What does the Bootstrap Trap?
Matt Locke & Andrew D Smith
B02
13:30, Wednesday 21 October 2015
Matt.Locke@UMACS.co.uk
AndrewDSmith8@Deloitte.co.uk
Agenda

• Aims
• What is the bootstrap?
• Back-test results
• Does the bootstrap underestimate extreme percentiles?
• Conclusions
Agenda

• Aims
• What is the bootstrap?
• Back-test results
• Does the bootstrap underestimate extreme percentiles?
• Conclusions
Aims of our investigation

We *are* aiming to:

- Get a better understanding of the strengths and limitations of the over-dispersed Poisson bootstrap (ODPB) as described by England & Verrall (2002)
- Compare the predictive distribution from ODPB against the actual outcomes, using generated data
- Investigate the robustness of the ODP bootstrap’s predictions when the model assumptions are violated
Aims of our investigation

We *are not* aiming to:

- Compare the performance of the ODP bootstrap with that of other mechanical or judgement based methods
  - Bootstrap methods applied to *paid* claims; we do not consider incurred triangles or frequency/severity models here
- Promote or discourage the use of:
  - The ODP bootstrap
  - Other mechanical methods
  - Judgement based methods
Agenda

- Aims
- What is the bootstrap?
- Back-test results
- Does the bootstrap underestimate extreme percentiles?
- Conclusions
Bootstrap in Stochastic Reserving

Delay to payment
1 2 3 4 5 6

Year of origin
2009 2010 2011 2012 2013 2014

Paid claims

Stochastic Calculations

Simulated runoff scenarios
Bootstrap Steps: Parameter Estimates

Parameters include:
- Development patterns
- Ultimate losses
- Variability

Proportion of Claims Paid

Delay (years)
Bootstrap Steps: Forecasting

• Also called “noise” or “process error”
• Simulating one or more future claim scenarios based on estimated parameters (we use gamma distributions here)
• This is a familiar approach for many other risks besides reserving uncertainty
Bootstrap Steps: Back-Casting

- Re-creating hypothetical historical claim scenarios based on estimated parameters
- May use re-sampled residuals (non-parametric) or analytical distributions (parametric)
How the Steps Fit Together

1. Estimate runoff pattern and residuals using basic chain ladder.
2. Back-cast by re-sampling residuals to capture parameter uncertainty.
3. Re-estimate runoff pattern.
4. Stochastic forecasts incorporate both parameter and process error.
Agenda

• Aims
• What is the bootstrap?
• Back-test results
• Does the bootstrap underestimate extreme percentiles?
• Conclusions
Back-Testing

• Compare actual outcomes to a predictive distribution
Back-Testing the ODP Bootstrap

Wait for the outcome

Bootstrap scenarios

x99

12 October 2015
Ranking the Outcomes

• Take 100 future claim scenarios
  • 1 actual outcome • and 99 from bootstraps •
• Sort into increasing order of outstanding claims
• Divide into 10 buckets, each containing 10 observations
• Suppose the actual outcome and the bootstrap are independent samples from the same distribution
• Then there is 1-in-10 chance the red lies in each bucket
The Back-Test

- Bootstrap multiple historical claim triangles
  - Multiple insurers
  - Multiple projection years
  - Or multiple random “realistic” triangles
- Aggregate the bucket counts across bootstraps
- Back-test passes if 10% of actual outcomes lie in each bucket
  - Within random sampling tolerance

Too many actual outcomes lie in the top and bottom bootstrap deciles, so bootstrap distribution too narrow

Different companies / years not independent so we do not know how significant this effect is
The Monte Carlo Back-Test (MCBT)

Stage 1

- One set of stage 2 parameters for each stage 1 back-cast triangle

Stage 2

- Parametric back-cast
- One set of stage 3 parameters for each stage 2 back-cast triangle

Stage 3

- Non-parametric back-cast
- One set of stage 3 parameters for each stage 2 back-cast triangle
Some previous studies focus on the worst 1%.

With 150,000 Monte Carlo observations we can use 100 buckets rather than 10. We know the observations are independent, which implies bars below 0.95 or above 1.05 are significant at 95% confidence.
Generating triangles – simple case

- **SIGNAL**: Assume a base development pattern
  - Use the same pattern for all origin years
- **NOISE**: Incremental claims in each cell generated from a gamma distribution with mean from pattern (with specified gamma vol.)
Generating triangles – simple case

• SIGNAL: Assume a base development pattern
  – Use the same pattern for all origin years

• NOISE: Incremental claims in each cell generated from a gamma distribution with mean from pattern (with specified gamma vol.)
Claims Development Patterns

Proportion of claims paid vs Delay (years)

- ROC08/10 (short)
- ROC08/15 (medium)
- ROC07 (long)
- Taylor/Ashe
- ROC08/20 (extra long)
## MCBT Results: Proportion > 99%-ile

<table>
<thead>
<tr>
<th>Gamma Volatility</th>
<th>Development pattern length</th>
<th>Short</th>
<th>Medium</th>
<th>Long</th>
<th>Extra Long</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td></td>
<td>1.1%</td>
<td>0.7%</td>
<td>0.7%</td>
<td>0.6%</td>
</tr>
<tr>
<td>Medium</td>
<td></td>
<td>1.5%</td>
<td>1.1%</td>
<td>0.8%</td>
<td>1.1%</td>
</tr>
<tr>
<td>High</td>
<td></td>
<td>1.9%</td>
<td>1.5%</td>
<td>1.1%</td>
<td>1.5%</td>
</tr>
<tr>
<td>Extra High</td>
<td></td>
<td>3.0%</td>
<td>2.7%</td>
<td>1.9%</td>
<td>2.7%</td>
</tr>
</tbody>
</table>

12 October 2015
Impact of Tail Length (High Volatility)
Impact of (Gamma) Volatility (Long Tail)

12 October 2015

Low
Medium
High
Extra High
Experiment: Omitting the Back-Cast

How much is the bootstrap adding?

- Perform the Monte Carlo Back Test using stochastic projection of the step 2 fitted parameters
- Therefore have no allowance for parameter uncertainty
Much Better to Bootstrap than Not

It's not conventional to apply bootstrap methods to e.g. Catastrophe models

No bootstrap

With bootstrap
Making Triangles More Realistic

- Scale parameter varies by delay
- Lumpy Distributions
- Change in development pattern by origin year
- Basic Gamma Model
- Payment acceleration / deceleration
- Superposed Inflation
- What else?
Generating triangles – simple case

• SIGNAL: Assume a base development pattern
  – Use the same pattern for all origin years

• NOISE: Incremental claims in each cell generated from a gamma distribution with mean from pattern (with specified gamma vol.)
Generating triangles – simple case

- **SIGNAL**: Assume a base development pattern
  - Use the same pattern for all origin years

- **NOISE**: Incremental claims in each cell generated from a gamma distribution with mean from pattern (with specified gamma vol.)

![Generated data](image-url)
Making Triangles More Realistic

• SIGNAL
  
  – Express the development pattern as hazard rates (compare force of mortality $\mu_x = -\log$ of survival rate $p_x$)
  
  – Transform these hazard rates for each origin year
    
    • Multiply by geometric random walks for origin period and calendar (or equivalently, raise survival rates to a power)
Making Triangles More Realistic

• SIGNAL: example
Making Triangles More Realistic

• SIGNAL: example
Making Triangles More Realistic

- SIGNAL: example

Origin and calendar year transform
Making Triangles More Realistic

- SIGNAL: example

Mean claims - including inflation
Origin Year and Calendar Year Effects

These figures relate to the long development pattern, and high gamma volatility
**Origin Year and Calendar Year Effects**

- **No effect**
- **10% random walk vol pa.** This is very volatile for inflation
- **20% random walk vol pa.** This says claims inflation is as volatile as a stock market. Real data is unlikely to be this volatile but it shows how far the assumptions have to be violated before the bootstrap really breaks down

12 October 2015
Lumpy Claims (ROC 2008, Method B)

Short Tail

Long Tail

Medium Tail

Extra Long Tail
The ROC pure compound Poisson / lognormal model (without inflation or changing patterns) has fatter tails than the gamma forecast distribution. The poor reported tail fit is partly a distribution approximation effect rather than a bootstrap effect.
Agenda

• Aims
• What is the bootstrap?
• Back-test results
• Does the bootstrap underestimate extreme percentiles?
• Conclusions
Extreme Percentile Underestimation

• Given how simple the concept is, the bootstrap does well for most of the distribution

• We replicate results of ROC and others that bootstrap does not perfectly capture extreme tails
  – In some instances the bootstrap distribution is less extreme than reality, and in others it is more extreme
  – We would not expect perfection

• Bootstrap is remarkably robust to moderate assumption violations
Some alternatives to the Bootstrap

• Bootstrap is not the only way to address parameter uncertainty

• Classical methods estimate parameters by maximum likelihood and derive standard errors from the Fisher information matrix

• Bayesian methods (England & Cairns, 2009)

• Single (non-bootstrap) forecast, with the standard deviation multiplied by an “adjustment factor”
  – Use the Monte Carlo Back Test to solve for the adjustment factor
Agenda

• Aims
• What is the bootstrap?
• Back-test results
• Does the bootstrap underestimate extreme percentiles?
• Conclusions
The Importance of Quantitative Testing

<table>
<thead>
<tr>
<th>Mechanical Methods</th>
<th>Rhetorical</th>
<th>Quantitative</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subjective methods</strong></td>
<td>Bootstrap takes account of parameter error&lt;br&gt;Assumptions do not hold in practice&lt;br&gt;Frequentist peg in a Bayesian hole</td>
<td><strong>Prob</strong>{outcome &gt; 99%-ile}&lt;br&gt;Robustness to model mis-specification&lt;br&gt;Unbiasedness / efficiency</td>
</tr>
<tr>
<td></td>
<td>Use underwriting knowledge, common sense, relevant for the board, practical decisions&lt;br&gt;Telling management what they want to hear, profit smoothing.</td>
<td>?</td>
</tr>
</tbody>
</table>

Wanted: outcome-based tests for subjective methods
Points for Discussion (1)

• What to do about known changes (premium cycles, legal developments, settlement processes, inflation etc)?
  – Should we strip them out of the data and put them back into the forecast? Or is that part of the noise we’re trying to measure and extrapolate?
  – The bootstrap allows for these mechanically, but only to the extent that these fluctuations affect the past and the future
Points for Discussion (2)

- Parameter and model errors pervade many risk models: premium risk, cat risk, reserving risk, credit risk, market risk etc.
  - For reserving risk we have the bootstrap. It’s not perfect but we’d give it 8/10 for capturing model and parameter uncertainty
  - For the other risks, we probably ignore model and parameter error, scoring 0/10
  - We can try to perfect the bootstrap, but should we prioritise other risks?
Acknowledgements

• We are grateful for the support from the Managing Uncertainty Qualitatively working party, the Managing Uncertainty with Professionalism working party and from our employers

• Special thanks to Sarah MacDonnell, Tom Wright and Peter England for detailed comments on earlier drafts

• All views expressed and any remaining errors are ours alone
Further Reading


England, P.D. and Verrall, R.J. (2002) Stochastic Claims Reserving in General Insurance (with discussion). British Actuarial Journal, 8, pp 443-544 http://www.actuaries.org.uk/sites/all/files/documents/pdf/sm0201.pdf. Note that this link is to the originally distributed sessional meeting paper. The is a crucial typographical error for the residual adjustment in Appendix 3 which is corrected in the paper finally published in the BAJ, and also in this paper.


Previous two papers available here: http://www.actuaries.org.uk/practice-areas/pages/general-insurance-reserving-oversight-committee-gi roc-0


The views expressed in this presentation are those of invited contributors and not necessarily those of the IFoA. The IFoA do not endorse any of the views stated, nor any claims or representations made in this presentation and accept no responsibility or liability to any person for loss or damage suffered as a consequence of their placing reliance upon any view, claim or representation made in this presentation.

The information and expressions of opinion contained in this publication are not intended to be a comprehensive study, nor to provide actuarial advice or advice of any nature and should not be treated as a substitute for specific advice concerning individual situations. Unpaid volunteers have produced this presentation to promote discussion in the public interest. You may copy, distribute, display and perform the work and make derivative works based on it provided you acknowledge the original authors.