GUY CARPENTER

6/7 June 2011

Pitfalls of Curve Fitting for Large Losses

Guy Carpenter

Amit Parmar, Michael Cane

Antitrust Notice

- The Casualty Actuarial Society is committed to adhering strictly to the letter and spirit of the antitrust laws. Seminars conducted under the auspices of the CAS are designed solely to provide a forum for the expression of various points of view on topics described in the programs or agendas for such meetings.

- Under no circumstances shall CAS seminars be used as a means for competing companies or firms to reach any understanding – expressed or implied – that restricts competition or in any way impairs the ability of members to exercise independent business judgment regarding matters affecting competition.

- It is the responsibility of all seminar participants to be aware of antitrust regulations, to prevent any written or verbal discussions that appear to violate these laws, and to adhere in every respect to the CAS antitrust compliance policy.

Agenda

- **Introduction**
- **Theoretical analysis**
  - Data sample size issues
  - Model uncertainty
  - Parameter error
  - Summary
- **Real-world analysis**
  - UK Motor market fitting
  - Individual clients versus market curve
- **Summary**
- **Questions**

Introduction

Introduction

What is curve fitting?

"Curve fitting is the process of constructing a curve, or mathematical function, that has the best fit to a series of data points, possibly subject to constraints"

- For today → Curve fitting is a method to model historic claims
  - We assume observed losses:
    - Follow a statistical distribution
    - Independent and identically distributed
    - Homogeneous

Introduction

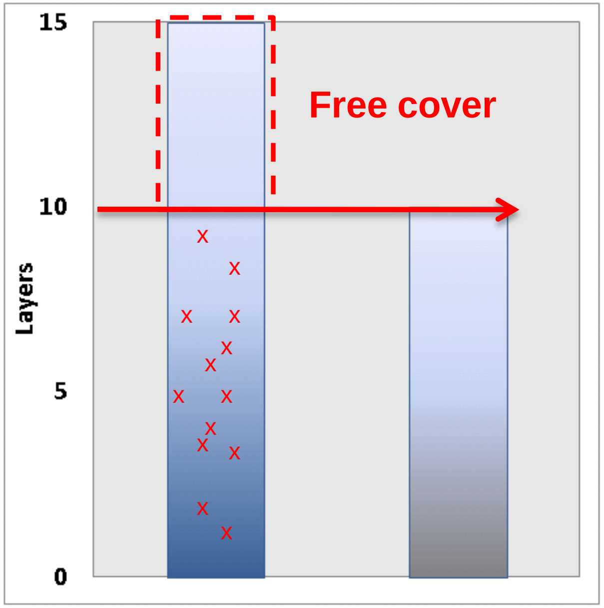
What is curve fitting used for?

- Understanding the historical data and simplifying data sets
- Modelling where there are few data points
- Understanding the potential tails of claims sets
- Reducing sample variation

Introduction

Why is curve fitting important for actuaries?

- Inherent advantages to knowing the frequency and severity rather than the expected loss
- Stochastic modelling
- Benchmarking exercises
- Helps with pricing layers above data points
- Helps alleviate free-cover problem in experience rating
- Exposure rating may not be possible
- Fundamental to the output of capital modelling



Introduction

How do we curve fit?

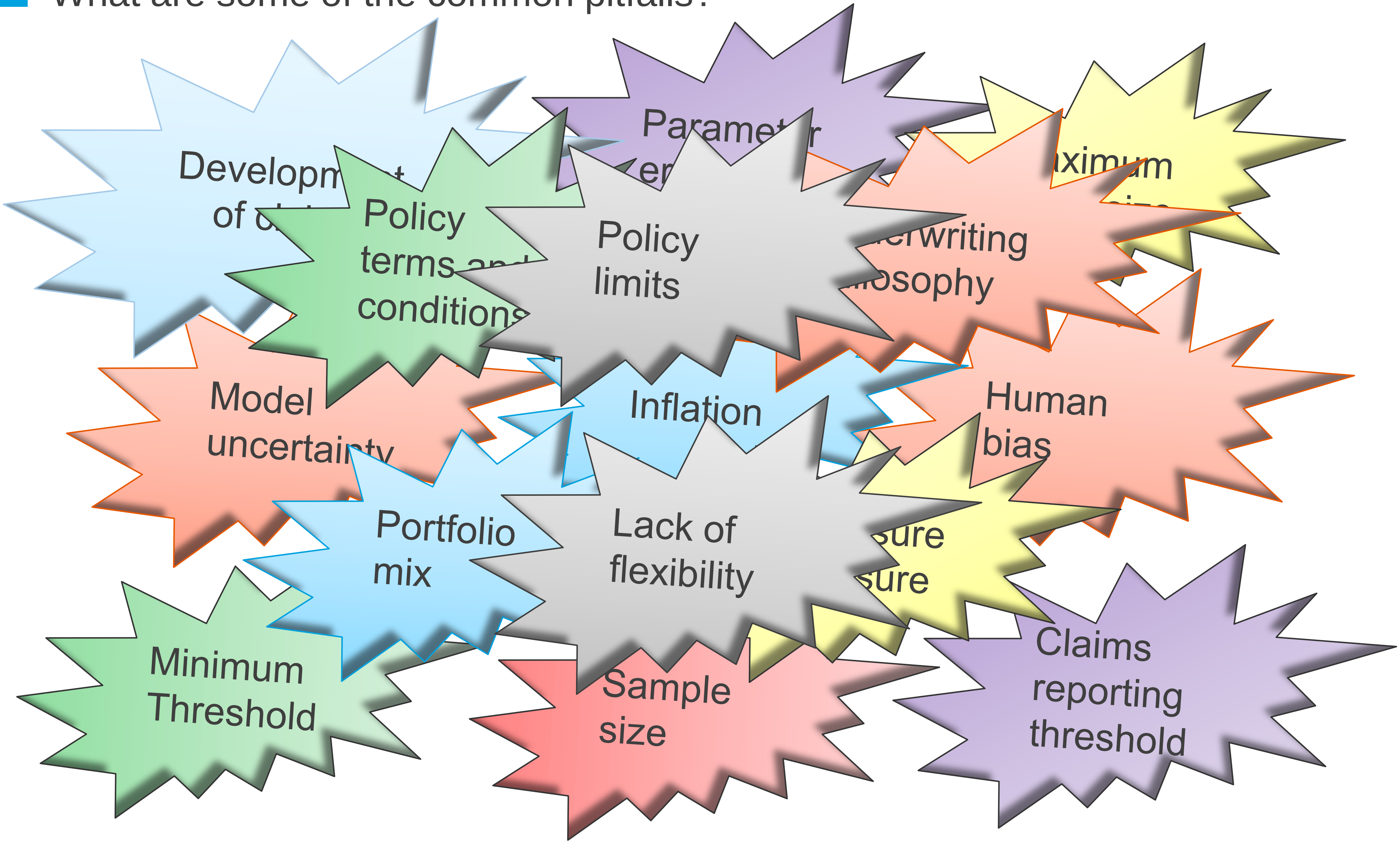
1. Consider a number of parametric probability distributions as contenders for explaining your claim set
   - Subjective list

2. Estimate parameters for each distribution
   - Method of moments
   - Maximum log-likelihood
   - Least squares estimation

3. Specify criteria for choosing fitted distribution
   - Goodness of fit tests
   - Inspection

Introduction

What are some of the common pitfalls?

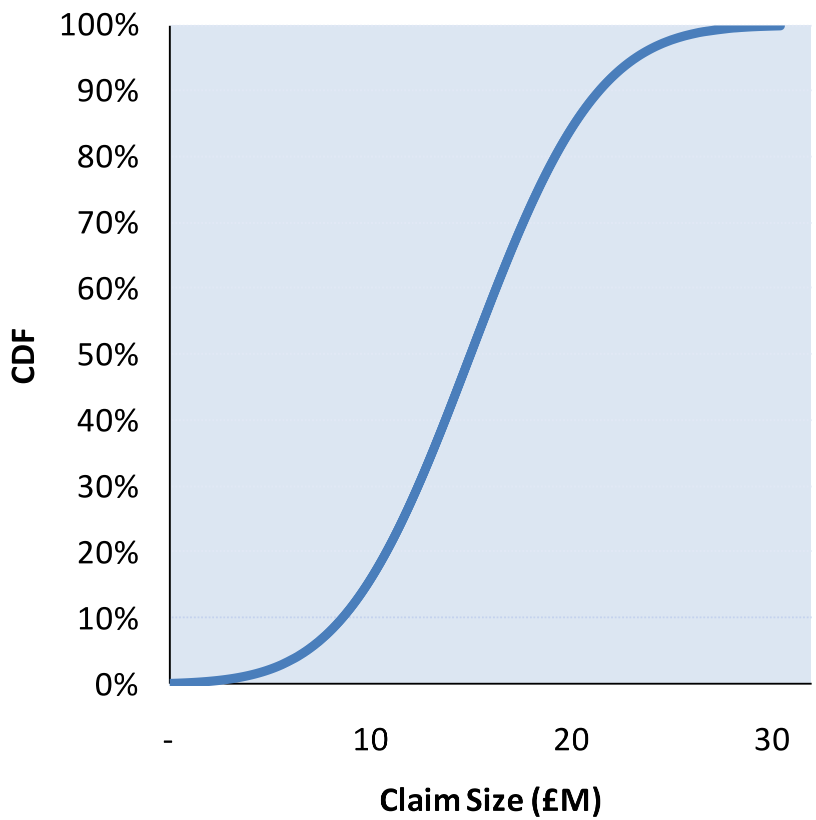


Theoretical analysis

Theoretical analysis

If we sample from:

- A known distribution
- With known parameters

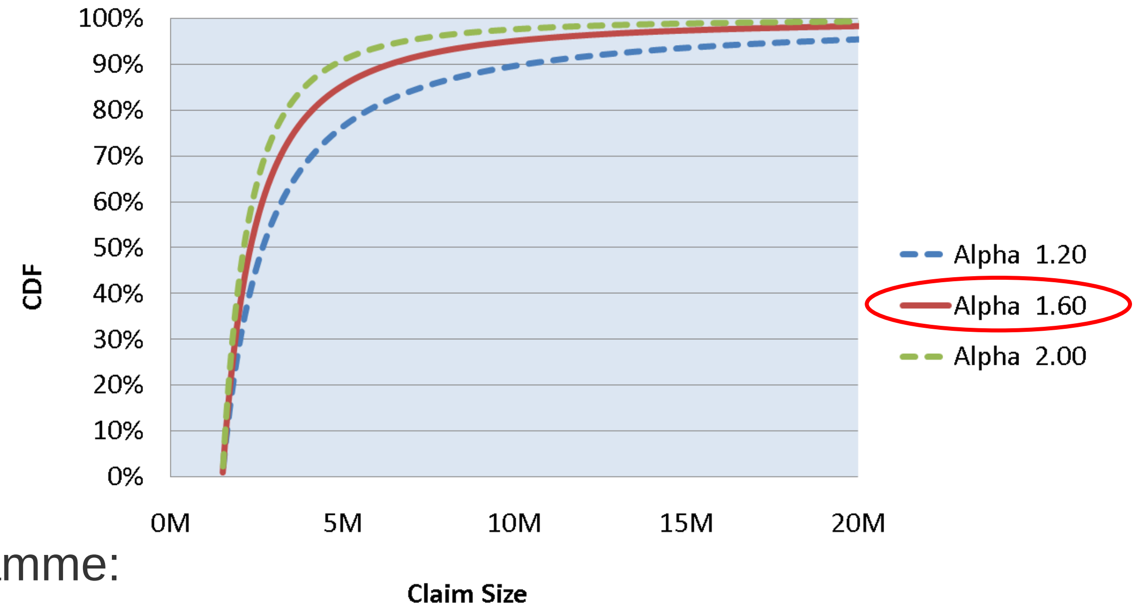


Is it possible to go wrong?

Lets find out…

Theoretical analysis
Our experiment

- Sample sizes
  - 30, 300 & 3000 ultimate claim data samples
- Distribution
  - Simple Pareto



- Parameters
  - Alpha = 1.6
  - Lambda = 1,500,000
- Reinsurance structure
  - Common motor programme:

    £3m xs £2m

    £5m xs £5m

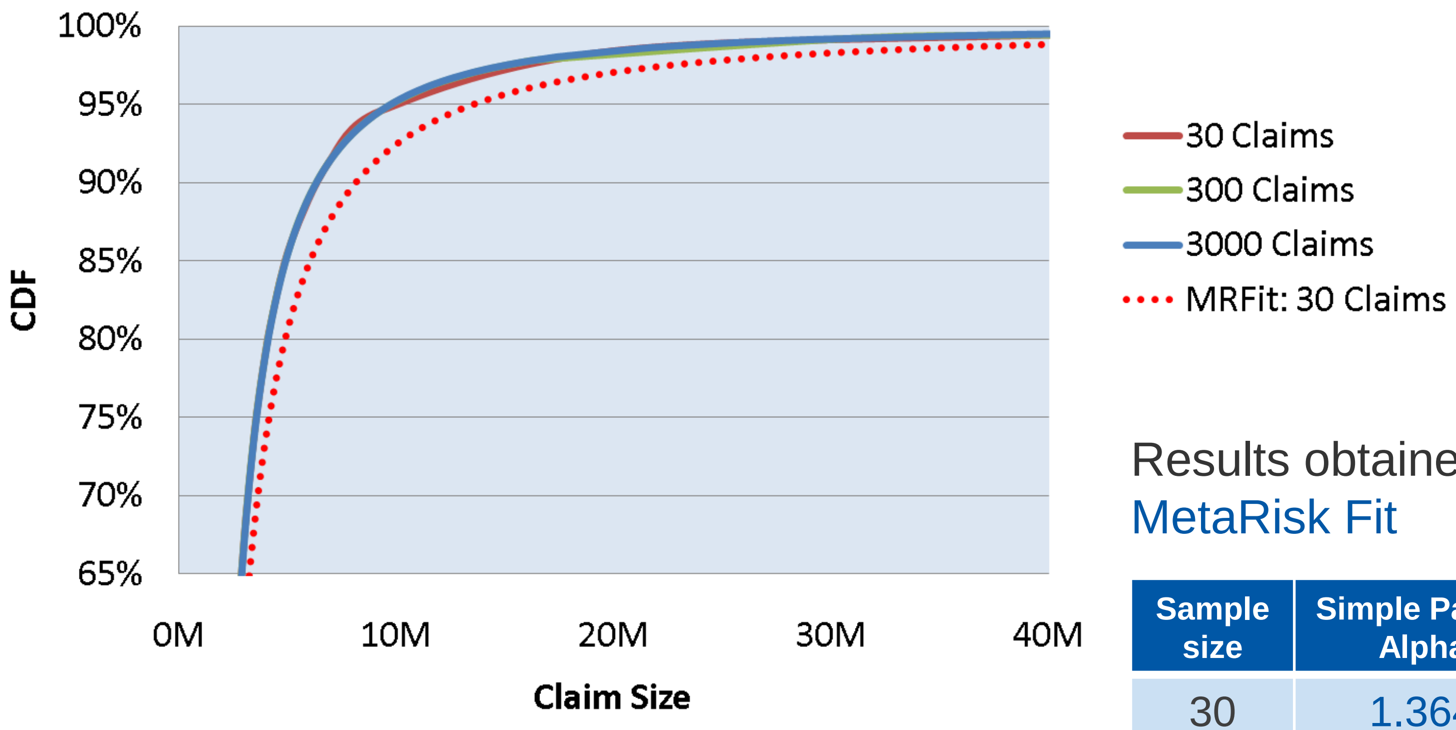
    £15m xs £10m

    Unlimited xs £25m

Data sample size issues

Theoretical analysis - Data sample size issues

What are the implications of insufficient data?



Results obtained using MetaRisk Fit

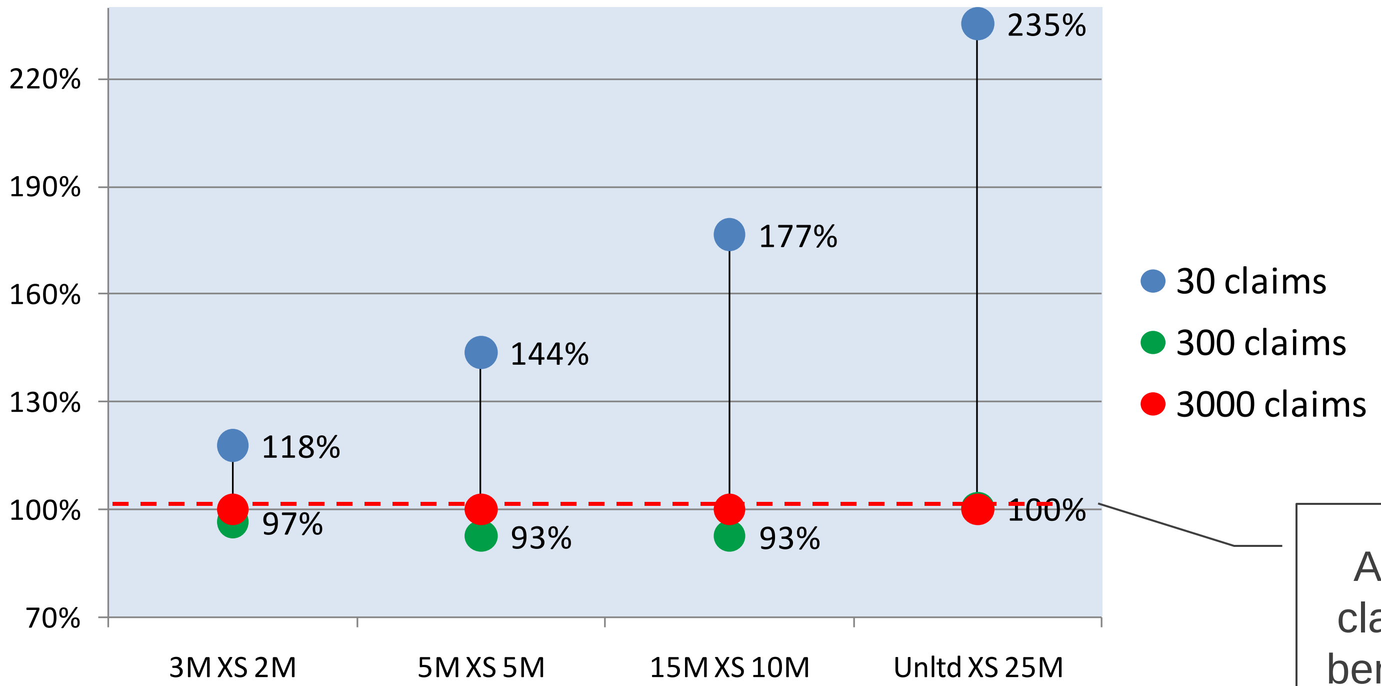
Sample size	Simple Pareto Alpha	CV
30	1.364	0.193
300	1.648	0.059
3000	1.596	0.019

How does the low sample size affect the pricing?

Theoretical analysis - Data sample size issues

Loss cost to the layer

Pricing using Simple Pareto distribution from each data set



Assume 3000 claims provides benchmark price

Significantly mis-priced with small data sample

Model uncertainty

Theoretical analysis – Model uncertainty

Suppose we have:

- Sufficient data: ✓
  - 3000 claim data sample

- What can go wrong?

- Distribution: ?
  - What are the chances of selecting the correct distribution?

What is the effect on our pricing?

Theoretical analysis – Model uncertainty

Possible severity distributions

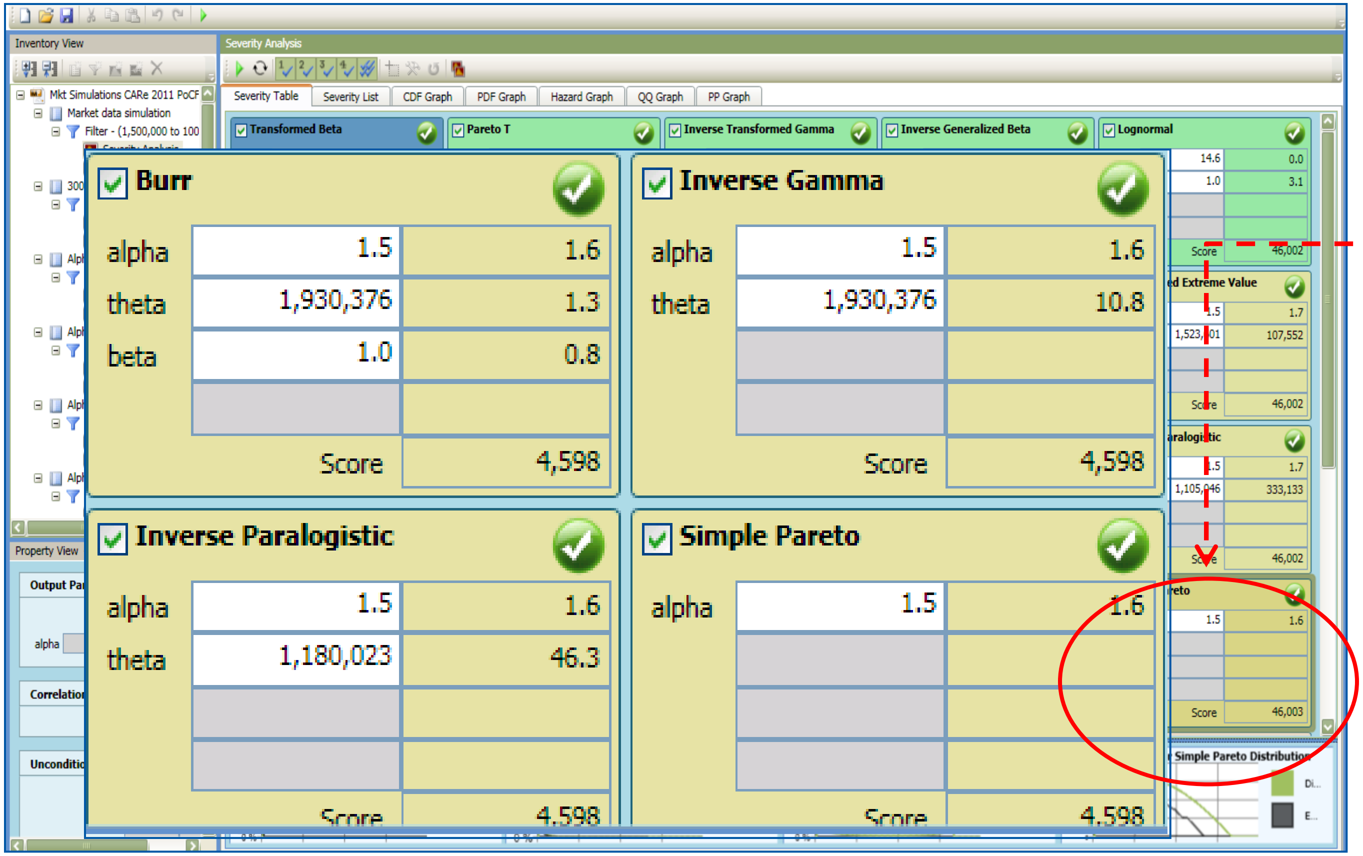
MetaRisk Fit – Severity distributions		
Simple Pareto	Lognormal	Pareto T
Extreme Value Limit	Generalized Cauchy	Inverse Transformed Gamma
Exponential	Normal	Split Simple Pareto
Inverse Paralogistic	Uniform	Transformed Gamma
Loglogistic	Generalized Extreme Value	Inverse Burr
Paralogistic	Extremal Pareto	Burr
Loggamma	Ballasted Pareto	Transformed Beta
Gamma	Power	Generalized Beta
Inverse Weibull	Beta	Inverse Generalized Beta
Inverse Gaussian	Inverse Beta	
Inverse Gamma	Generalized Pareto	

Key: **1-Parameter** **2-Parameter** **3-Parameter** **4-Parameter**

Common distributions used to conduct our analysis

Theoretical analysis – Model uncertainty

Chances of getting the wrong distribution with sufficient data



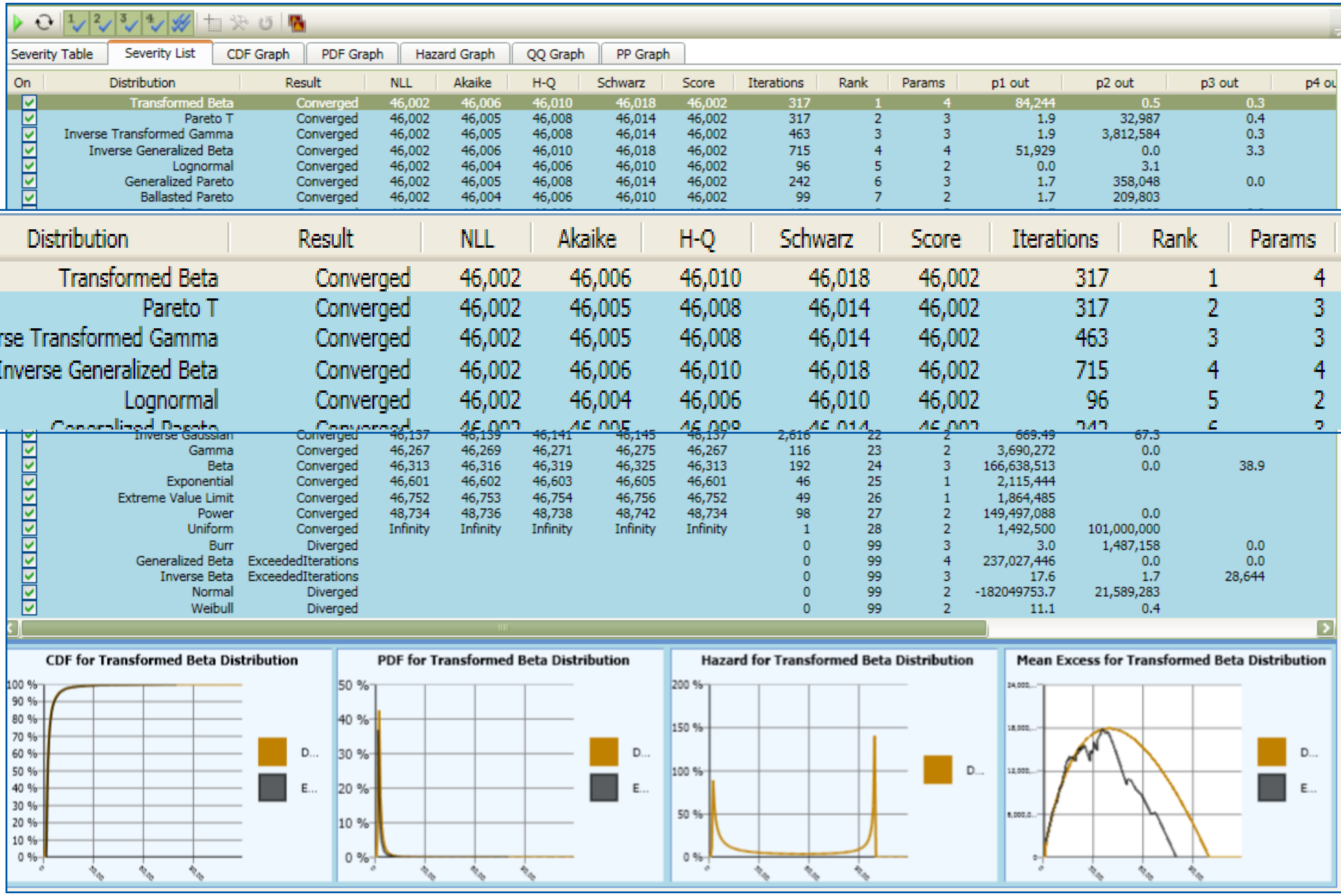
MetaRisk Fit:

Simple Pareto is 1 of the 28 distributions

Theoretical analysis – Model uncertainty

How are the ranks calculated?

3000 claims

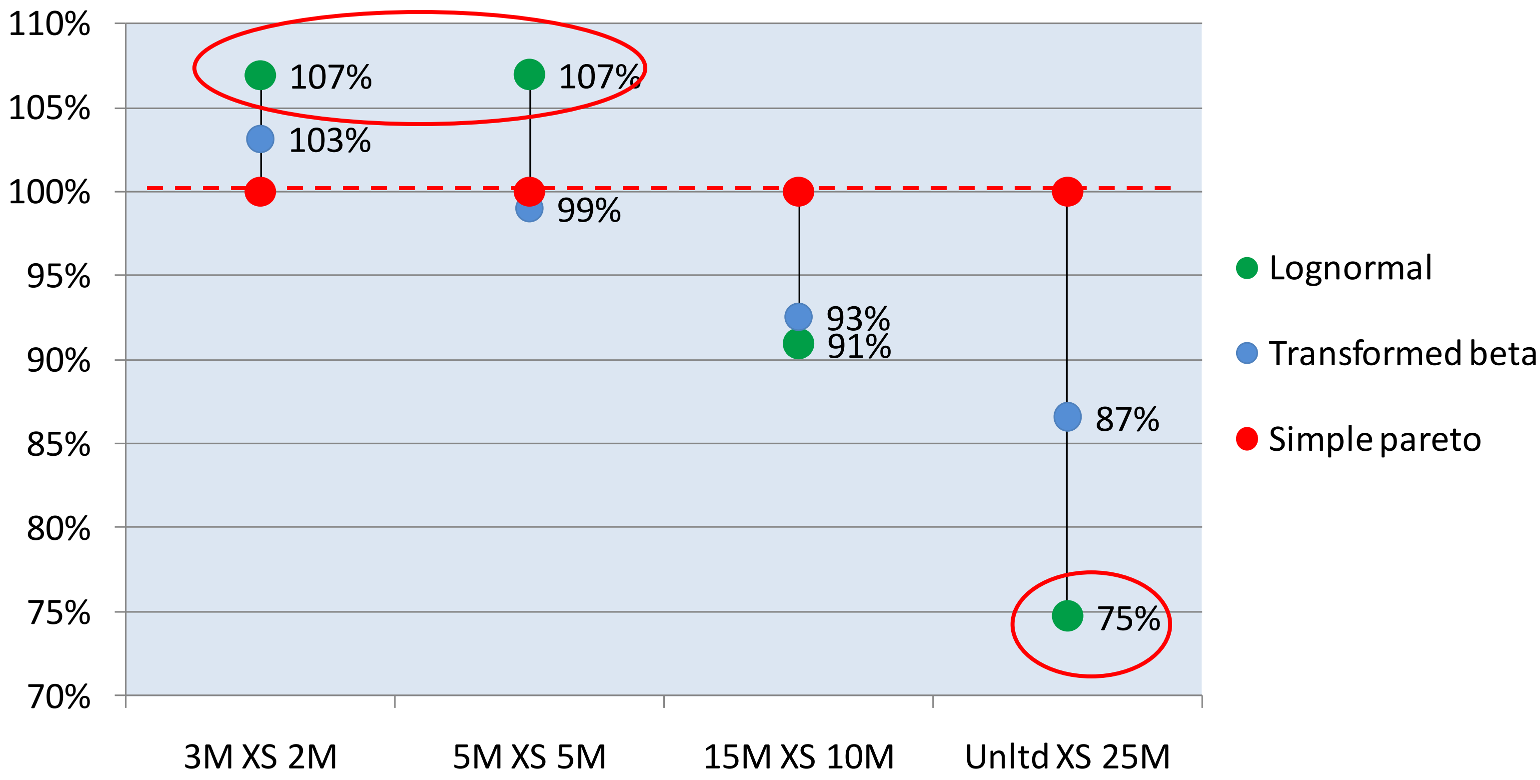


On	Distribution	Result	NLL	Akaike	H-Q	Schwarz	Score	Iterations	Rank	Params	p1 out	p2 out	p3 out	p4 ou
✓	Transformed Beta	Converged	46,002	46,006	46,010	46,018	46,002	317	1	4	84,244	0.5	0.3	
✓	Pareto T	Converged	46,002	46,005	46,008	46,014	46,002	317	2	3	1.9	32,987	0.4	
✓	Inverse Transformed Gamma	Converged	46,002	46,005	46,008	46,014	46,002	463	3	3	1.9	3,812,584	0.3	
✓	Inverse Generalized Beta	Converged	46,002	46,006	46,010	46,018	46,002	715	4	4	51,929	0.0	3.3	
✓	Lognormal	Converged	46,002	46,004	46,006	46,010	46,002	96	5	2	0.0	3.1		
✓	Generalized Pareto	Converged	46,002	46,005	46,008	46,014	46,002	242	6	3	1.7	358,048	0.0	
✓	Ballasted Pareto	Converged	46,002	46,004	46,006	46,010	46,002	99	7	2	1.7	209,803		
✓	Inverse Gaussian	Converged	46,137	46,139	46,141	46,145	46,137	2,616	22	2	669.49	67.3		
✓	Gamma	Converged	46,267	46,269	46,271	46,275	46,267	116	23	2	3,690,272	0.0		
✓	Beta	Converged	46,313	46,316	46,319	46,325	46,313	192	24	3	166,638,513	0.0	38.9	
✓	Exponential	Converged	46,601	46,602	46,603	46,605	46,601	46	25	1	2,115,444			
✓	Extreme Value Limit	Converged	46,752	46,753	46,754	46,756	46,752	49	26	1	1,864,485			
✓	Power	Converged	48,734	48,736	48,738	48,742	48,734	98	27	2	149,497,088	0.0		
✓	Uniform	Converged	Infinity	Infinity	Infinity	Infinity	Infinity	1	28	2	1,492,500	101,000,000		
✓	Burr	Diverged						0	99	3	3.0	1,487,158	0.0	
✓	Generalized Beta	ExceededIterations						0	99	4	237,027,446	0.0	0.0	
✓	Inverse Beta	ExceededIterations						0	99	3	17.6	1.7	28,644	
✓	Normal	Diverged						0	99	2	-182049753.7	21,589,283		
✓	Weibull	Diverged						0	99	2	11.1	0.4		

On	Distribution	Result	NLL	Akaike	H-Q	Schwarz	Score	Iterations	Rank	Params	p1 out
✓	Transformed Beta	Converged	46,002	46,006	46,010	46,018	46,002	317	1	4	84,244
✓	Pareto T	Converged	46,002	46,005	46,008	46,014	46,002	317	2	3	1.9
✓	Inverse Transformed Gamma	Converged	46,002	46,005	46,008	46,014	46,002	463	3	3	1.9
✓	Inverse Generalized Beta	Converged	46,002	46,006	46,010	46,018	46,002	715	4	4	51,929
✓	Lognormal	Converged	46,002	46,004	46,006	46,010	46,002	96	5	2	0.0

Theoretical analysis – Model uncertainty
Expected loss to the layer

3000 claims

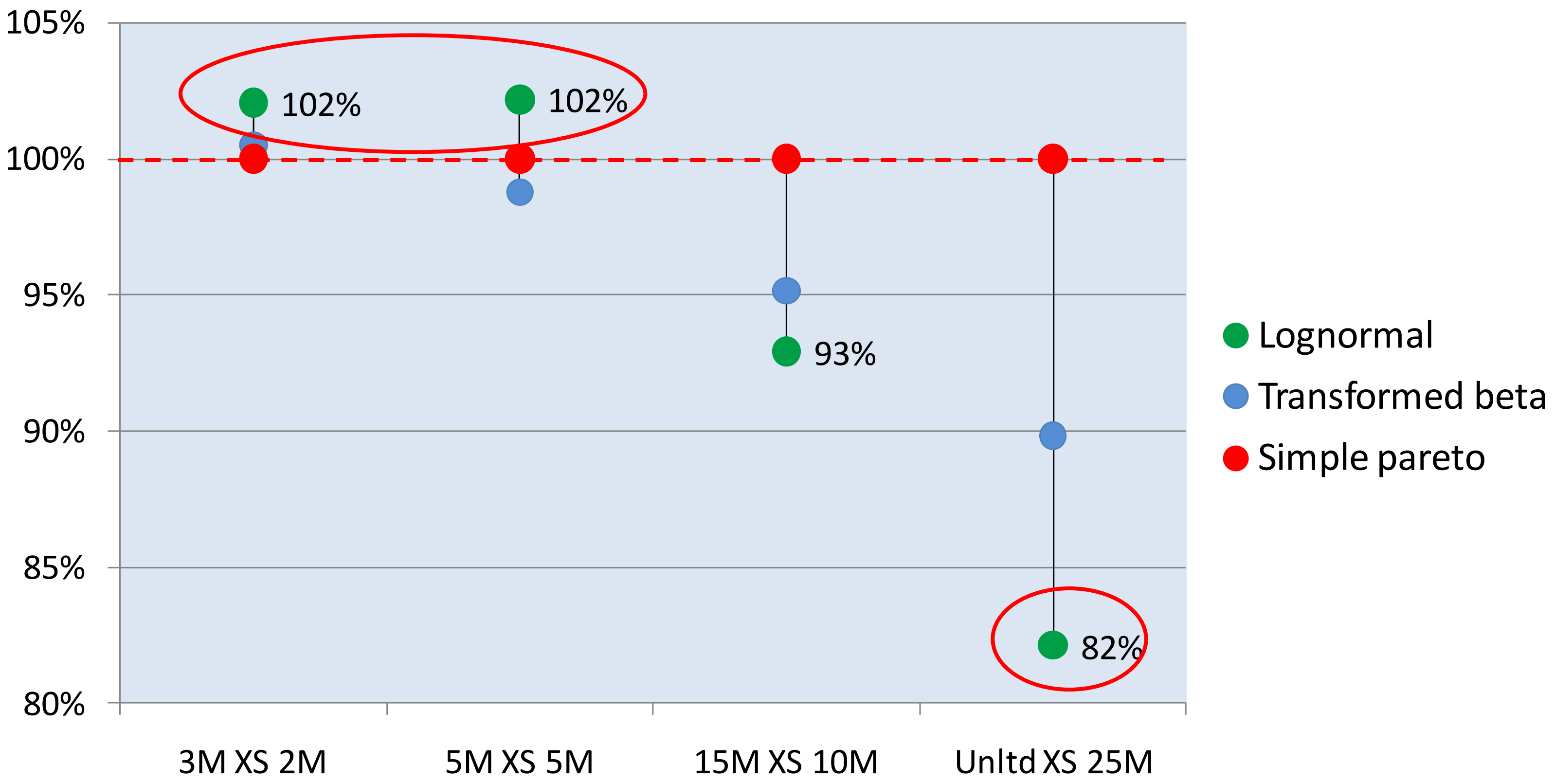


Lognormal: Over-pricing for lower layers; Under-pricing for higher layers

Theoretical analysis – Model uncertainty

Standard deviation of loss to the layer

3000 claims



Lognormal also underestimates volatility on the higher layers

Parameter error

Theoretical analysis – Parameter error

Suppose we have:

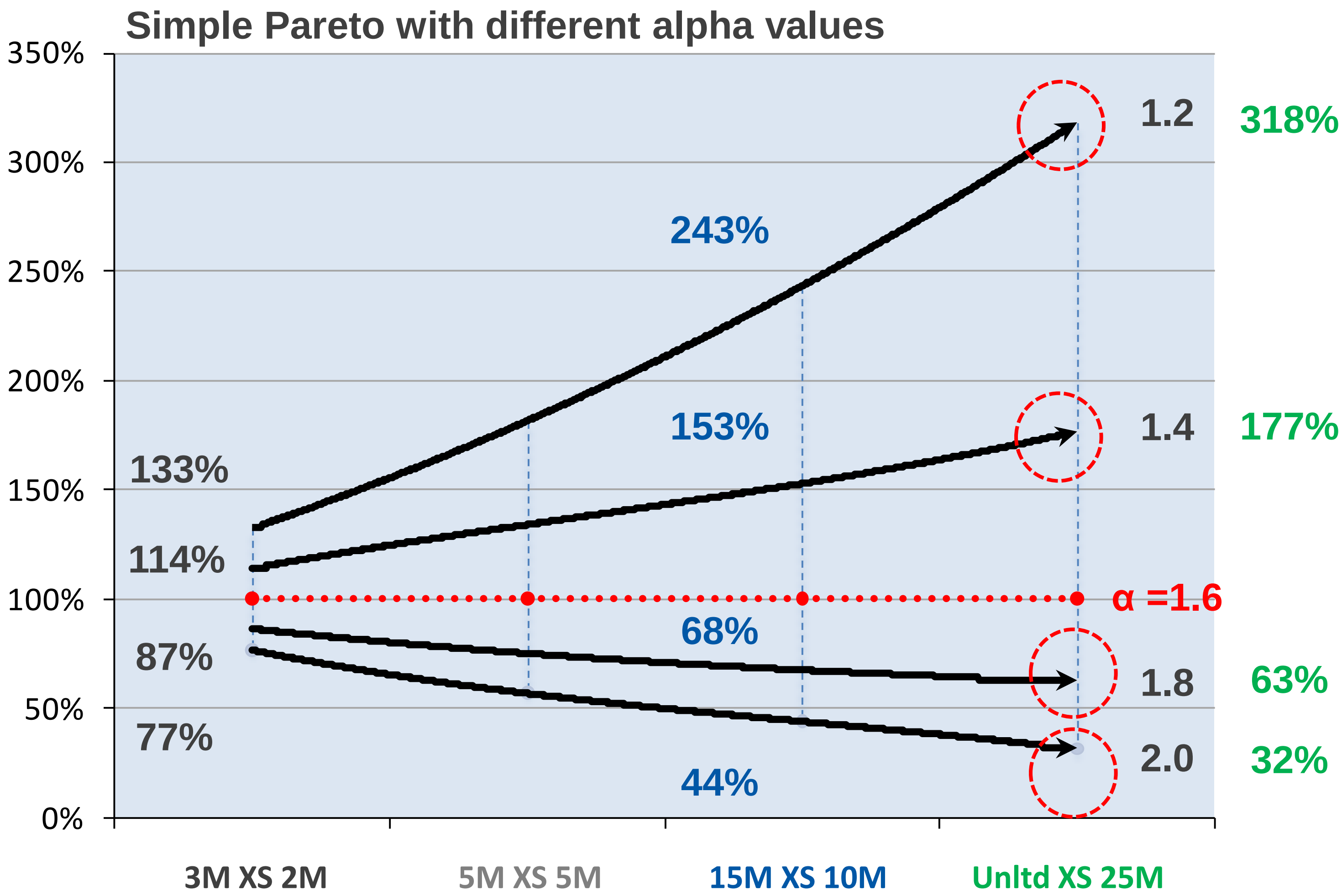
- Sufficient data:
  - 3000 claim data sample
- Correct distribution:
  - Simple Pareto

- What can go wrong?
- Incorrect parameters:
  - Instead of α = 1.6
  - We could pick lower or higher values

What is the effect on our pricing?

Theoretical analysis – Parameter error

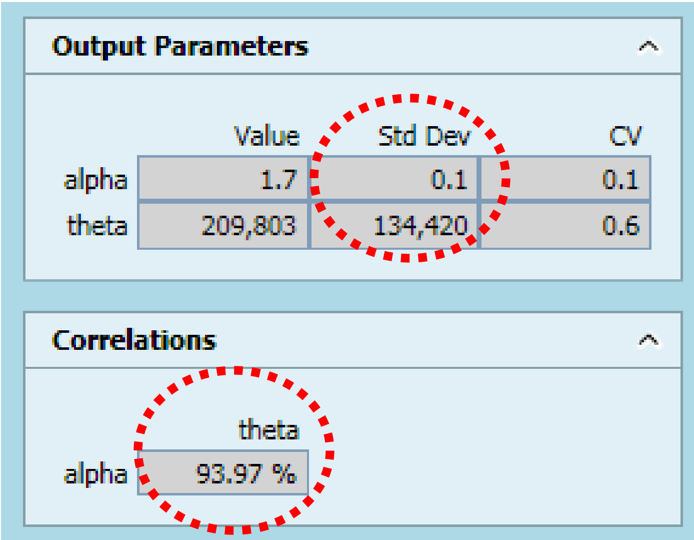
The funnel of uncertainty



How can we deal with this volatility?

Theoretical analysis – Parameter error

Quantifying parameter error



Output Parameters

	Value	Std Dev	CV
alpha	1.7	0.1	0.1
theta	209,803	134,420	0.6

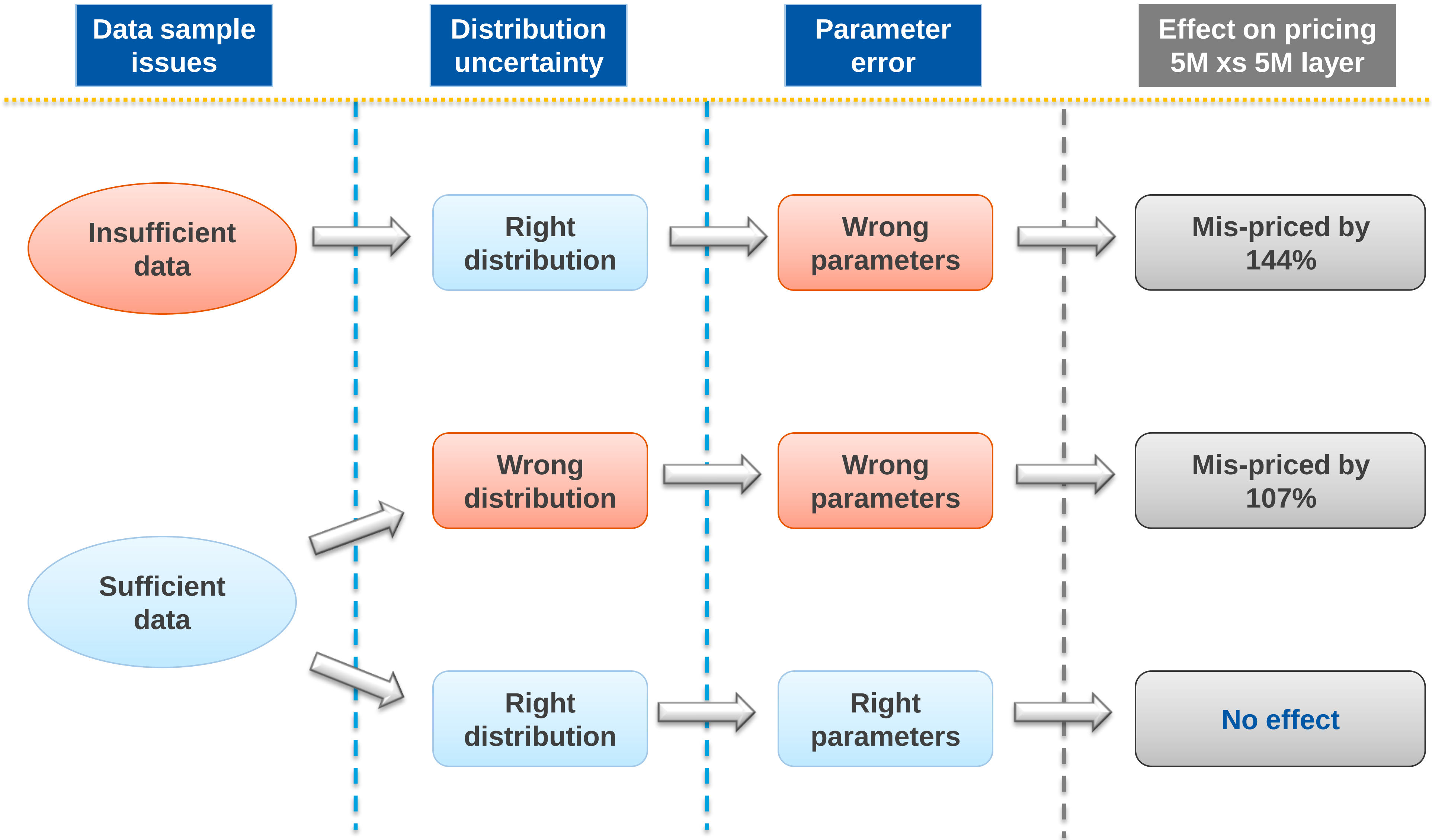
Correlations

	theta
alpha	93.97 %

MetaRisk Fit extract -
Ballasted Pareto, 3000 claims

- Parameter error is effectively measuring sample size error
- Distortion is accentuated in multi-parameter distributions
- Parameter standard deviation and correlation quantifies parameter uncertainties
- We simulate parameters for each run of the model *e.g.*, year of simulation
- We assume a lognormal distribution for parameter uncertainty

Theoretical analysis
Summary



Real-world analysis

UK Motor Market

Real-world analysis
Setting the scene

Case Study: UK Motor Market

- Benchmarking is particularly important in Europe:
  - No industry data collectors such as ISO / NCCI
- Homogenous line of business
- We have access to approximately 60% of motor market data in the UK
- Unlimited reinsurance coverage
  - Not loss limited
  - Low deductibles
- Compulsory line of business

Real-world analysis
Market data statistics

Market data summary statistics	
Number of companies	20
Analysis threshold	£1,700,000
Total number of claims	1,285
Average claim number (per client)	72
Minimum claim number	9
Maximum claim size	£30,235,668
Basis	Report Year
Years selected	2000 – 2007

Real-world analysis

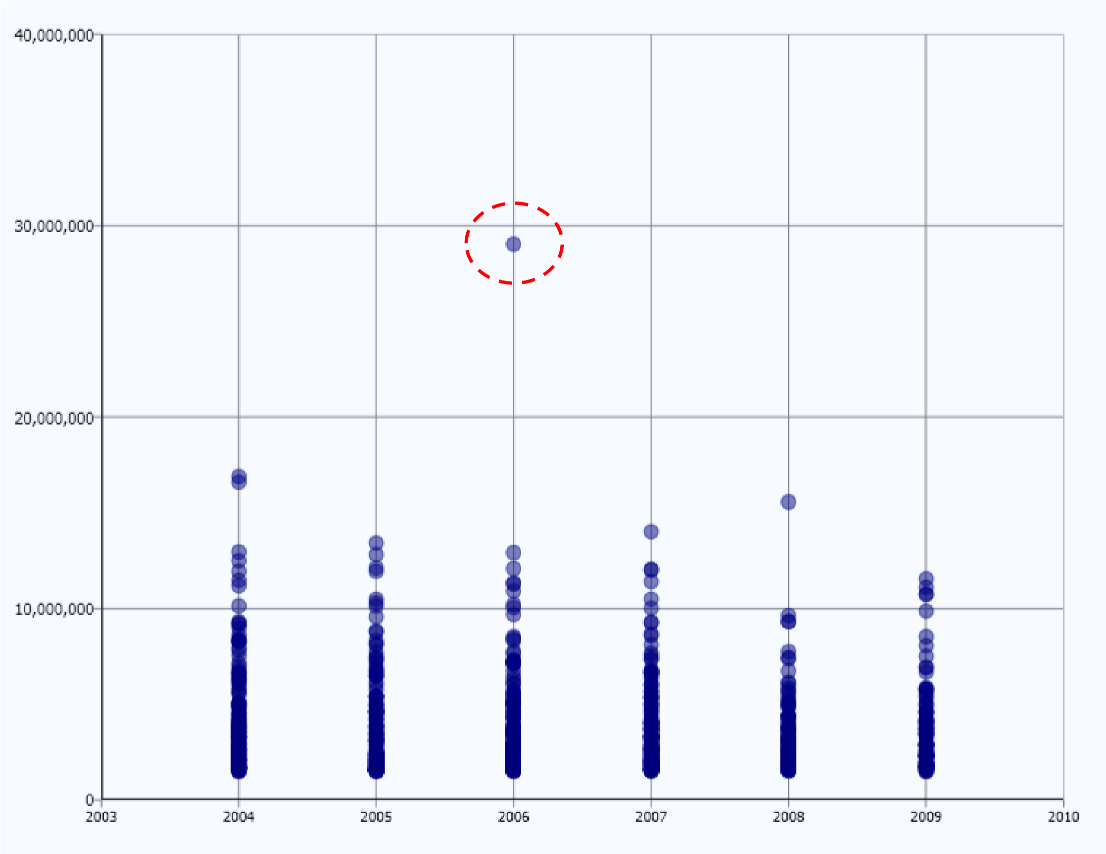
What data to fit to?

- Minimum threshold: £750,000
- Recent years: Uncertainty increases with LDF assumption
- Older years: Uncertainty increases with inflation assumption
- Inflation: 7.5% pa

Report year	1	2	3	4	5	6	7	8	9	10	11
2000	118	241	398	607	799	960	1,113	1,285	1,407	1,536	1,554
2001	131	295	521	731	919	1,092	1,277	1,412	1,548	1,571	
2002	172	407	629	846	1,046	1,243	1,387	1,532	1,556		
2003	253	484	712	936	1,161	1,315	1,470	1,495			
2004	240	480	724	973	1,149	1,319	1,344				
2005	251	513	771	963	1,150	1,175					
2006	279	547	751	960	989						
2007	296	512	745	777							
2008	243	502	536								
2009	298	333									
2010	43										

Real-world analysis
Largest claim effect

- The largest observed claim has a big influence on the fit

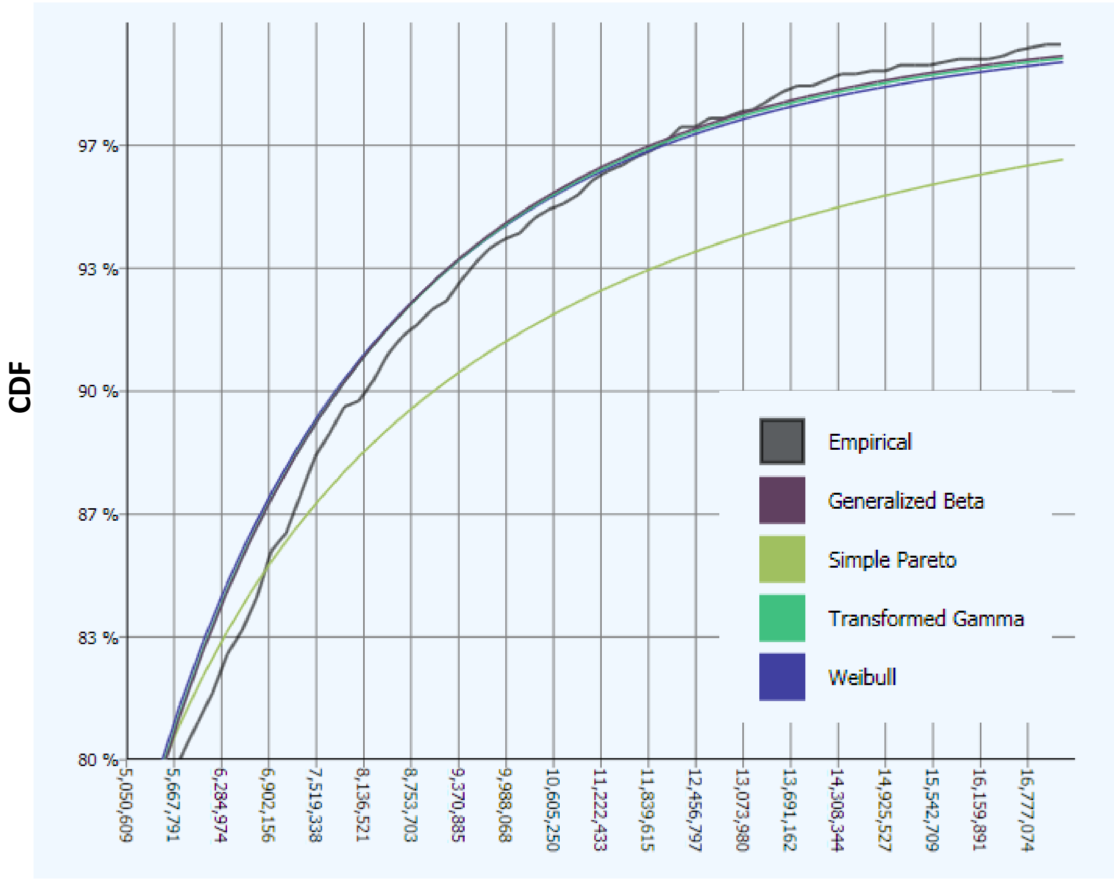


How do we deal with such outliers?

- Remove
- Ignore
- Weighting
- Transform

Real-world analysis

Market empirical vs. possible best fit curves



Distribution	No. of Parameters
Simple Pareto	1
Weibull	2
Transformed Gamma	3
Generalised Beta	4

Real-world analysis
What selection criteria to use?

Mathematical tests

- Goodness-of-fit tests such as:

$$1.\ \text{Natural Log-Likelihood}$$

$$2.\ \text{Akaike} = NLL + K + \frac{K(K+1)}{n-K-1}$$

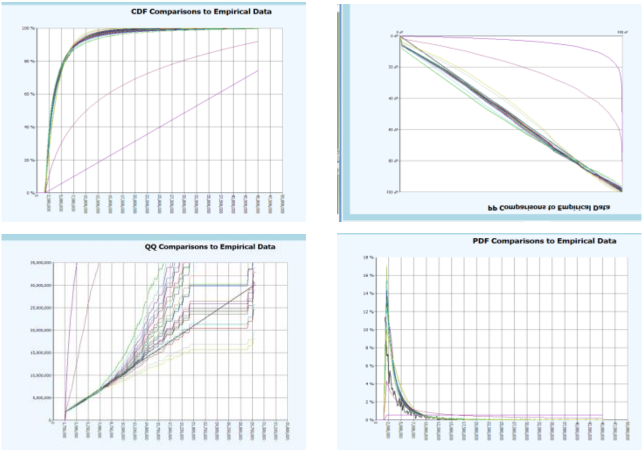
$$3.\ \text{HQ} = NLL + \frac{K.\ln(\ln(n))}{2} \quad for\ n > e$$

$$4.\ \text{Schwartz} = NLL + \frac{K.\ln(n)}{2}$$

$$Where: \text{n} = \text{number of data points}$$

$$\text{K} = \text{number of parameters}$$

By eye – visual judgement



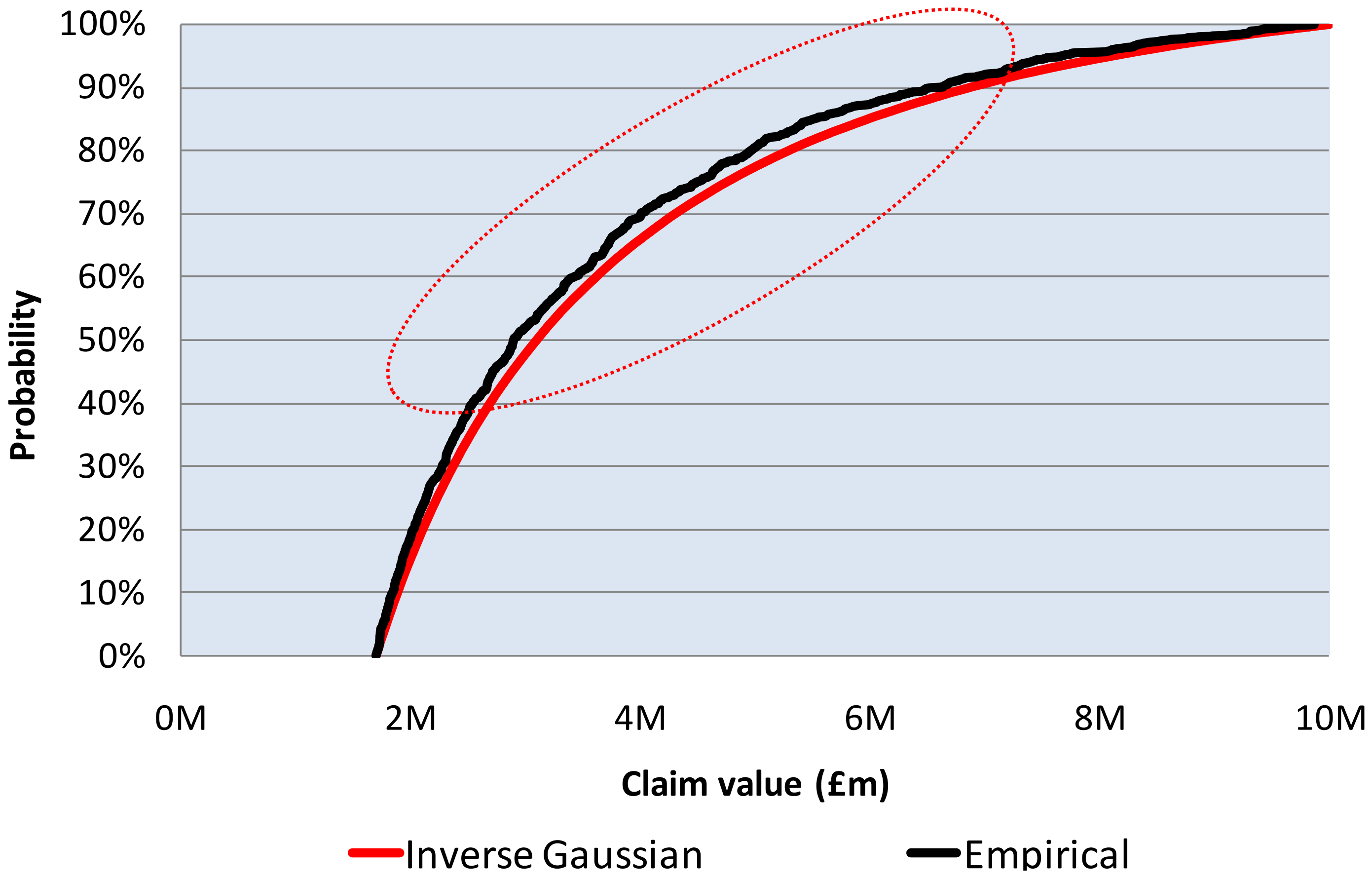
- E.G.,
  - CDF
  - PDF
  - QQ Graph
  - PP Graph

Choosing the market curve
Possible criteria

- Good fit versus over parameterisation
  - Use an information criteria like the H-Q test
- Higher number of parameters may lead to less predictive power
- Parameter CV should be low
- Parameters should be significantly different from zero
- Interpretability of the model and parameters
- Where is the curve going to be used ?

Curve-fitting is subjective; it is an art not a science

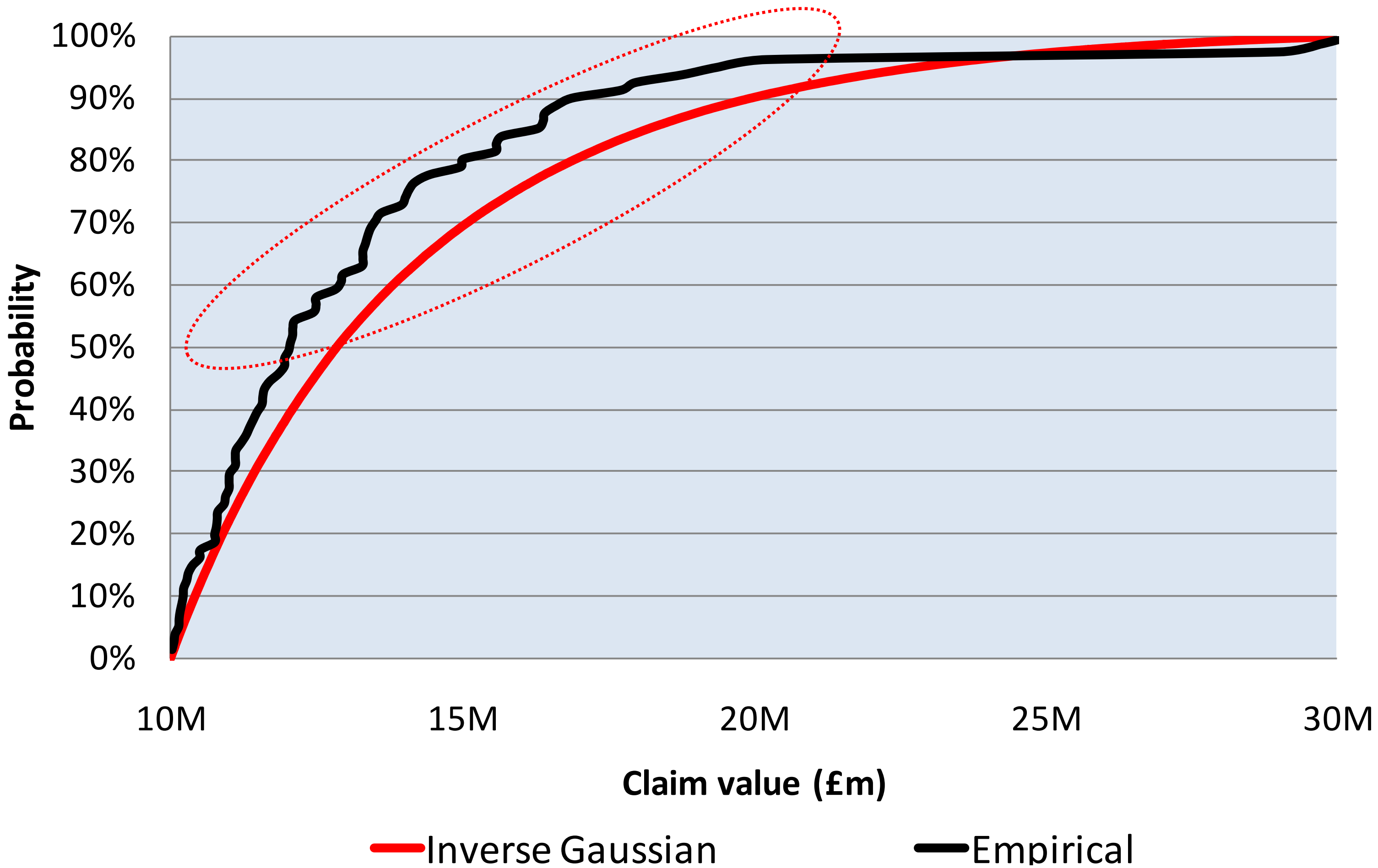
Real-world analysis
What part of curve to fit to?



Inverse Gaussian – good fit to the body of the distribution (0 - £10M)

Real-world analysis

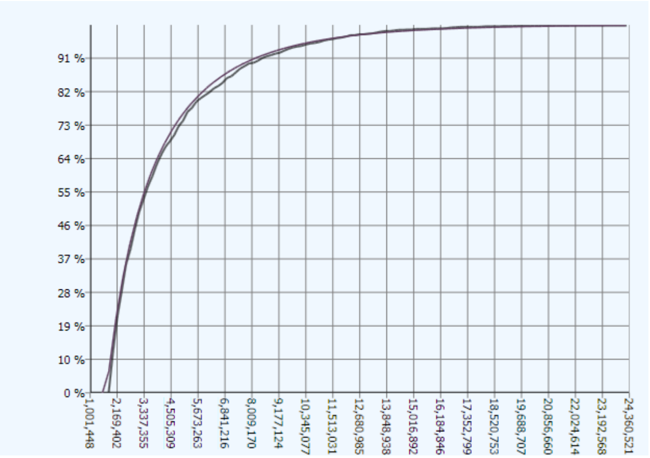
What part of curve to fit to?



Although, the fit is heavier at the tail (£10M - £30M)

Generalised Beta

- Has a good fit when looking at the CDF graph
- Best performing in tests



BUT…

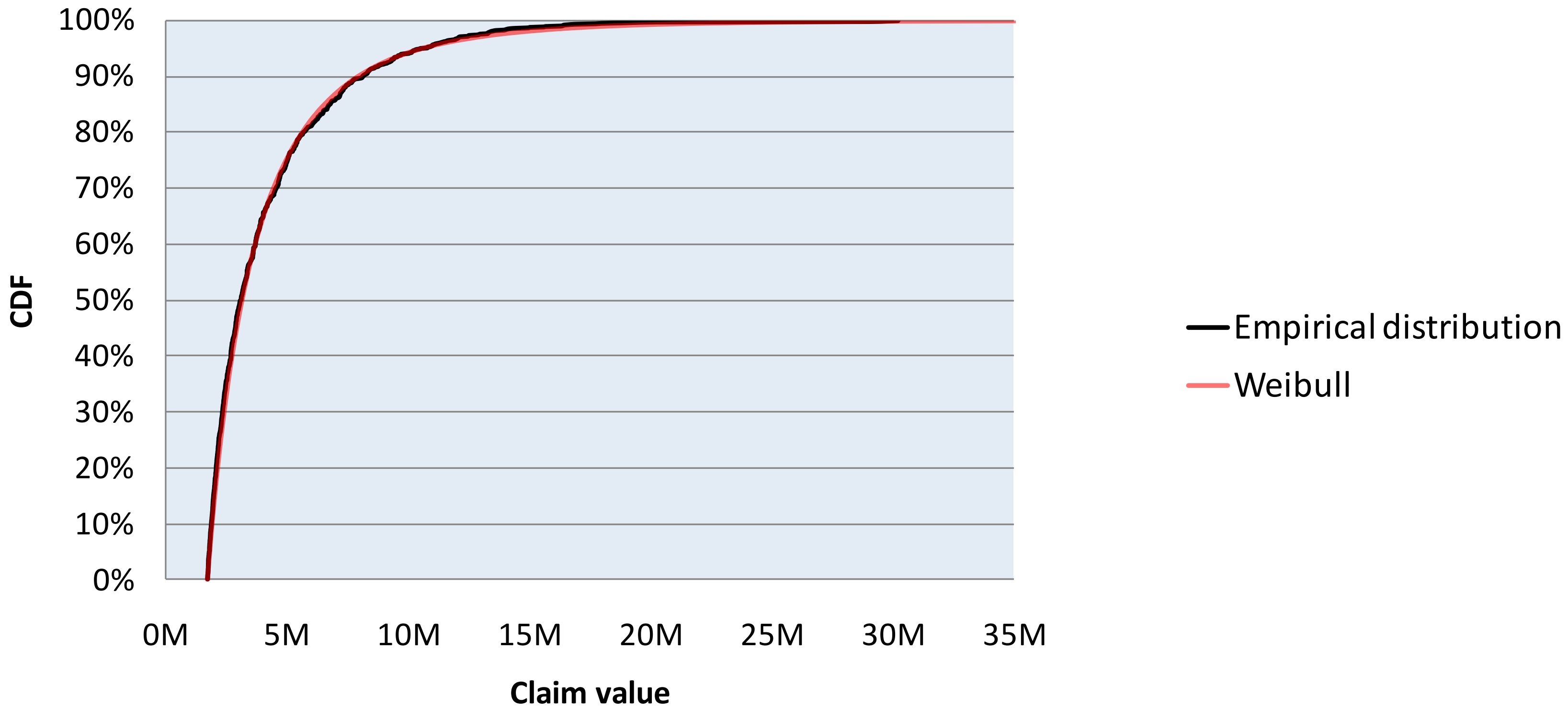
- CVs of parameters are too high
- Beta value is too low

Parameters			
Name	**Value**	**Std Dev**	**CV**
theta	207,219,025	146,990,162	0.71
tau	0.72	0.14	0.20
beta	0.000000114	0.000000083	0.73
eta	22.67	17.13	0.76

Selected Distribution: Weibull

Real-world analysis

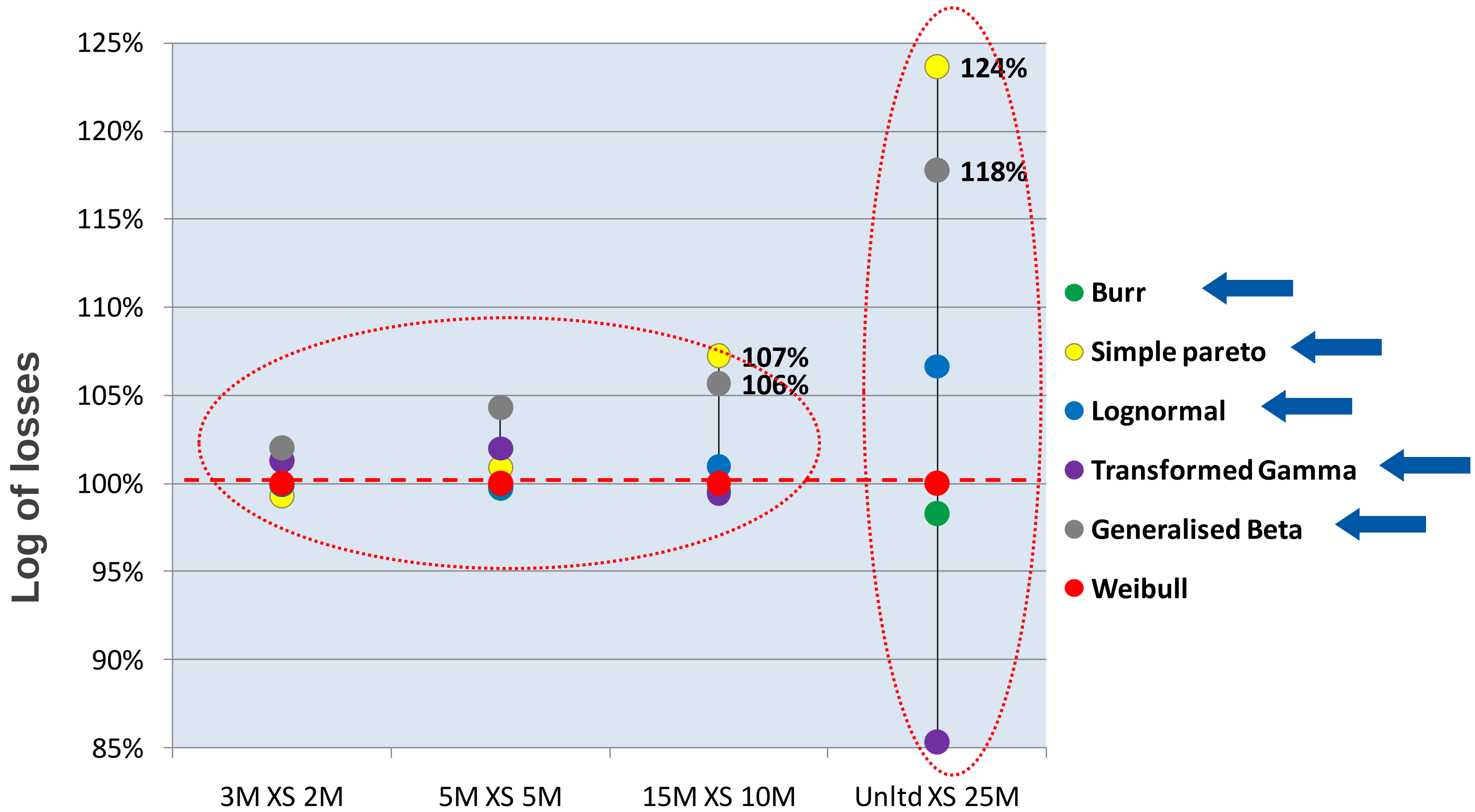
Best fit selected – Weibull distribution



Parameters			
Name	Value	Std Dev	CV
theta	548,690	206,009	0.38
beta	0.53	0.05	0.10

Correlations	
	beta
theta	0.99

Real-world analysis
Effect on the layers



Burr, Lognormal & Transformed Gamma similar to Weibull

Simple Pareto & Generalised Beta: Over-pricing for higher layers

Individual clients' versus market curve

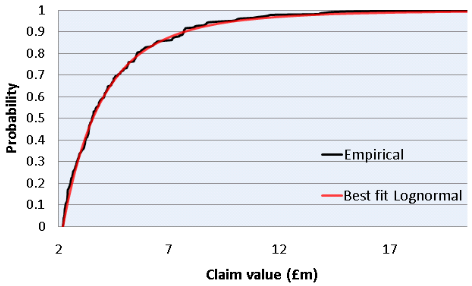
Real-world analysis - Individual clients' vs. Market curve

Individual client data statistics

Attribute	Client A	Client B	Client C	Market
Total number of claims	293	52	12	1,515
Analysis threshold	£1,700,000			
Maximum claim size	£29,731,529	£16,415,791	£12,090,704	£30,235,668
Minimum claim size	£1,709,255	£1,736,425	£1,727,721	£1,702,032
Average claim size	£3,955,290	£4,138,872	£4,853,976	£4,209,709
Basis	Report Year			
Years	2000 - 2007			

Real-world analysis
Client empirical vs. best fit

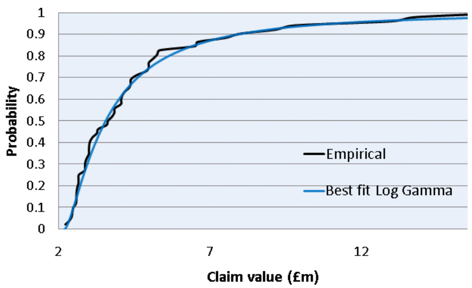
Client A



Parameters			
Name	Value	Std Dev	CV
mu	14.28	0.26	0.02
sigma	0.89	0.11	0.12

Correlations	
	beta
theta	-0.94

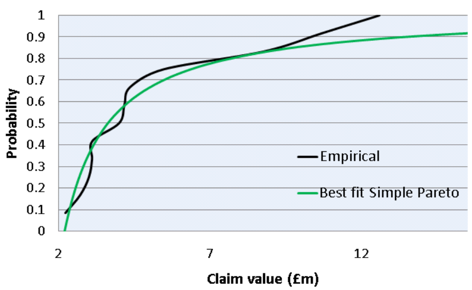
Client B



Parameters			
Name	Value	Std Dev	CV
alpha	2.35	0.51	0.22
tau	1.70	0.31	0.18

Correlations	
	beta
theta	0.87

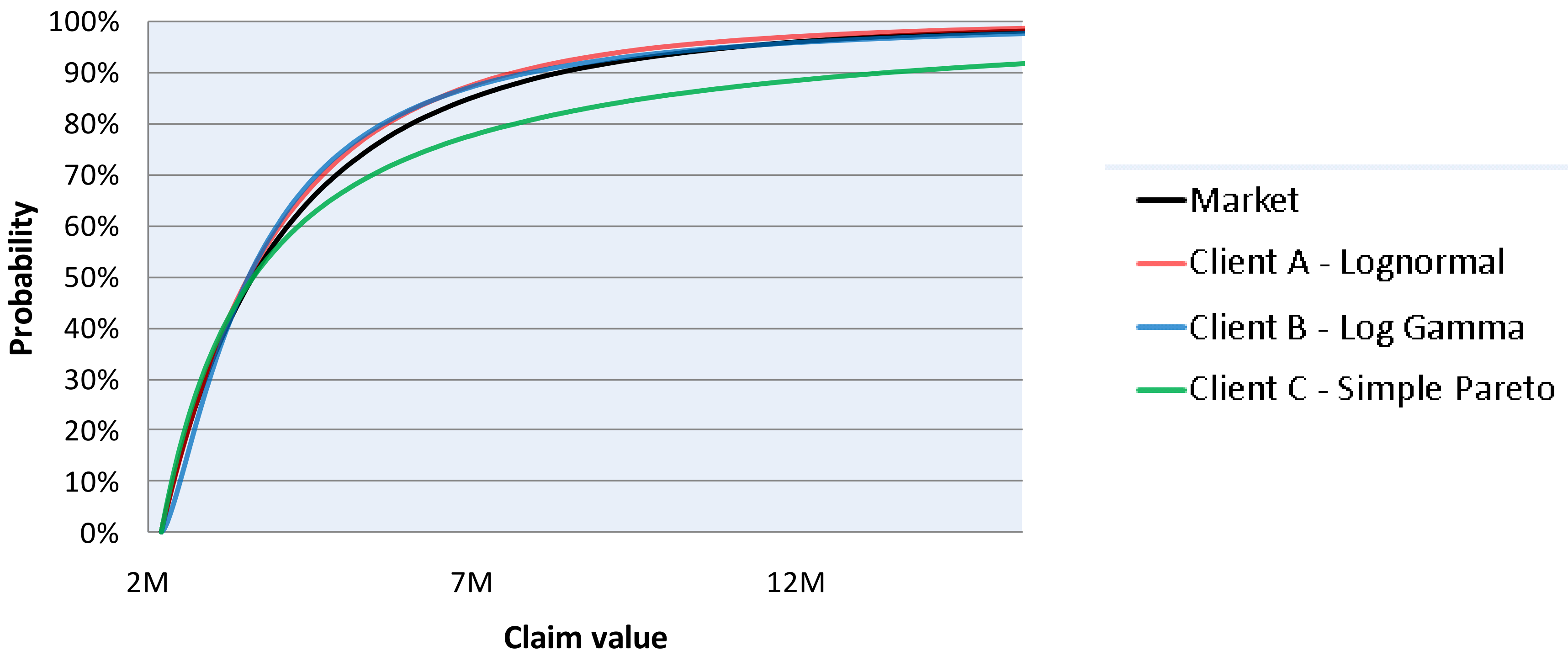
Client C



Parameters			
Name	Value	Std Dev	CV
alpha	1.07	0.38	0.35

Real-world analysis

Market Curve vs. Clients' best fit



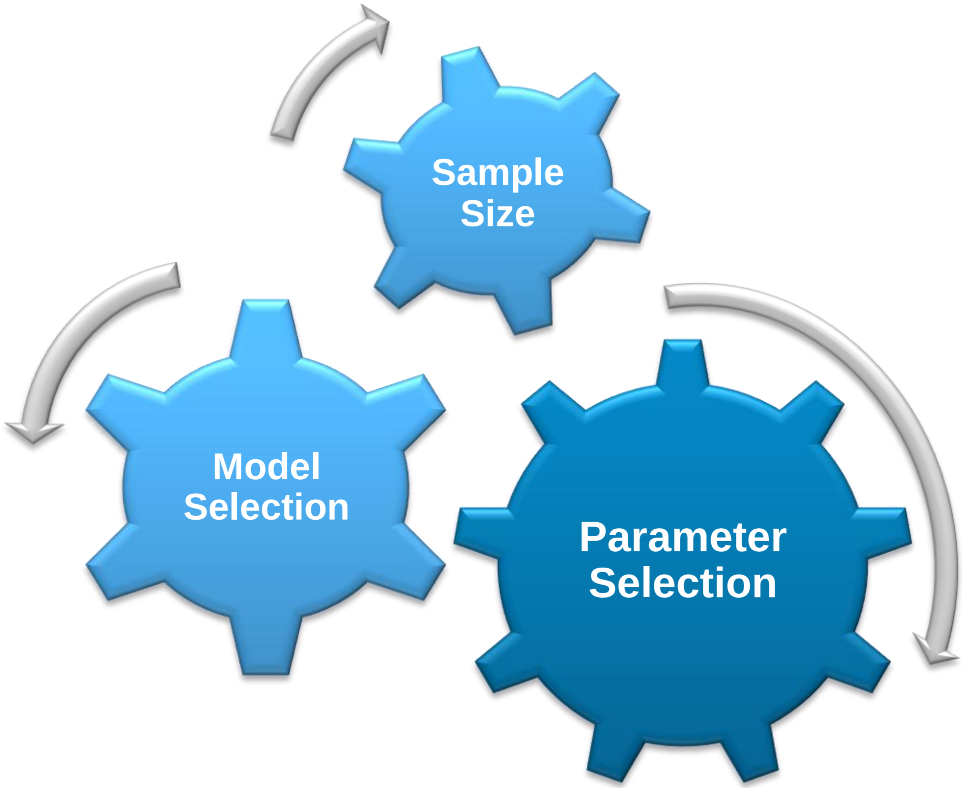
Layers	3M XS 2M	5M XS 5M	15M XS 10M	Unltd XS 25M
Market	100%	100%	100%	100%
Client A	96%	83%	74%	84%
Client B	97%	88%	126%	394%
Client C	103%	161%	437%	2558%

Summary

Summary
Key messages

Bad news: Difficult to 'hide' from the pitfalls of curve fitting

- Multiplicative effect
- Implications where curves are most needed
- Model selection has least impact



Good news:

'Ultimately curve-fitting is where science and art meet'

Any questions?

Important Disclosure

Guy Carpenter & Company, LLC provides this report for general information only. The information and data contained herein is based on sources we believe reliable, but we do not guarantee its accuracy, and it should be understood to be general insurance/reinsurance information only. Guy Carpenter & Company, LLC makes no representations or warranties, express or implied. The information is not intended to be taken as advice with respect to any individual situation and cannot be relied upon as such. Please consult your insurance/reinsurance advisors with respect to individual coverage issues.

Readers are cautioned not to place undue reliance on any calculation or forward-looking statements. Guy Carpenter & Company, LLC undertakes no obligation to update or revise publicly any data, or current or forward-looking statements, whether as a result of new information, research, future events or otherwise. The rating agencies referenced herein reserve the right to modify company ratings at any time.

GUY CARPENTER

Statements concerning tax, accounting or legal matters should be understood to be general observations based solely on our experience as reinsurance brokers and risk consultants and may not be relied upon as tax, accounting or legal advice, which we are not authorized to provide. All such matters should be reviewed with your own qualified advisors in these areas.

www.guycarp.com