Pitfalls of Curve Fitting for Large Losses

Guy Carpenter

Amit Parmar, Michael Cane
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
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?

- Development of data
- Policy terms and conditions
- Parameter estimation
- Policy limits
- Portfolio mix
- Inflation
- Minimum threshold
- Sample size
- Claims reporting threshold
- Maximum bias
- Human bias
- Sure sure
- Sure sure
- Rewriting philosophy
Theoretical analysis
Theoretical analysis

If we sample from:
- A known distribution
- With known parameters

Is it possible to go wrong?

Let's 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

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Simple Pareto Alpha</th>
<th>CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>1.364</td>
<td>0.193</td>
</tr>
<tr>
<td>300</td>
<td>1.648</td>
<td>0.059</td>
</tr>
<tr>
<td>3000</td>
<td>1.596</td>
<td>0.019</td>
</tr>
</tbody>
</table>

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

Significantly mis-priced with small data sample

Assume 3000 claims provides benchmark price
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**

<table>
<thead>
<tr>
<th>MetaRisk Fit – Severity distributions</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Simple Pareto</strong></td>
<td><strong>Lognormal</strong></td>
</tr>
<tr>
<td>Extreme Value Limit</td>
<td>Generalized Cauchy</td>
</tr>
<tr>
<td>Exponential</td>
<td>Normal</td>
</tr>
<tr>
<td>Inverse Paralogistic</td>
<td>Uniform</td>
</tr>
<tr>
<td>Loglogistic</td>
<td>Generalized Extreme Value</td>
</tr>
<tr>
<td>Paralogistic</td>
<td>Extremal Pareto</td>
</tr>
<tr>
<td>Loggamma</td>
<td><strong>Ballasted Pareto</strong></td>
</tr>
<tr>
<td>Gamma</td>
<td>Power</td>
</tr>
<tr>
<td>Inverse Weibull</td>
<td>Beta</td>
</tr>
<tr>
<td>Inverse Gaussian</td>
<td>Inverse Beta</td>
</tr>
<tr>
<td>Inverse Gamma</td>
<td>Generalized Pareto</td>
</tr>
</tbody>
</table>

**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
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

Simple Pareto with different alpha values

How can we deal with this volatility?
Theoretical analysis – Parameter error
Quantifying parameter error

- 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

MetaRisk Fit extract - Ballasted Pareto, 3000 claims
Theoretical analysis

Summary

- **Data sample issues**
  - Insufficient data
  - Sufficient data

- **Distribution uncertainty**
  - Right distribution
  - Wrong distribution

- **Parameter error**
  - Wrong parameters
  - Right parameters

- **Effect on pricing 5M xs 5M layer**
  - Mis-priced by 144%
  - Mis-priced by 107%
  - 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
### Market data summary statistics

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of companies</td>
<td>20</td>
</tr>
<tr>
<td>Analysis threshold</td>
<td>£1,700,000</td>
</tr>
<tr>
<td>Total number of claims</td>
<td>1,285</td>
</tr>
<tr>
<td>Average claim number (per client)</td>
<td>72</td>
</tr>
<tr>
<td>Minimum claim number</td>
<td>9</td>
</tr>
<tr>
<td>Maximum claim size</td>
<td>£30,235,668</td>
</tr>
<tr>
<td>Basis</td>
<td>Report Year</td>
</tr>
<tr>
<td>Years selected</td>
<td>2000 – 2007</td>
</tr>
</tbody>
</table>
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
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

<table>
<thead>
<tr>
<th>Distribution</th>
<th>No. of Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple Pareto</td>
<td>1</td>
</tr>
<tr>
<td>Weibull</td>
<td>2</td>
</tr>
<tr>
<td>Transformed Gamma</td>
<td>3</td>
</tr>
<tr>
<td>Generalised Beta</td>
<td>4</td>
</tr>
</tbody>
</table>
Real-world analysis
What selection criteria to use?

Mathematical tests

- Goodness-of-fit tests such as:

  1. Natural Log - Likelihood

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

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

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

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

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Name</th>
<th>Value</th>
<th>Std Dev</th>
<th>CV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>theta</td>
<td>207,219,025</td>
<td>146,990,162</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>tau</td>
<td>0.72</td>
<td>0.14</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>beta</td>
<td>0.0000000114</td>
<td>0.0000000083</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td>eta</td>
<td>22.67</td>
<td>17.13</td>
<td>0.76</td>
</tr>
</tbody>
</table>
Real-world analysis
Best fit selected – Weibull distribution

Parameters

<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
<th>Std Dev</th>
<th>CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>theta</td>
<td>548,690</td>
<td>206,009</td>
<td>0.38</td>
</tr>
<tr>
<td>beta</td>
<td>0.53</td>
<td>0.05</td>
<td>0.10</td>
</tr>
</tbody>
</table>

Correlations

<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>theta</td>
<td>0.99</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Client A</th>
<th>Client B</th>
<th>Client C</th>
<th>Market</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of claims</td>
<td>293</td>
<td>52</td>
<td>12</td>
<td>1,515</td>
</tr>
<tr>
<td>Analysis threshold</td>
<td></td>
<td></td>
<td></td>
<td>£1,700,000</td>
</tr>
<tr>
<td>Maximum claim size</td>
<td>£29,731,529</td>
<td>£16,415,791</td>
<td>£12,090,704</td>
<td>£30,235,668</td>
</tr>
<tr>
<td>Minimum claim size</td>
<td>£1,709,255</td>
<td>£1,736,425</td>
<td>£1,727,721</td>
<td>£1,702,032</td>
</tr>
<tr>
<td>Average claim size</td>
<td>£3,955,290</td>
<td>£4,138,872</td>
<td>£4,853,976</td>
<td>£4,209,709</td>
</tr>
<tr>
<td>Basis</td>
<td></td>
<td></td>
<td></td>
<td>Report Year</td>
</tr>
<tr>
<td>Years</td>
<td></td>
<td></td>
<td></td>
<td>2000 - 2007</td>
</tr>
</tbody>
</table>
Real-world analysis
Client empirical vs. best fit

### Client A

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Name</th>
<th>Value</th>
<th>Std Dev</th>
<th>CV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mu</td>
<td>14.28</td>
<td>0.26</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>sigma</td>
<td>0.89</td>
<td>0.11</td>
<td>0.12</td>
</tr>
</tbody>
</table>

**Correlations**
- beta: -0.94

### Client B

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Name</th>
<th>Value</th>
<th>Std Dev</th>
<th>CV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>alpha</td>
<td>2.35</td>
<td>0.51</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>tau</td>
<td>1.70</td>
<td>0.31</td>
<td>0.18</td>
</tr>
</tbody>
</table>

**Correlations**
- beta: 0.87

### Client C

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Name</th>
<th>Value</th>
<th>Std Dev</th>
<th>CV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>alpha</td>
<td>1.07</td>
<td>0.38</td>
<td>0.35</td>
</tr>
</tbody>
</table>

---

Guy Carpenter
Real-world analysis
Market Curve vs. Clients’ best fit

<table>
<thead>
<tr>
<th>Layers</th>
<th>3M XS 2M</th>
<th>5M XS 5M</th>
<th>15M XS 10M</th>
<th>Unltd XS 25M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Market</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Client A</td>
<td>96%</td>
<td>83%</td>
<td>74%</td>
<td>84%</td>
</tr>
<tr>
<td>Client B</td>
<td>97%</td>
<td>88%</td>
<td>126%</td>
<td>394%</td>
</tr>
<tr>
<td>Client C</td>
<td>103%</td>
<td>161%</td>
<td>437%</td>
<td>2558%</td>
</tr>
</tbody>
</table>

- Market
- Client A - Lognormal
- Client B - Log Gamma
- Client C - Simple Pareto
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.

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.

This document or any portion of the information it contains may not be copied or reproduced in any form without the permission of Guy Carpenter & Company, LLC, except that clients of Guy Carpenter & Company, LLC need not obtain such permission when using this report for their internal purposes.

The trademarks and service marks contained herein are the property of their respective owners.

© 2011 Guy Carpenter & Company, LLC
All Rights Reserved