Onset thresholds by the Extreme Value Theory: A practical application in non-life insurance

Inmaculada Peña
MAPFRE Spain

CAE Fall 2014 Meeting
Zurich, 31 October 2014

AGENDA

1. Introduction: framework, motivation, proposal and goals
2. Methodology: Extreme Value Theory
3. Empirical analysis
4. Conclusion

1. INTRODUCTION

Framework:

Sustainability of the insurance business: appropiate premiums to provide for the future losses.

Actuaries are aware of the potential risk inherent in large claims.

Topic of interest in many fields in insurance, reinsurance and finance.

Motivation:

Insurance companies must control and quantify the risks, specially those whose impact is considerable such as the large claims.

1. INTRODUCTION

Proposal:

Applying a scientific methodology, the Extreme Value Theory, to find an appropiate method to explain extremal events and to set the optimal threshold.

Goals:

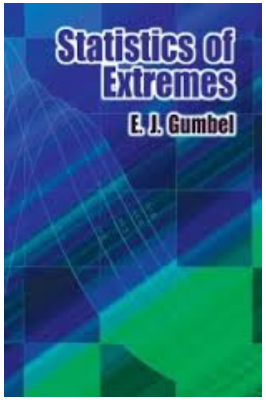
- To analyze the tail distribution for the large claims separately of the rest of claims.
- To choose the optimum threshold.
- To obtain the statistical model of large claims for Liability products, testing for the appropriate statistical distribution of claim data.

2. METHODOLOGY

THE EXTREME VALUE THEORY (EVT)

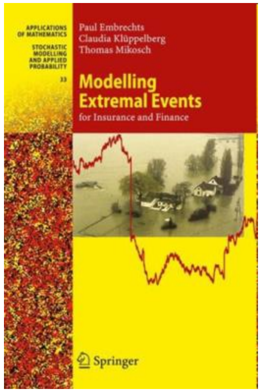
A serie of statistical and probabilistic techniques that allow modelling with a precise mathematical description the very low or quite high values of a variable, the tail distribution.

Textbook summary with applications to insurance and finance:



(1958)

E.Gulius Gumbel



(1997)

Paul Embrechts

2. METHODOLOGY

Relevance:

EVT in modelling extremal events is the counterpart of Central Limit Theorem (CLT) for sums.

- CLT is concerned about small fluctuations around the mean. The mean of a large iid random variables will be distributed according to the Normal distribution.

- EVT provides asymptotic behavior of the extreme realizations. The maximum of a large iid random variables will be distributed according to Gumbel, Fréchet or Weibull distributions.

EVT allows us to deepen our knowledge of models and tools concerning the tail distributions.

Problem:

Extremes events are rare so it implies no much data.

2. METHODOLOGY

Why use EVT?

Principal methodology for extreme values.

Applications in different fields, importance in the insurance business.

To predict extremal events and therefore to protect against adverse effects caused by them.

Common applications:

Choose the threshold value.

To analize the tail distribution and checking for the goodness of fit.

Estimating the probability of occurrence that a claim exceeds an amount.

2. METHODOLOGY

- **Generalized Extreme Value Distribution (GEV):**

GEV is the limiting distribution for the maximum values.

Considering a sequence of iid variables $\{X_i\}_{n \epsilon N}$ with a common distribution G(x).

$G_\xi(x)$ describes the limiting distribution of normalised maximum.

The Jenkinson-Von Mises representation provides a unified formula of one-parameter ξ for the three possible limiting distributions, Fréchet, Weibull and Gumbel.

$$G_\xi(x) = \begin{cases} exp\left\{-(1+\xi x)^{-1/\xi}\right\} & if \;\; \xi \neq 0 \\ exp\{-exp\{-x\}\} & if \;\; \xi = 0 \end{cases}$$

2. METHODOLOGY

- **Generalized Pareto Distribution (GPD):**

GPD is the limiting distribution for the tails.

Focused on those events that exceed a given threshold *u*.

$$Pr\{X > u + y \mid X > u\} \sim H_{\xi}(y)$$

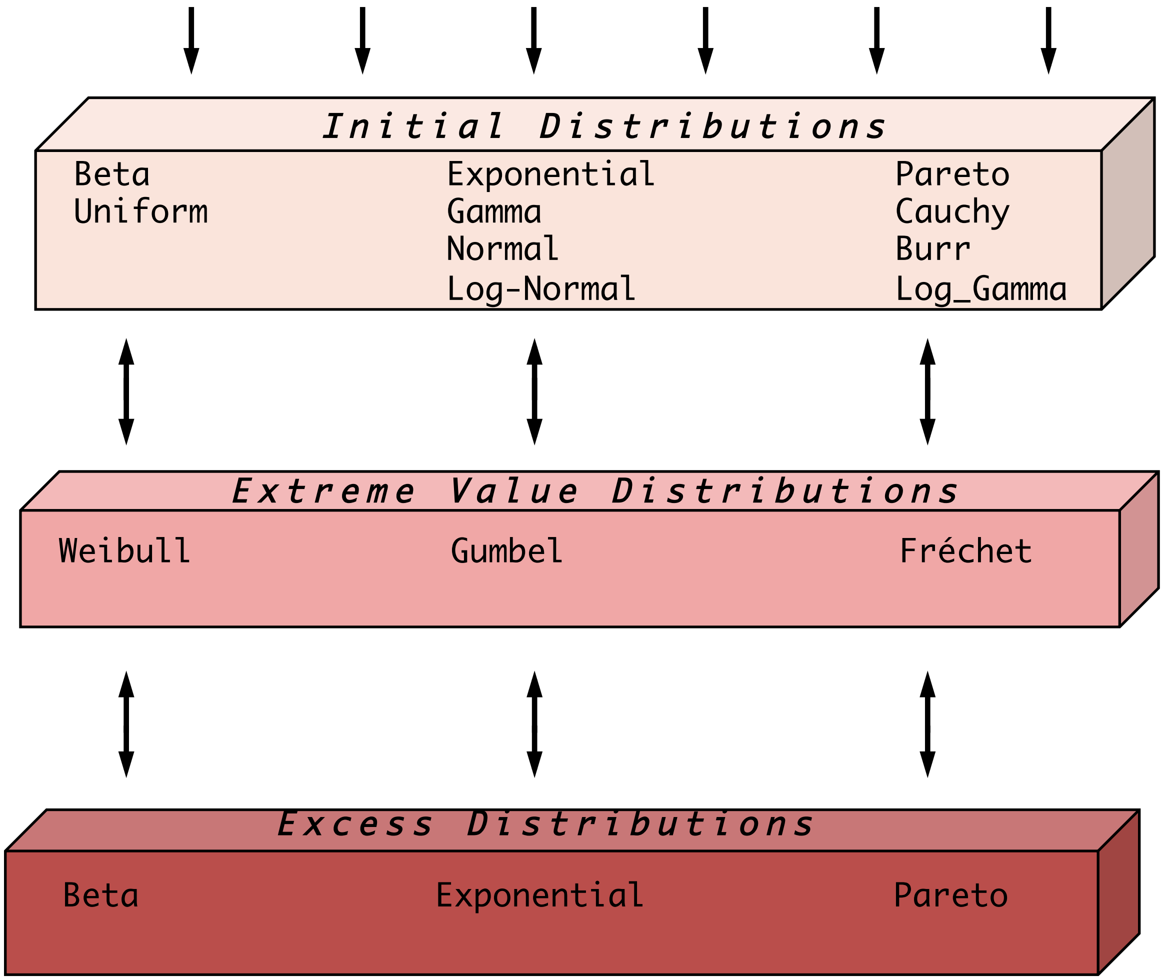
$H_{\xi}(y)$ the asymptotics excess distribution

Theorem of Pickands-Balkema-De Haan:

If and only if F $\in$ MDA(G_{ξ}) $\xi \in$ R, the GPD is the associated distribution function of exceedances over the *u* threshold.

$$H_{\xi}(y) = \begin{cases} 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{-\frac{1}{\xi}} & if \;\; \xi \neq 0 \\ 1 - exp\left(-\frac{y}{\tilde{\sigma}}\right) & if \;\; \xi = 0, \;\; y \epsilon D(\xi, \tilde{\sigma}) \end{cases}$$

2. METHODOLOGY



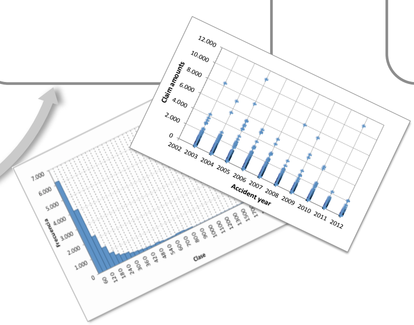
3. EMPIRICAL ANALYSIS

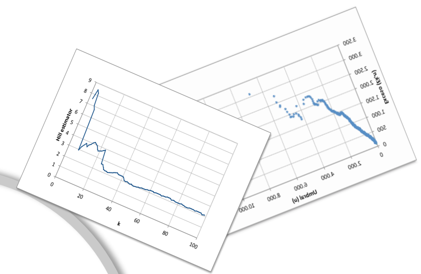
SCOPE OF DATA

- Liability products
- Comparative study
- Date and timeline
- Inflation CPI
- Correction factor

EXPLORATORY ANALYSIS

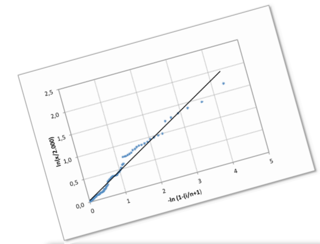
- Statistic summary
- Tool graphs:
  - Histograms
  - Amount of claims per AY





TAIL ANALYSIS

- Empirical Mean Excess Function
- Hill plot



THRESHOLD SELECTION & EXCESS MODEL

- Assuming Pareto family
- Parameters Estimation: maximum likelihood
- Checking model fit:
  – QQ-Pareto
  – Chi-square test

IL α=2,18

PL α=1,79

3. EMPIRICAL ANALYSIS

PHASE 1: SCOPE OF DATA

- Data-set: corresponds to Industrial and Professional Liability insurance claims.
- Source data: one direct insurance company.
- Comparative study.
- Horizon of 10 years, period 2003-2012 inclusive.
- Inflation update (Spanish consumer price index).
- Correction factor: data in monetary units (m.u.).
- Screening: settled claims selecting only losses ≥ 35 m.u.

3. EMPIRICAL ANALYSIS

PHASE 2: EXPLORATORY ANALYSIS

- Statistic summary of the claim amounts:

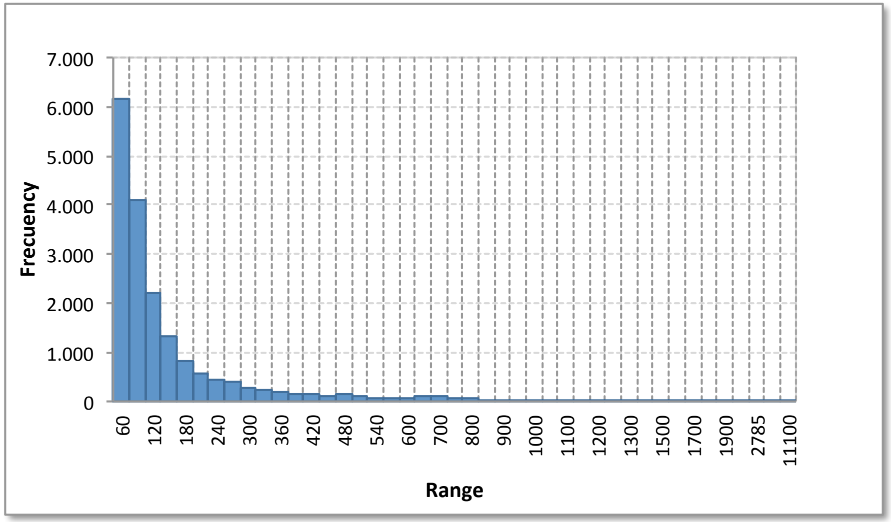
Statistic	Industrial Liability	Professional Liability
N	18.362	2.007
Minimum	35,01	35,13
1st quartile	52,81	56,26
Median	80,26	89,94
Mode	45,20	83,90
Mean	161,08	232,04
Standard Dev.	2,33	11,54
Skewness	12,76	8,76
3rd quartile	150,08	211,63
Maximum	10.821,93	9.293,08
$x_{0,99}$	1.185,49	2.349,58

3. EMPIRICAL ANALYSIS

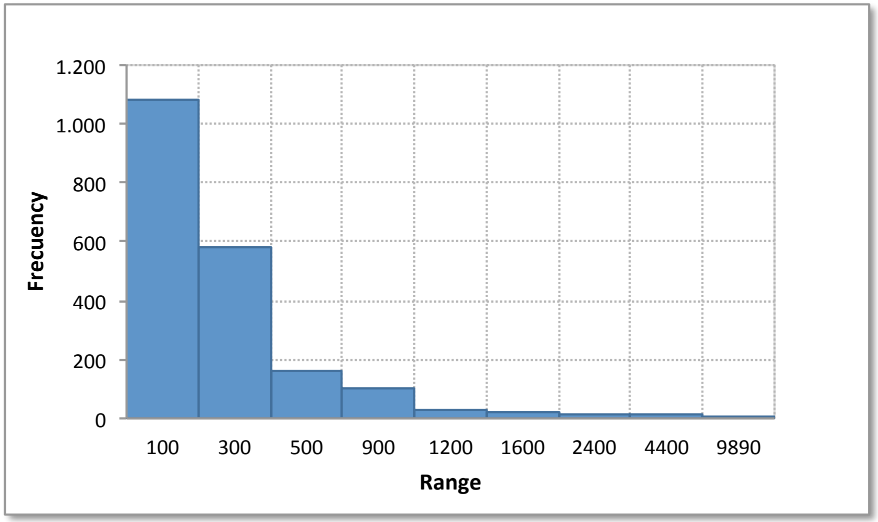
PHASE 2: EXPLORATORY DATA ANALYSIS

- Frequency histograms of the claim amounts:

Industrial Liability



Professional Liability

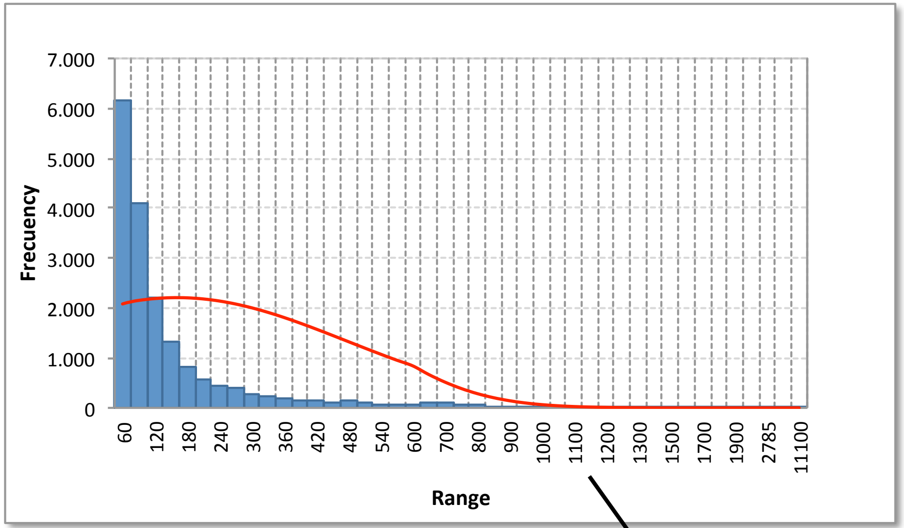


3. EMPIRICAL ANALYSIS

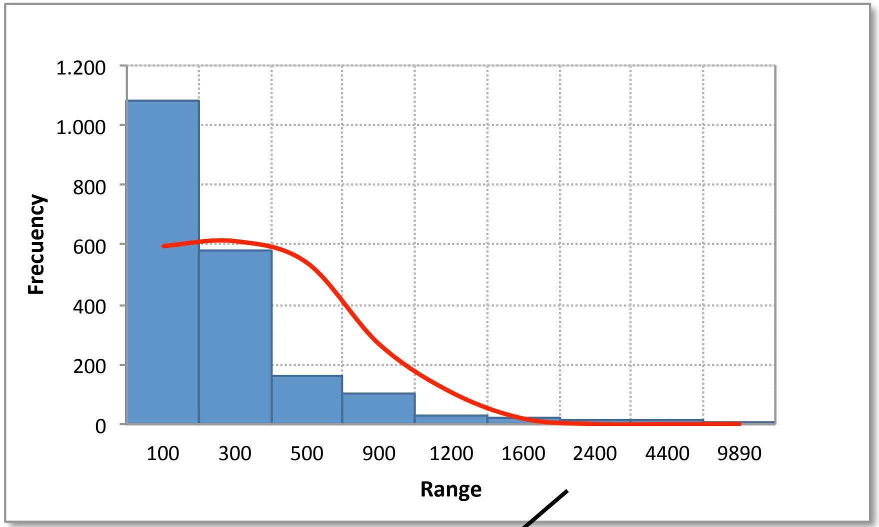
PHASE 2: EXPLORATORY DATA ANALYSIS

- Frequency histograms of the claim amounts:

Industrial Liability



Professional Liability



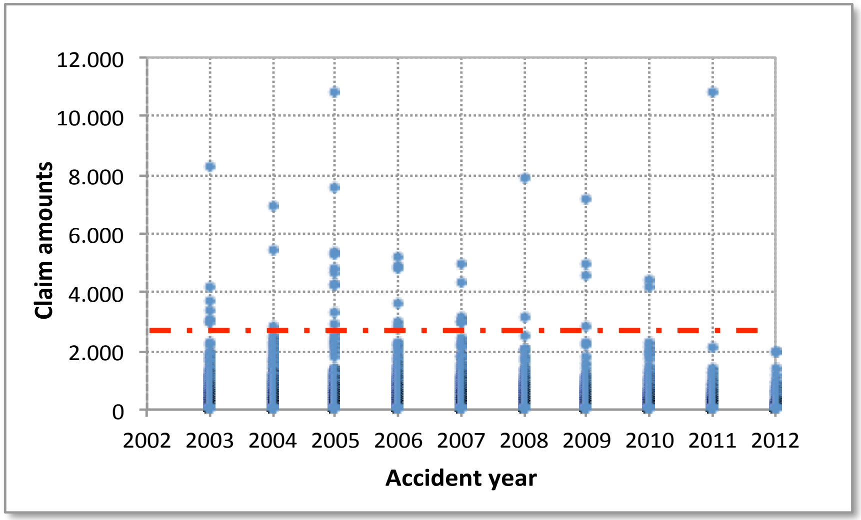
Heavy right-hand tails

3. EMPIRICAL ANALYSIS

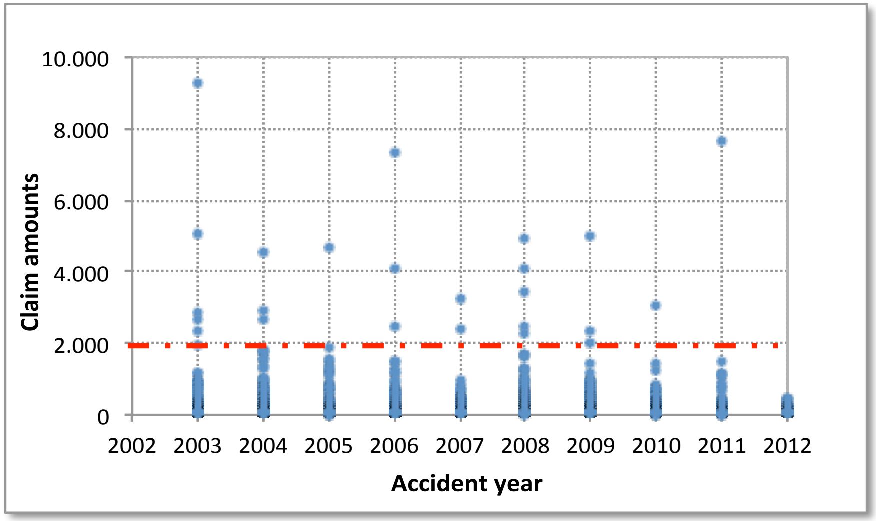
PHASE 2: EXPLORATORY DATA ANALYSIS

- Graphics of amount of claims per accident year.

Industrial Liability



Professional Liability



3. EMPIRICAL ANALYSIS

PHASE 3: TAIL ANALYSIS

Analysis of heavy-tailed events by two graphical methods:

1. **Mean excess plot (ME-plot):**

   Represents the empirical mean excess function $E_{k,n}$ that refers to the mean excess over the threshold value u.

   If $u = X_{k+1}$, then $E_{k,n}$ is the mean of the k upper order statistics such as:

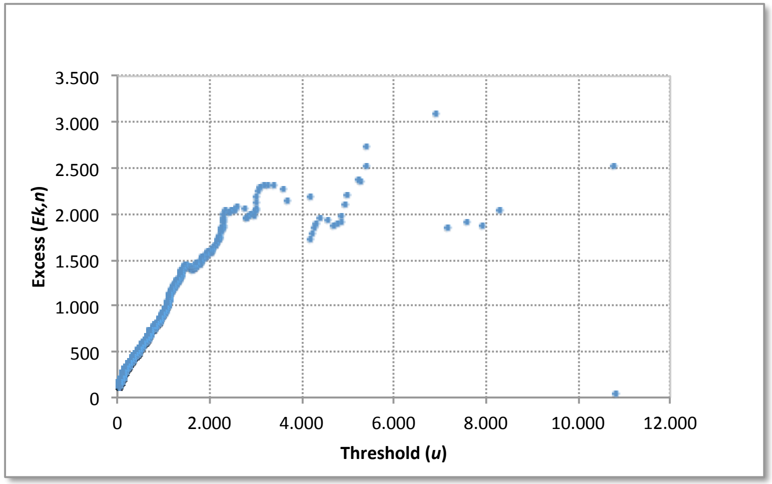
$$E_{k,n} = \widehat{e_n}(X_k) = \frac{\sum_{i=1}^{k} X_i}{k} - X_{k+1} \quad k = 1, \dots, n-1$$

   Graphic points of ME-plot $\{(X_{k,n}, e_n(X_{k,n})) : k = 1, \dots, n\}$

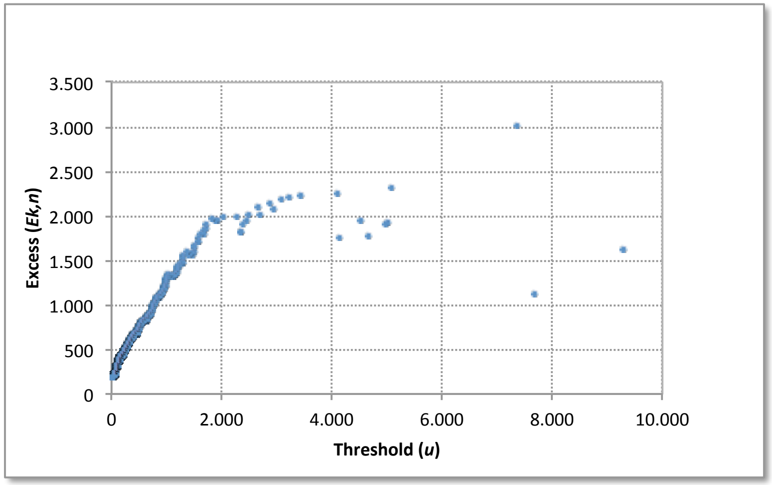
   ME-plot at work: provides a graphical test for tail behaviour and the step to confirm the GDP at the tail.

3. EMPIRICAL ANALYSIS

Industrial Liability



Professional Liability



Interpretation

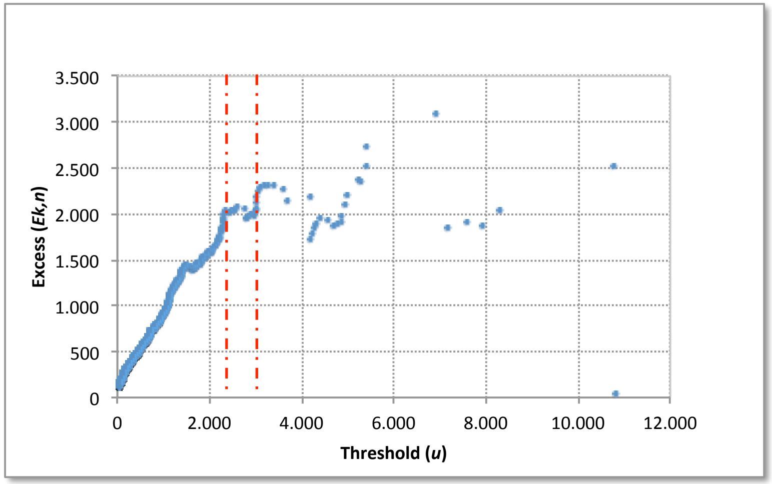
- ✓ $e_n(u)$ tends to infinite: typical heavy-tailed distribution.
- ✓ Straight line with positive slope corresponds to the Pareto distribution.
- ✓ Similar behaviour for both products.

Onset of threshold

- ✓ Change of slope.
- ✓ Selecting a range of threshold values.

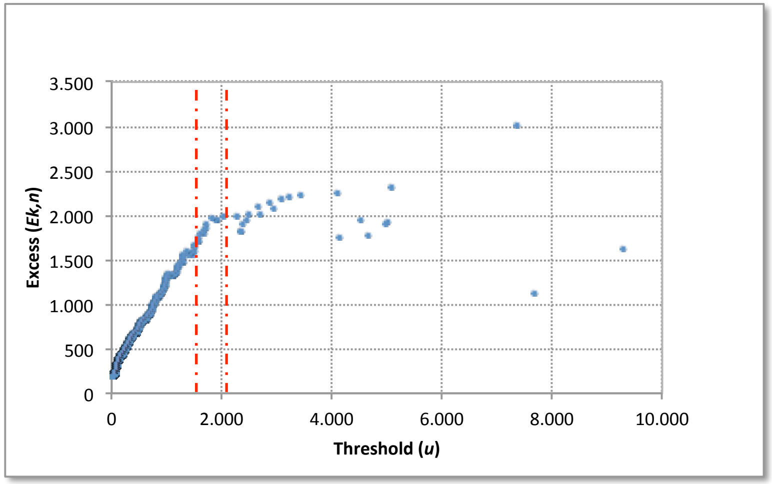
3. EMPIRICAL ANALYSIS

Industrial Liability



Threshold	Exceedances
2.200,38	61
2.895,32	37
3.013,02	33

Professional Liability



Threshold	Exceedances
1.492,10	37
2.017,10	23

3. EMPIRICAL ANALYSIS

2. **Hill plot:**

Based on the Hill estimator: it is the conditional maximum likelihood estimator for the heavy-tailed distributions.

$$\hat{\alpha}^{(H)} = \hat{\alpha}^{(H)}_{k,n} = \frac{k}{\sum_{j=1}^{k} ln \frac{X_{(n-j+1)}}{X_{(n-k)}}}$$

✓ Why use Hill estimator?

Robustness: because it only depends on the shape of the distribution tail, not on the entire distribution.

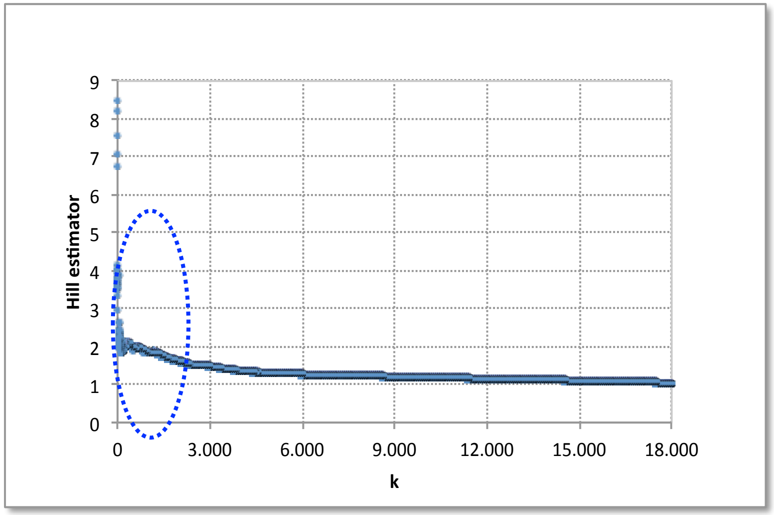
Simplicity of the formula.

✓ Does it always work?

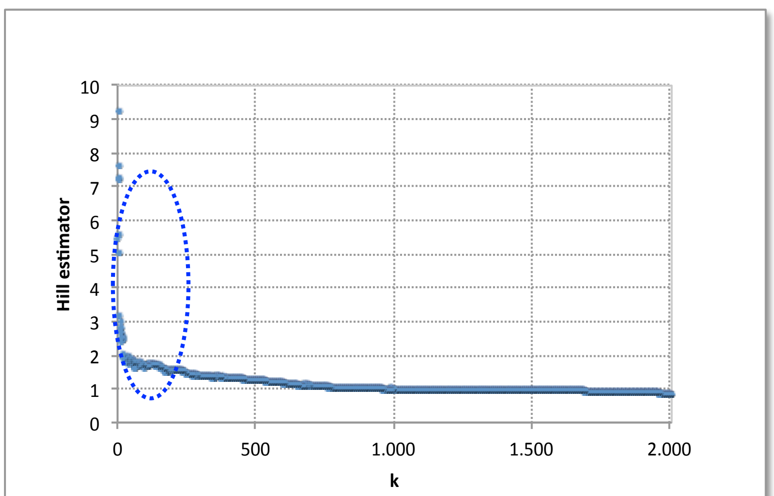
Hill plot is effective when the distribution is Pareto or close to Pareto.

3. EMPIRICAL ANALYSIS

Industrial Liability



Professional Liability



Interpretation

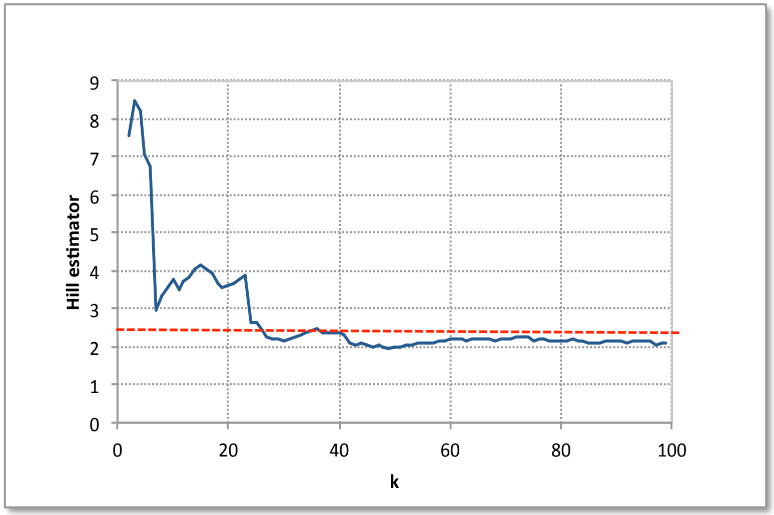
- ✓ Hill estimator keeps constant
- ✓ Similar behaviour for both cases

Onset of threshold

- ✓ Region where the estimated values are approximately constant over a range of k-values
- ✓ Best goodness of fit to Pareto distribution.

3. EMPIRICAL ANALYSIS

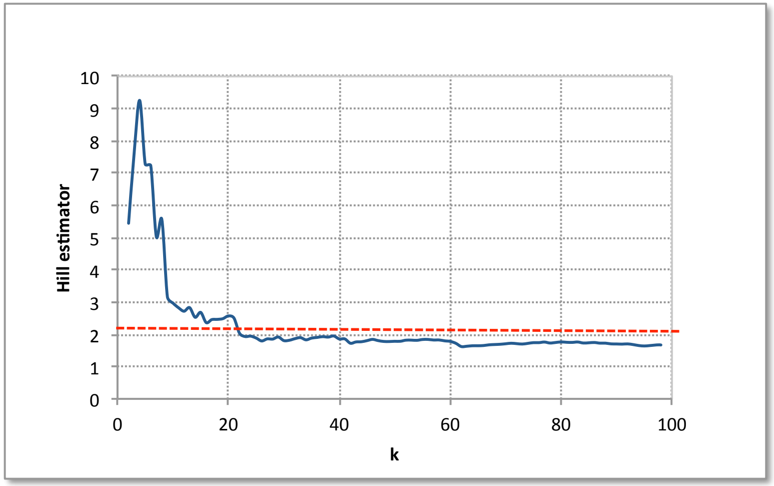
Industrial Liability



✓ Zoom in 0,5% tail size

Threshold	Exceedances	Pareto alpha
1.995,74	74	2,17
2.298,75	51	2,04
2.747,11	40	2,32

Professional Liability



✓ Zoom in 5% tail size

Threshold	Exceedances	Pareto alpha
936,76	75	1,74
1.201,16	49	1,79
1.919,86	22	2,03

3. EMPIRICAL ANALYSIS

PHASE 4: THRESHOLD SELECTION & EXCESS MODEL

- Assuming the underlying distribution behaves as a Pareto distribution.
- Estimating the parameters by the maximum likelihood method. (Hill estimator)
- Checking the goodness of fit to Pareto by two methods:

1. **Chi-square test:**

   The optimal threshold depends on the p-value: the higher p-value is the most robust the threshold value will be.

2. **Quantile Pareto plot.**

3. EMPIRICAL ANALYSIS

1. Chi-square test:

   Comparison of threshold values:

Industrial Liability

Threshold	Exceedances	Pareto alpha	p-value	Tail size
1.995,74	74	2,17	0,7360	0,4%
2.200,38	**61**	**2,18**	**0,7633**	**0,3%** ✓
2.298,75	51	2,04	0,6027	0,3%
2.747,11	40	2,32	0,7498	0,2%
2.895,32	37	2,37	0,4855	0,2%
3.013,02	33	2,32	0,4056	0,2%

Professional Liability

Threshold	Exceedances	Pareto alpha	p-value	Tail size
936,76	75	1,74	0,5752	4%
1.201,16	**49**	**1,79**	**0,7720**	**2%** ✓
1.492,10	37	1,90	0,3704	2%
2.017,10	23	2,03	0,1272	1%
1.919,86	22	2,03	0,1185	1%

3. EMPIRICAL ANALYSIS

2. Quantile Pareto plot

- Hypothesis testing is designed to assess the strength of the evidence against the null hypothesis.

  H_0: the Pareto distribution provides the correct statistical model for the claims data.

  H_1: the Pareto distribution does not provide the correct model for the claims data.

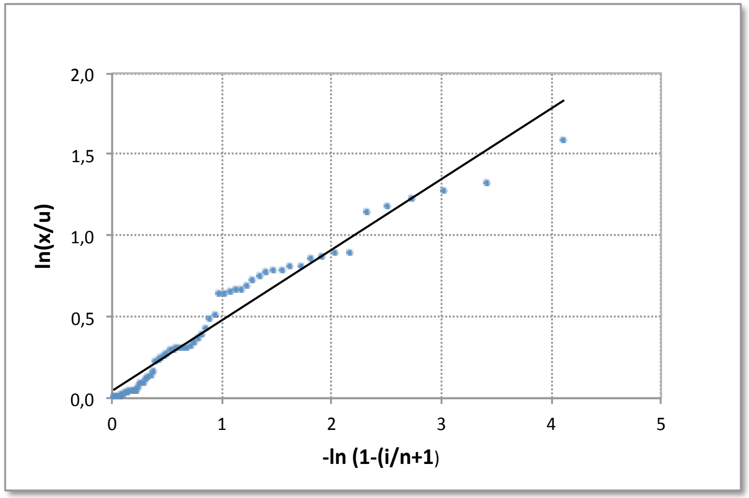
- Plotting Q-Q plot:

  Quantiles *p* order.

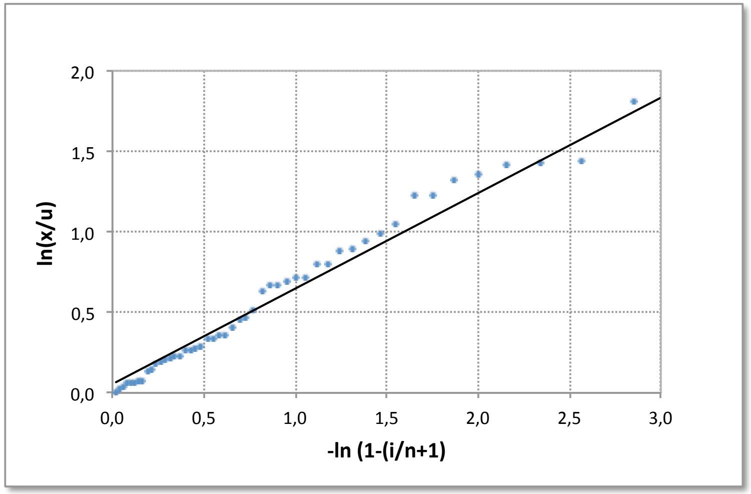
  Graphic points: $\left\{-\ln\left(1-\frac{i}{n+1}\right), \ln\left(\frac{X_i}{u}\right) : i = 1, \ldots, n\right\}$

3. EMPIRICAL ANALYSIS

Industrial Liability



Professional Liability



Quantile Pareto Plot

Product	Optimum threshold	Exceedances	R^2
Industrial Liability	2.200	61	96%
Professional Liability	1.201	49	97%

Interpretation:

- ✓ Similar behaviour for both products
- ✓ The points plotted fall on the reference line
- ✓ The null hypothesis is not rejected: Pareto distribution provides the correct model for the tail

3. EMPIRICAL ANALYSIS

- Comparison of parameters: our results vs experience of the SST.

Parameters α_i of the Pareto distribution[1]		
LoB	for β =CHF 1 million	for β =CHF 5 million
Motor vehicle liability	2.50	2.80
Comprehensive motor vehicle insurance	1.85	1.85
Property	1.40	1.50
Liability	1.80	2.00
Compulsory accident insurance	2.00	2.00
Health	3.00	3.00
Transport	1.50	1.50
Financial and surety	0.75	0.75
Others	1.50	1.50

Product	Pareto α_i
Industrial Liability	2.18
Professional Liability	1.79

[1] FINMA (2006): *Technical document on the Swiss Solvency Test*. Swiss Financial Market Supervisory Authority. Online. [Accessed 8 October 2014].

3. EMPIRICAL ANALYSIS

SUMMARY OF RESULTS

- **Exploratory analysis:**

  The analysis of data shows heavy-tailed distributions.

- **Tail analysis:**

  Mean Excess plot and Hill plot define a range of thresholds.

- **Threshold selection and excess model:**

  Testing the goodness of fit provides the criteria to select the optimal threshold and the Pareto distribution as a good excess model.

  The estimation of the parameters gives information about the tail distribution.

  The α parameters correspond with similar LoB.

4. CONCLUSION

- Extreme Value Theory summarizes the scientific methodology on actuarial modelling of extremal events, those are our large claims.
- Large claims must be analyzed separately of the rest of claims to provide more accurate estimates.
- The practical application presents a precise description of the large claims:
  - ✓ The underlying distributions have heavy tails
  - ✓ Choose the optimum threshold
  - ✓ The Pareto distribution provides a good model for the excesses over a threshold
  - ✓ The parameters correspond to similar LoB

4. CONCLUSION

- The Extreme Value Theory provides techniques that allow a better risk control that ensures the solvency of the insurance companies.

- This methodology corresponds to a theorical view but it is paramount to attain the best model that the actuaries should apply their practical approach.

Thank you for your attention

ipenasa@mapfre.com

A