GUY CARPENTER

12 October 2011

Pitfalls of Curve Fitting for Large Losses

Guy Carpenter

Amit Parmar, Michael Cane

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**

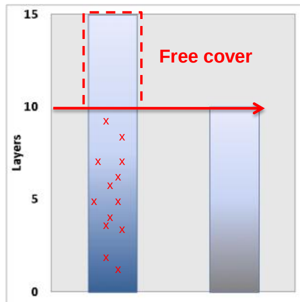
Introduction

Introduction
What is curve fitting used for?

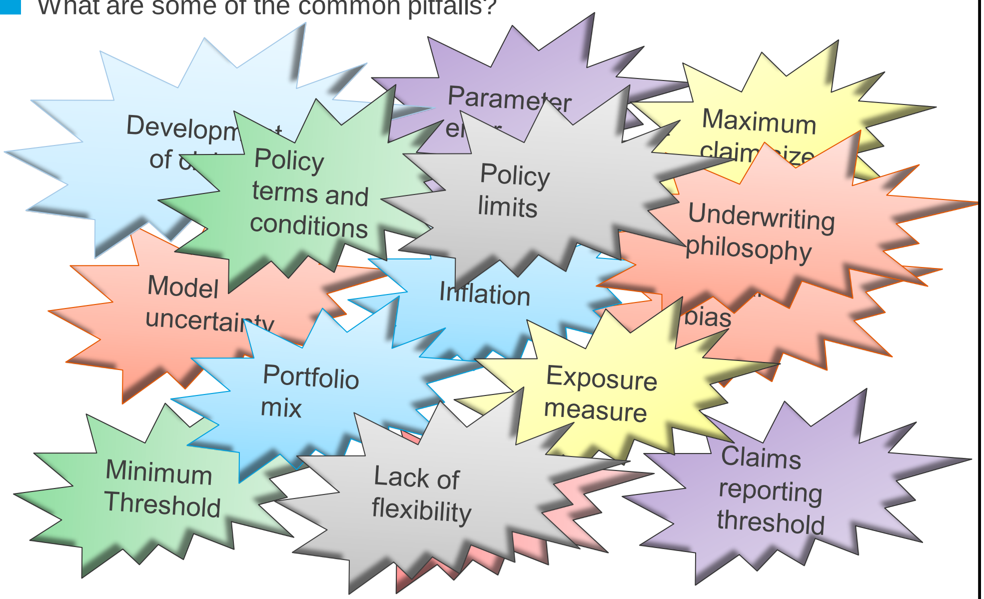
- Understanding the historical data and simplifying data sets
- Modelling where there are few data points
- Understanding potential extremes of the data (via tails)
- Reducing sample variation

Introduction
Why is curve fitting important for actuaries?

- Stochastic modelling
- Benchmarking exercises
- Helps alleviate free-cover problem in experience rating
- Exposure rating may not be possible
- Fundamental input to the capital model
- Advantages to separating the frequency and severity



Introduction
What are some of the common pitfalls?

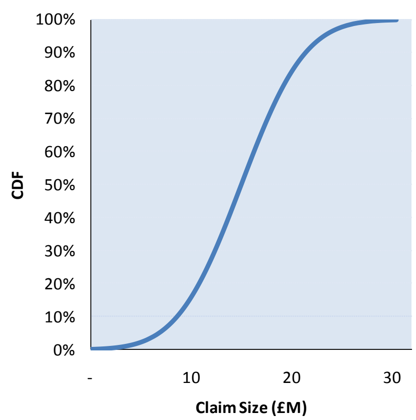


Theoretical analysis

Theoretical analysis

If we sample from:

- A known distribution
- With known parameters



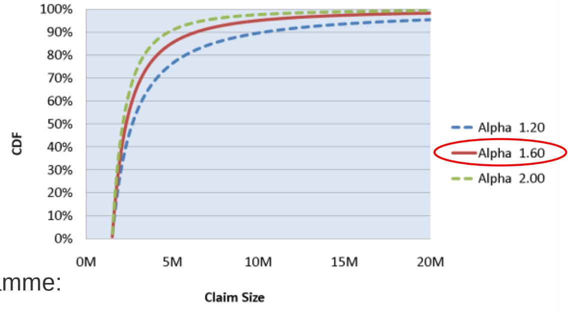
Is it possible to go wrong?

Lets find out…

Theoretical analysis
Our experiment

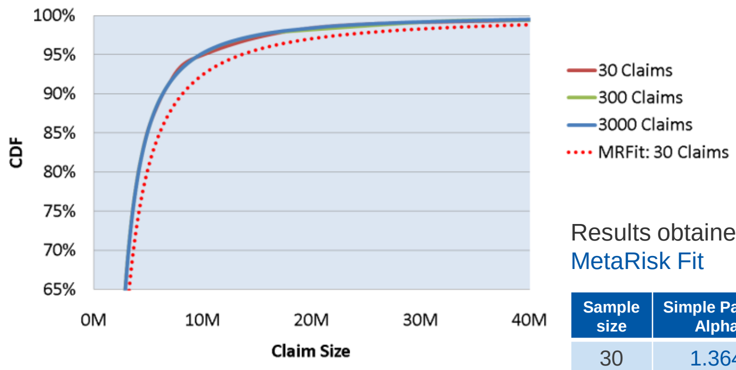
- Sample sizes
  - 30, 300 & 3000 ultimate claim data samples
- Distribution
  - Simple Pareto
- Parameters
  - Alpha = 1.6
  - Lambda = 1,500,000
- Reinsurance structure
  - Common motor programme:

    £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?



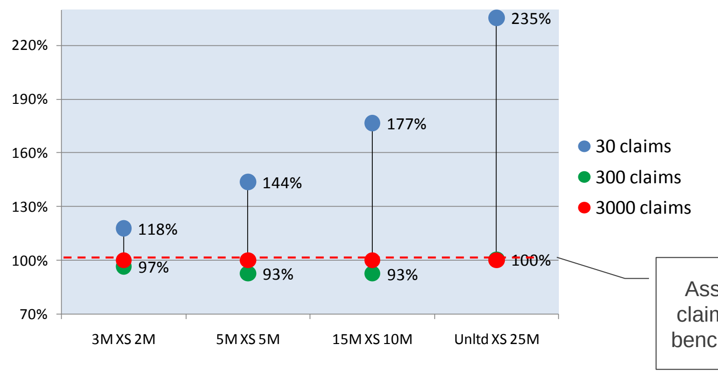
Results obtained using MetaRisk Fit

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?

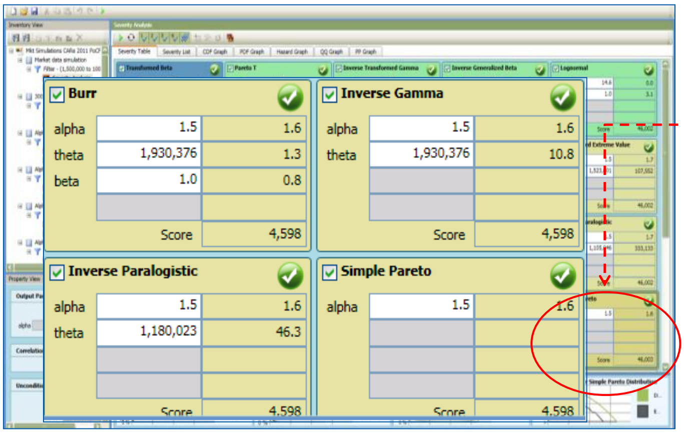
Theoretical analysis – Model uncertainty
Possible severity distributions

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



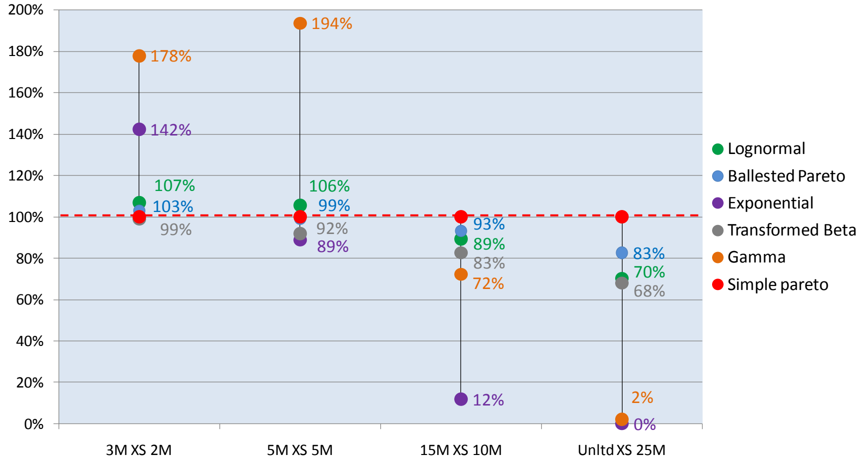
MetaRisk Fit:

Simple Pareto is 1 of the 28 distributions

Theoretical analysis – Model uncertainty

Expected loss to the layer

3000 claims

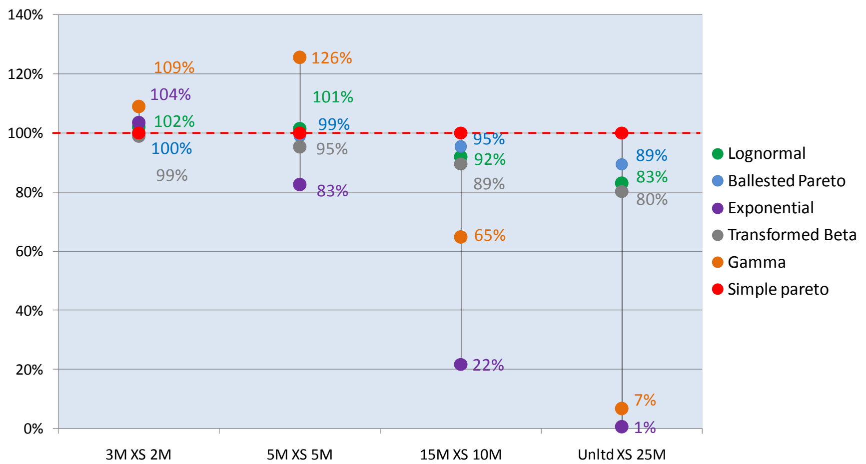


Lognormal: Over-pricing for lower layers; Under-pricing for higher layers

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

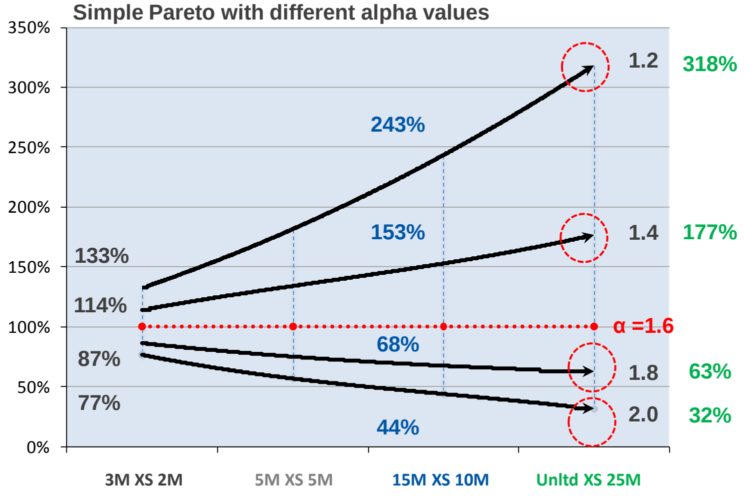
Suppose we have:

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

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

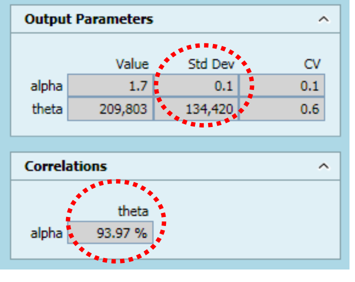
What is the effect on our pricing?

Theoretical analysis – Parameter error
The funnel of uncertainty



How can we deal with this volatility?

Theoretical analysis – Parameter error
Quantifying parameter error



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

Data sample issues	Distribution uncertainty	Parameter error	Effect on pricing 5M xs 5M layer
Insufficient data	Right distribution	Wrong parameters	Mis-priced by 144%
Sufficient data	Wrong distribution	Wrong parameters	Mis-priced by 106%
Sufficient data	Right distribution	Right parameters	No effect

Real-world analysis
UK Motor Market

Real-world analysis
Setting the scene

Case Study: UK Motor Market

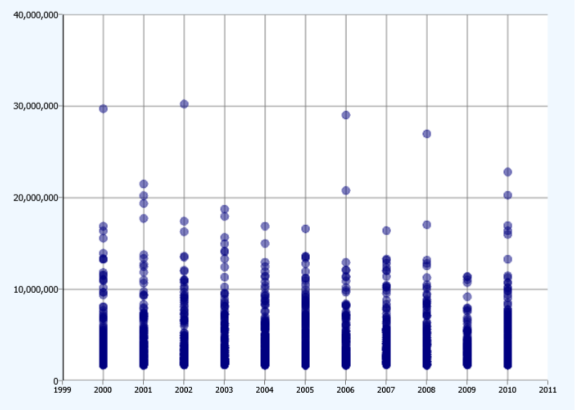
- Benchmarking is particularly important in Europe:
  - No industry data collectors such as ISO / NCCI
- Homogenous line of business
- We have access to approximately 60% of motor market data in the UK
- Unlimited reinsurance coverage
  - Not loss limited
  - Low deductibles
- Compulsory line of business

Real-world analysis
Market data statistics – for developed claims

Market data summary statistics	
Number of companies	21
Analysis threshold	£1,700,000
Total number of claims	1750
Average claim number (per client)	83
Minimum claim number	11
Maximum claim size	£30,235,668
Basis	Report Year
Years selected	2000 – 2010
Inflation	7.5% pa

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
What selection criteria to use?

Mathematical tests

- Goodness-of-fit tests such as:

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

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

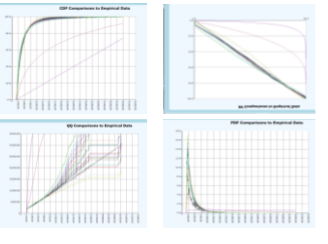
$$3.\ \frac{\text{HQ}}{2} = NLL + \frac{K.\ln(\ln(n))}{2} \quad for\ n > e$$

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

$Where$: n = number of data points

K = number of parameters

By eye – visual judgement



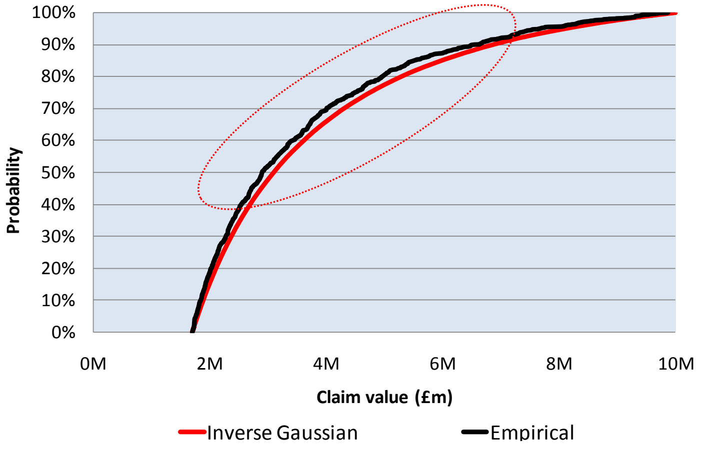
- E.G.,
  - CDF
  - PDF
  - QQ Graph
  - PP Graph

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

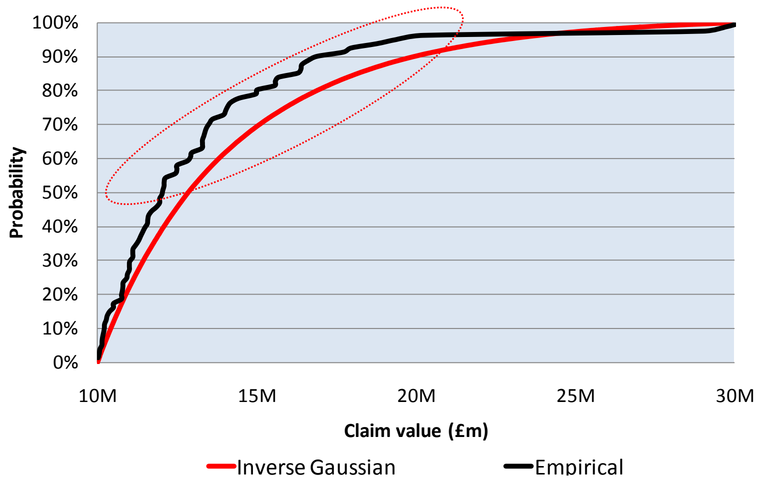
Real-world analysis
What part of curve to fit to?



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

Real-world analysis

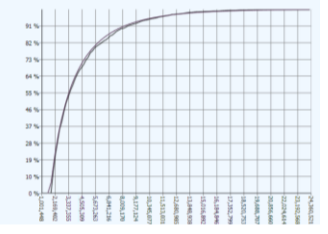
What part of curve to fit to?



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

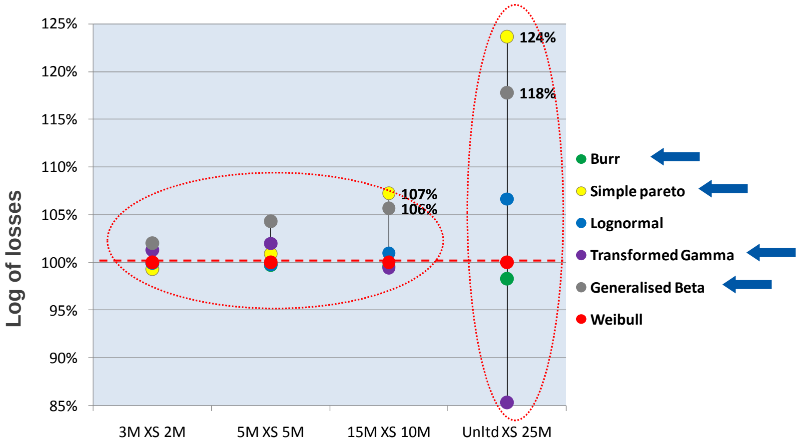


BUT…

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

Selected Distribution: Weibull

Real-world analysis
Effect on the layers



Burr, Lognormal & Transformed Gamma similar to Weibull

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

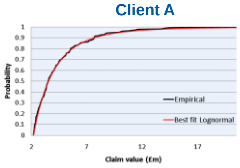
Individual clients' versus market curve

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 - 2010			

Real-world analysis
Client empirical vs. best fit

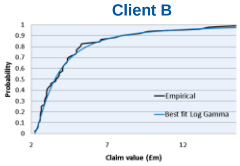
Client A



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

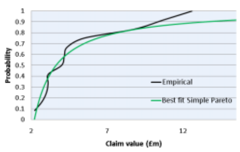
Client B



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

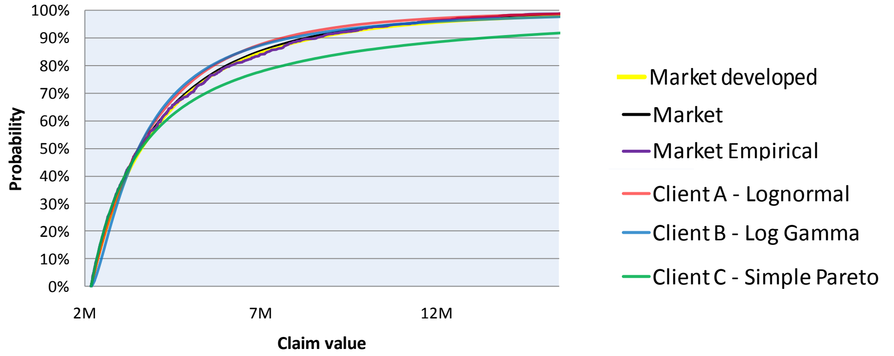
Correlations	
	beta
theta	0.87

Client C



Parameters			
Name	Value	Std Dev	CV
alpha	1.07	0.38	0.35

Real-world analysis
Market Curve vs. Clients' best fit



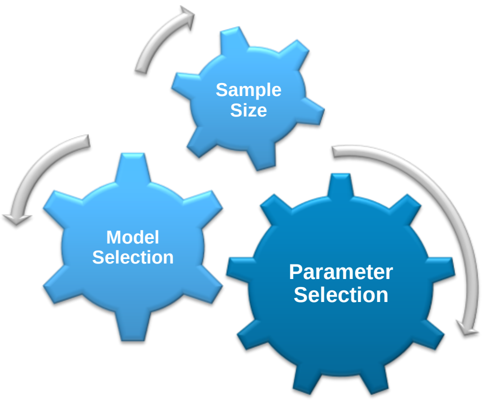
Layers	3M XS 2M	5M XS 5M	15M XS 10M	Unltd XS 25M
Market Dev	100%	100%	100%	100%
Market	99%	95%	86%	65%
Market Empirical	98%	100%	66%	27%
Client A	95%	79%	64%	55%
Client B	96%	83%	109%	256%
Client C	102%	153%	378%	1660%

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?