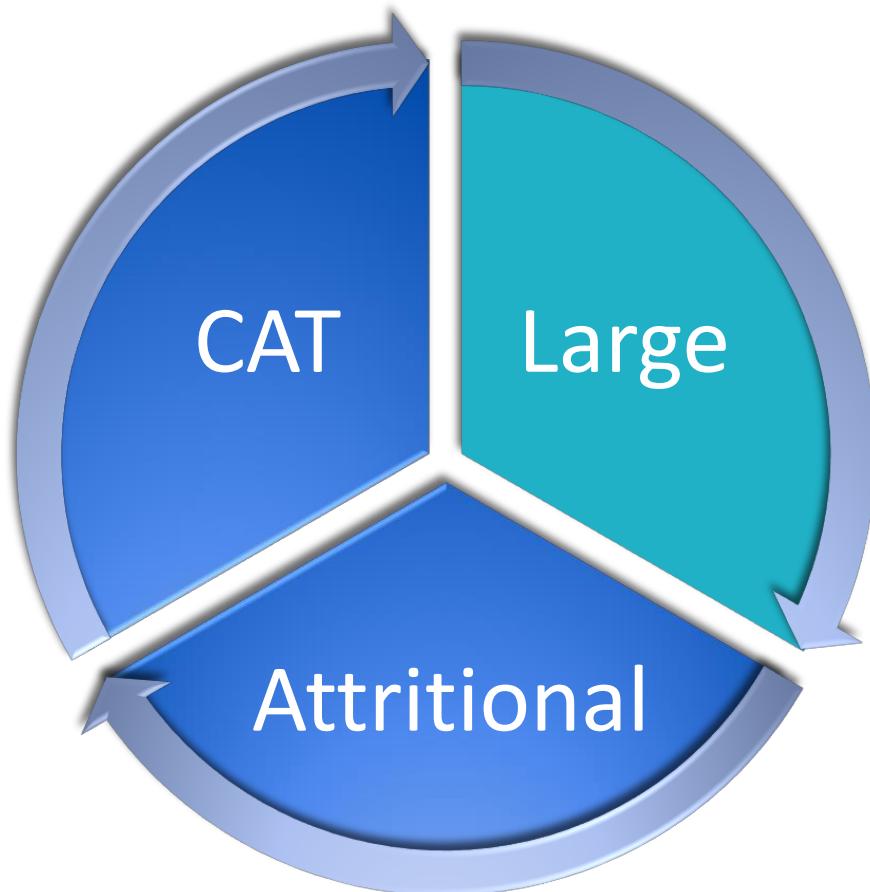
Clustering



DO NOT FORGET

- Identify CAT events in your data (hint: occurrence date)
- Remember: CAT claim in property is often a large number of small claims. If you find a large loss, will you include it in the CAT? This should be documented!!
- Compare them to the modeling outputs
- Pay attention to the simulation process

CLA split





Large losses

Extreme value theory:

Describe the unusual rather than the usual

- What is a large loss?



Poisson process

$$S^{Large}(t) = \sum_{i=1}^{N(t)} X_i$$

Where $\{N(t): t \geq 0\}$ is Poisson process with rate λ and
 $\{X_i: i \geq 1\}$ i.i.d. $\sim Pareto(A, \alpha)$ independent of $\{N(t): t \geq 0\}$

Frequency-Severity approach:

- Frequency: Poisson $P(N(t) = k) = e^{-\lambda} \frac{\lambda^k}{k!}$
- Severity: Pareto $P(X \leq x) = \begin{cases} 1 - \left(\frac{x}{A}\right)^{-\alpha} & \text{if } x > A \\ 0 & \text{if } x \leq A \end{cases}$



Motivation

Many reasons to use this process:

1. Poisson distribution is a good candidate for a frequency
(non negative, integer, easy to use,...)
2. Pareto distribution has interesting properties:
 - Shifted distribution: $P(X \leq A) = 0$
 - Heavy tail distribution: $\int_{0-}^{\infty} e^{\lambda x} dF(x) = \infty \quad \forall \lambda > 0$
3. Direct extension : Binomial-negative – Generalized Pareto process



What if...

The large loss threshold is not equal to the A of the Pareto?

- Should these losses be added to attritional?
 - The large loss behavior is often very different from attritional losses
- Should the A be reduced?
 - Force a Pareto where there is no evidence of such a distribution
- ✓ Use a mixture distribution e.g.
 - ✓ Other distribution (empirical / exponential / ...) between the large loss threshold and A
 - ✓ Pareto above A

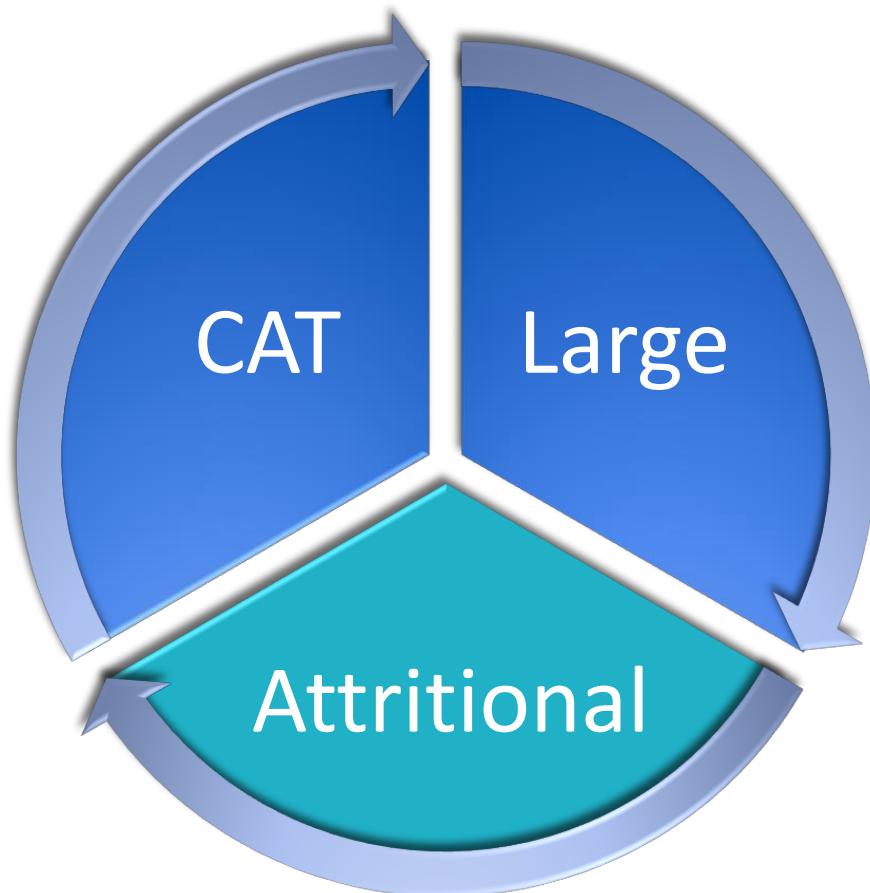


Alternatively

When we have a profile we can use it to derive exposition curves ... Remember ?

That's actually what you might do !

CLA split





Attritional losses

- Represent the mass of « small » claims

$$A(t) = Total(t) - Large(t) - CAT(t)$$

→ Function of the large loss threshold

- Often expressed on a loss ratio basis (LR)
- Lognormal model

$$A(t) \sim LogNorm(\mu, \sigma^2)$$

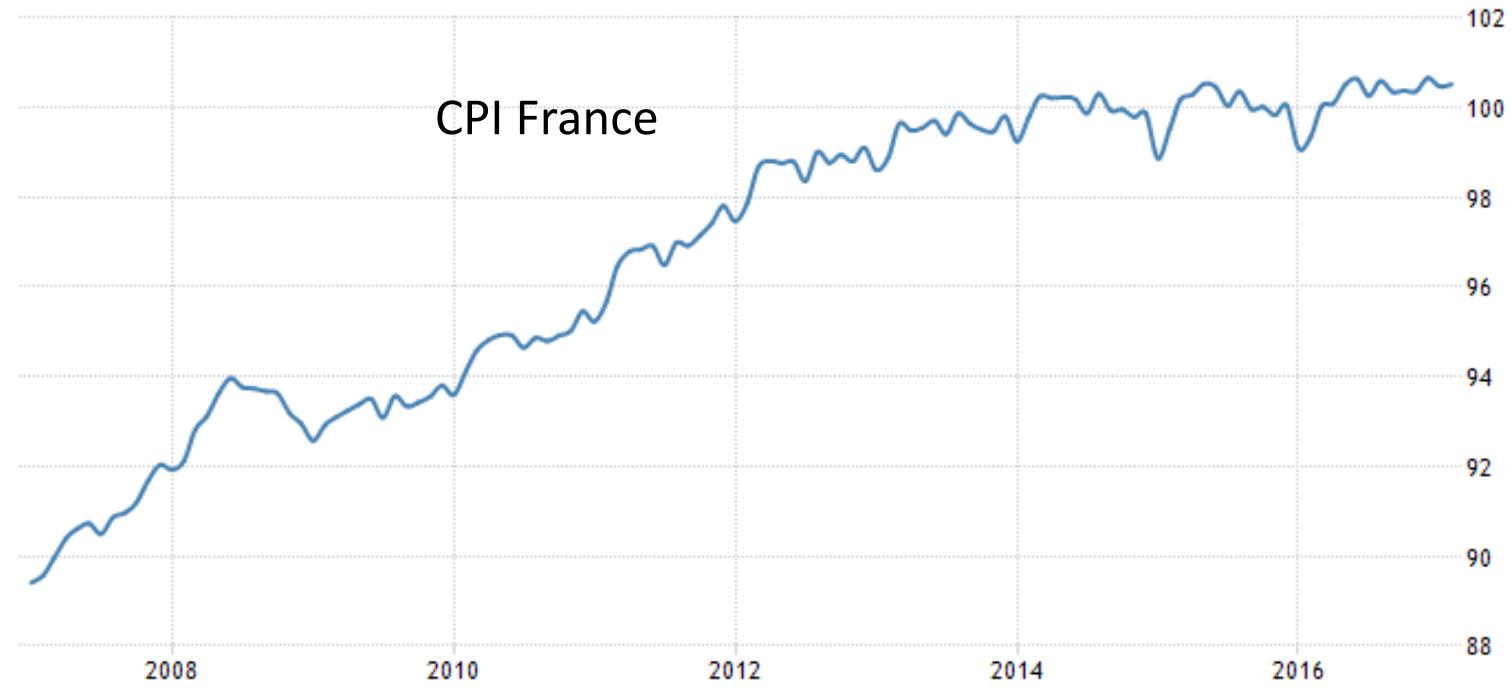


In Your case

- You have to spot CAT losses, and remove them from the sample → They'll be modelled by the Cat Model given
- You're then left with a mix of Attritional and Large Losses
- You have to determine a Threshold for separating the two
 - Based on a forfaiary criteria ?
 - Based on the volatility of the remaining attritional ?
 - Based on the reinsurance structures you have to price ?
- ... And you can make a LAC split and treat data accordingly ...

To be treated

- Claim indexation (Liability claims ... Wages !? Consumer Price ? Property claims Your choice!)
- Premium indexation (e.g. consumer price)



➤ To be found and forecasted!



To be treated (2)

- Claim data=> how to apply indexation? How to index the Sum Insured ? Are they useful ?
- CLA split?
- As-if data?
- Are all data comparable ? Are there any trend in frequency, severity, Loss Ratios ? What to retain as a final model ?

DO NOT FORGET TO JUSTIFY THE ASSUMPTIONS

Questions - suggestions?



"Harris, when I said 'any questions' I was using
only a figure of speech."