The Actuarial Profession
making financial sense of the future

26 August 2011

Extending the Mack Bootstrap
Hypothesis Testing and Resampling Techniques

Joseph Lo, Aspen*
foreword by Stephen Postlewhite

Foreword

Many general insurance actuaries have a love-hate relationship with the Mack bootstrap. The methodology has usefully provided them with a tool to assess reserve uncertainty that is compatible with the traditional chain ladder method and as such has become a market standard for reserving risk assessment. However, they are confronted regularly with the way that the methodology affords them with little control over the resulting distributions. Moreover, the independence assumptions in the method are not always met, potentially distorting and likely understating risk assessments. The lack of practitioner consideration in this area has been a concern echoed by various regulators recently.

In light of this, we judged it important to find other methods that are more suitable for Aspen's reserving risk modelling. With his responsibility in actuarial research and development, Jo was asked to lead the internal Actuarial and Risk Management effort towards this goal. A collection of methods were subsequently tested and further developed, enriching our internal reserving risk modelling toolbox.

Those methods that are aimed at the independence assumptions of the Mack bootstrap are presented in this paper, helpfully supported by step-by-step demonstrations and impact studies on actual data. I commend this paper to practitioners for furtherance of the profession's debate in this important modelling area.

Stephen Postlewhite, MA,FIA

Head of Risk, Aspen

Acknowledgement

The author is grateful for the assistance in the research and the comments on the paper provided by his colleagues in Aspen's risk management, capital modelling and actuarial teams.

* Joseph Lo, London, U.K.; joseph.lo@aspen.co

Introduction and overview

The Mack bootstrap is now a popular tool for assessing reserve uncertainty among practitioners. In 1993, Thomas Mack published what is now known as the "Mack model" (Mack, 1993). The model focuses on the first two moments of the reserve distribution, allowing the practitioners to derive the prediction error of their chain ladder reserve estimates. Its independence assumptions allow this to be performed without resort to computer simulations. The model also helpfully allows a split of the prediction error into estimation error and forecast error. As well as (Mack, 1993), the reader may also find (England & Verrall, Stochastic claims reserving in general insurance, 2002), (Wüthrich & Merz, 2008) useful references for the Mack model and for wider reserving risk modelling.

Since then, extensions have been made to the Mack models. The key extension is the work done by England and Verrall for the past decade or so. They explored the use of bootstrapping in the Mack model. In so doing, the Mack model is given a place in many stochastic capital models in the insurance industry. The author refers the reader to their 2006 paper (England & Verrall, Predictive distributions of outstanding liabilities in general insurance, 2006) for further details of the Mack bootstrap, its merits against other techniques, practical implementation discussions and its historical development.

The purpose of this paper is first to present how one may perform hypothesis testing on data triangles to see when and how the independence assumptions may not hold. The features that give rise to rejection of the independence assumptions are called *exceptions*. An example of this is a calendar period that has significantly high claim developments across multiple origin periods.

The paper then presents practical techniques to perform bootstrapping to *accommodate* these exceptions. These techniques form a way to extend the Mack bootstrap methodology. When hypothesis testing is done on the extended model, the same features would look less exceptional and more like "business as usual". One of the techniques is known to some practitioners as *sieve resampling* or *partitioned resampling*. Mark Shapland and Jessica Leong have briefly alluded to this as *stratified sampling* in their practical paper on bootstrapping under the Over-dispersed Poisson ("ODP") and GLM frameworks (Shapland & Leong, 2010). The author has not yet seen the second technique in use in the industry, and we shall call this *exception resampling*.

Having dealt with the estimation error, the calendar period driver is introduced to incorporate calendar period exceptions into the forecast error projection. The approach gives primacy to calendar period claim emergence, and then allows secondary dependencies that could be helpfully imposed between origin periods and between calendar period drivers.

Although it is possible to adjust the forecast error framework to give importance to other dimensions (e.g. development period), the calendar period dynamic can be striking and occupy a special place for assessing risk over fixed time horizons. The fixed time horizon modelling is especially urgent for the industry as it prepares for

Solvency II compliance (see, for example, Slide 12 of (Orr & Hawes, 2010) for a suggestion of one-year reserving risk modelling method: the calendar period drivers can contribute to the one-year claim emergence step ["step 3" on the slide] of that suggestion significantly). Origin period and development period nuances can be projected through specific development or variance parameters.

Examples are included to demonstrate the techniques. The recent popularity of publishing loss development triangles has helped researchers to test proposed models with actual data (see for example, (Busse, Mueller, & Dacorogna, 2010)). This paper takes advantage of this, and make use of data from: ACE (ACE Limited, 2010); Arch (Arch Capital Group Ltd., 2010); Axis (Axis Capital Holdings Limited, 2010); and XL (XL Capital Ltd., 2010). For ease of reference, the actual triangles used in this paper are included in the Appendix. It goes without saying that the paper does not attempt to perform full analyses of the reserves or claims emergence of these companies, let alone make any comment on the companies themselves. The contents of the paper must not be relied upon for such analyses or comments. The paper uses the triangles as an aid to demonstrate to the readers how the proposed methods might be adapted for use with their own data, of which they would have a much higher degree of understanding. As a consequence, reserves that are displayed alongside the published triangles should differ from what we have in this paper: after all, it is very rare that companies book reserves direct from a chain ladder model without tail factors!

These example triangles were selected as they were actual data that were publicly available. The author and his colleagues at Aspen initially examined the techniques on their own in-house data: they reached general conclusions that were similar to those in this paper.

In using these examples, we gain an idea of how much of a difference the independence assumptions make. In some cases, this could be significant. A key conclusion is that the practitioner should examine the independence assumptions carefully when applying the Mack bootstrap.

The remainder of the paper is divided into four sections.

- **Definitions and terminology**. This sets up the necessary vocabulary and mathematical symbols for the discussions in the paper. It contains a very broad brush overview of the original Mack model and Mack bootstrapping.

- **The independence assumptions and estimation error**. We discuss how the Mack independence assumptions could be contradicted by some triangles and what the practitioner might want to do about this in relation to estimation error. The section then discusses sieve resampling and exception resampling. Step-by-step implementation examples are presented and impacts on various triangles are discussed. Further generalisations to these techniques are then followed by some further remarks, where possible future next steps and research topics are indicated.

- **The independence assumptions and forecast error**. This is a smaller section discussing a technique of incorporating calendar period exceptions as drivers into the forecast. Again, a step-by-step example is presented. Possible secondary dependencies are examined. Again, it ends with some further remarks, including indications of possible next steps and research topics.

- **Conclusions**. This is a summary of key themes in the paper. It urges the practitioners to continue the debate on how to deal with the failure of independence assumptions when using the Mack bootstrap model, as the extent of underestimation of prediction error could be high.

The bibliography is followed by an appendix of all the triangles used in this paper, along with their development factors, variance parameters and residuals, for ease of reference.

Definitions and terminology

In this paper, we deal with annual origin cohorts and annual development periods. The origin cohorts range from $i = 1$ to $i = I$. The development periods range from $j = 1$ to $j = I$. Calendar years are denoted by k, with the calendar year that had the last full year's worth of data being $k = I$. We do not discuss tail factors in this paper for fear of cluttering mathematical notations. It is possible to extend the concepts and examples in this paper to take account of tails.

Cumulative claims from origin year i and development year j is denoted by $C_{i,j}$. Applications could be made to paid or reported data. It is useful to denote the set of information already available to us (i.e. the "top half" of the triangle), we call this $D = \{C_{i,j} : i + j \leq I + 1\}$.

Link ratios (or development factors) between successive cumulative amounts are defined to be $\lambda_{i,j} = C_{i,j+1} / C_{i,j}$.

The original Mack model

The original Mack method as described in (Mack, 1993) makes the following assumptions that are consistent with the volume-weighted chain ladder assumptions:

1. There are development factors f_j, varying by development period j such that $E(C_{i,j+1} \mid C_{i,1}, \ldots, C_{i,j}) = f_j C_j$
2. For any different origin periods, with $i \neq i'$, $\{C_{i,1}, \ldots, C_{i,I}\}$and $\{C_{i',1}, \ldots, C_{i',I}\}$ are independent
3. There are variance parameters σ_j^2, varying by development period j such that $\mathrm{Var}(C_{i,j+1} \mid C_{i,1}, \ldots, C_{i,j}) = \sigma_j^2 C_j$

The following unbiased estimators (i.e. estimators whose expected values are the parameters to be estimated) are assumed:

4. For f_j, the estimator is the volume-weighted average factor $\widehat{f_j} = \frac{\sum_{i \leq I-j} C_{i,j}\lambda_{i,j}}{\sum_{i \leq I-j} C_{i,j}}$.

5. For σ_j^2, the estimator is given by $\widehat{\sigma_J^2} = \frac{1}{I-j-1}\sum_{i \le I-j} C_{i,j}\left(\lambda_{i,j} - \widehat{f_J}\right)^2$.

The above assumptions are thought to be natural assumptions to make, assuming the volume-weighted chain ladder. In particular, if the variance structure fails Assumption 3, then there would be some other ("better") estimator, linear in the $\lambda_{i,j}$, having a smaller variance than that associated with $\widehat{f_J}$ in Statement 4. From this, we can have estimators for $C_{i,j}$ for $i + j > I + 1$.

6. The estimator given by $\widehat{C_{\iota,J}} = C_{i,I+1-i}.\widehat{f_{I+1-\iota}}.\widehat{f_{I+1-\iota+1}} \ldots \widehat{f_{J-1}}$ is unbiased for $C_{i,j}$, for $i + j > I + 1$.

The unbiasedness of this estimator relies on the estimators, $\widehat{f_J}$'s, being pairwise uncorrelated for distinct development periods.

7. $\text{mse}\left(C_{i,J}\right) := E\left(\left(\widehat{C_{\iota,J}} - C_{i,j}\right)^2 | D\right) = \text{Var}\left(C_{i,J}|D\right) + \left(\widehat{C_{\iota,J}} - E\left(C_{i,J}|D\right)\right)^2$, for each origin period i.

The *mean square error* (mse) of the ultimate can be split up into two components: the variance of the ultimate given the information we currently have, and a measure of the variability of the estimator for the ultimate against the mean ultimate. The first component, $\text{Var}\left(C_{i,J}|D\right)$, is a measure of the *process error* or *forecast error*. Roughly, this is related to the uncertainty driven by natural randomness. The second component, $\left(\widehat{C_{\iota,J}} - E\left(C_{i,J}|D\right)\right)^2$, is a measure of the *estimation error* or *parameter error*. It is related to the uncertainty related to the best estimate itself. Equivalently, in our context, this is the uncertainty related to our development factor estimates due to the limited amounts of data available to us. The standard deviation of the reserves from the *i*th origin period using a Mack bootstrap model would be approximately this mean square error. (See (England & Verrall, 2006) and (Shapland & Leong, 2010) for discussions of the need for adjustments in implementing the Mack bootstrap.)

The Mack bootstrap

The Mack bootstrap adapts the Mack model for simulation under a Monte Carlo framework. England and Verrall's 2006 paper (England & Verrall, 2006) is an important article in this area, and we shall follow it in this subsection. There are two parts to the method – corresponding to the estimation error and the forecast error, as mentioned above.

The estimation error part is taken care of by the traditional statistical technique of bootstrapping, which can be performed on well-defined statistical models. For the Mack model, this consists of:

a. Calculating a set, **R**, of standardised residuals, $r_{i,j}$. These can be roughly interpreted as the deviations of the observed cumulative amounts in the next step, $C_{i,j+1}$, against the expected cumulative amounts in the next step $\widehat{C_{\iota,J+1}}$, standardised by dividing by the standard deviation of the next step. That is to say, $r_{i,j} = \sqrt{\theta_j}\frac{(C_{i,j+1}-C_{i,j}\widehat{f_j})}{\sqrt{\widehat{\sigma_j^2}C_{i,j}}}$. The factor $\sqrt{\theta_j}$ is used so that the standard deviations of the simulated outcomes from Step b below are close to those from the analytical formula. We follow England and Verrall, setting $\theta_j = \frac{I-j}{I-j-1}$. From the point of view of link ratios, $r_{i,j} = \sqrt{\theta_j}.\sqrt{C_{i,j}}\frac{(\lambda_{i,j}-\widehat{f_j})}{\sqrt{\widehat{\sigma_j^2}}}$.

b. Repeatedly resampling **R** for each (i, j) in the upper half of the triangle, allowing replacements to give $r_{i,j}^{\mathrm{B}}$.

c. Backing out *pseudo link ratios*, $\lambda_{i,j}^{\mathrm{B}} = \widehat{f_j} + r_{i,j}^{\mathrm{B}}\frac{\sqrt{\widehat{\sigma_j^2}}}{\sqrt{C_{i,j}}}$.

d. Re-performing the volume-weighted average calculation to obtain a set of bootstrapped development factors, $\widehat{f_j^{\mathrm{B}}} = \sum_{i=1}^{I-j} C_{i,j}\lambda_{i,j}^{\mathrm{B}} \Big/ \sum_{i=1}^{I-j} C_{i,j}$

The forecast distribution is simulated stepwise throughout the bottom half of the triangle. Each cell, $C_{i,j+1}$, in the bottom half of the triangle has an assumed distribution – such as the empirical residual distribution, normal, gamma or lognormal – with mean and variance conditional on the previous value, $C_{i,j}$. Specifically, the conditional mean is $\widehat{f_j^{\mathrm{B}}}C_{i,j}$ and the conditional variance is $\widehat{\sigma_j^2}C_{i,j}$.

The independence assumptions and estimation error

This section considers the independence characteristics of Mack model – Assumptions 1, 2 and 3 – in relation to assessing the estimation error. As we are dealing with estimation error, the proposed extension technique extends Step b of the Mack bootstrap.

Assumption 2 suggests that claims from different origin cohorts develop independently of one another. Assumptions 1 and 3 are saying that, within a cohort, the development at each step depends solely on the latest cumulative figure on that cohort. The three statements translate to the implementation Step b of allowing the standardised residuals from anywhere in the triangle to be applied to anywhere else in the triangle.

However, in some triangles, these three statements are not met, and Step b could then be invalid: and it is important to consider responses to deal with this.

The section is divided into the following subsections. The first three (labelled †) are designed to let the reader assess the concepts without being bogged down in technically involved discussions of the concepts.

- **Examples of exceptional features†**. Real life examples of how the independence assumptions could be broken are presented here, grouped in three dimensions: development period, origin period and calendar period.

Their discussion will be based on how the standardised residuals relate to one another.

- **What should we do with the exceptions in assessing estimation error?†** We briefly discuss the various broad options the practitioner could take when faced with exceptions – when the independence assumptions are contradicted. We acknowledge the validity of different options in different circumstances. At the same time, we indicate the paper is devoted to one of the options: that of *extending* the Mack bootstrap to accommodate the exceptions.

- **High level view of the proposed extensions†**. The two broad concepts discussed in this subsection are meant to be used iteratively to arrive at an extension of the Mack bootstrap that is just far enough to accommodate the exceptions. Hypothesis testing is used to identify exceptions under the various models, original or extended. The extended resampling techniques are used to actually extend the model to accommodate these exceptions.

- **Identification of exceptional features**. The hypothesis testing to identify failure of the independence assumptions is discussed in detail, with a step-by-step example.

- **Sieve resampling**. The sieve resampling technique is discussed with an example.

- **One-step exception resampling of one exceptional feature: theoretical considerations**. The exception resampling discussion is spread over five subsections. The last three (labelled with *****) could be omitted in a first reading. The first introduces a general theoretical framework, with an example of resampling from just one feature.

- **One-step exception resampling of one exceptional feature: examples**. This demonstrates the theoretical considerations with several examples. There is a detailed step-by-step example in this subsection.

- **One-step simultaneous exception resampling***. It is possible to resample more than one exceptional feature simultaneously. This is especially helpful when the exceptional features are from the same dimension (e.g. they are all calendar period exceptions). This subsection indicates how this could be done. An example is presented, extending one from the previous subsection.

- **Small exceptions and parametric bootstrapping***. Exceptional features with only a few residuals could be particularly difficult to deal with – or, indeed, may not be desirable to accommodate. We discuss how assuming the residuals come from a distribution could help their accommodation.

- **Superimpositions of exceptionally resampled residuals***. This is the final of the four subsections that discuss exception resampling in detail. It describes how exception resampling could be superimposed onto previously resampled residuals. A chain of superimpositions could be made to accommodate exceptional features from different dimensions.

- **Further remarks**. A series of concluding remarks and further technical comments are made. Strengths and limitations of the techniques are presented. We also indicate possible next steps and research topics for the interested readers to engage in.

Examples of exceptional features†

In the development period dimension, it is well known that individual large claims could affect the developments through small random fluctuations in the reporting or payment timing. The uncertainty surrounding individual case reserve estimates could give rise to adjustment swings through the development of the claim cohort. These are two examples which can give rise to negative correlations between successive pairs of residuals.

The Axis liability reinsurance incurred triangle gives residuals that are highly negatively correlated (with coefficient -100%) between the second and third development periods.

Negative Correlations between Successive Development Periods

Axis, Liability Reinsurance

Residuals from the Incurred Data

UWY	1	2	3	4	5
2003	220%	52%	-15%	120%	123%
2004	52%	-138%	137%	-117%	-70%
2005	63%	163%	-139%	44%	
2006	6%	-27%	41%		
2007	-30%	-30%			
2008	-63%				

Correlation between the 2nd and 3rd periods: -100%

On the other hand, there are cases where claims develop in "runs" – a large development tends to be followed by another large one, and vice-versa. In this case, there are positive correlations between successive residuals.

The Arch third party occurrence incurred triangle: there is a high correlation (98%) between the residuals of the third and fourth development periods.

Positive Correlations between Successive Development Periods

Arch, 3rd Party Occurrence Insurance

Residuals from the Incurred Data

AY	1	2	3	4	5	6
2002	93%	37%	111%	119%	-32%	128%
2003	-15%	188%	-87%	-130%	-122%	-59%
2004	155%	70%	127%	92%	118%	
2005	-23%	-14%	34%	-18%		
2006	-32%	-94%	-114%			
2007	115%	-97%				
2008	-149%					

Correlation between the 3rd and 4th periods: 98%

The development behaviour of early development can be substantially different from that of late development. Significant positive or negative skewness, for example, may be observed in the residuals of the earlier developments, due to new claims being reported there. Significant negative skewness may be observed in the later developments, if, say, significant subrogations are possible for the class.

The Arch third party claims made paid data have highly skewed residuals, with a skewness in excess of 200%, in the first development period. We observe significant negative skewness of -142% in the first development period of the ACE North American workers' compensation incurred data. Note that the skewness is 0% for a symmetric distribution.

Positive Skewness in a Developemnt Period

Arch, 3rd Party Claims Made Insurance

Reisudals from the Paid Data

AY	1	2	3	4	5	6
2002	48%	5%	-82%	-34%	-12%	-140%
2003	7%	-156%	-81%	184%	156%	22%
2004	-44%	-75%	72%	-1%	-74%	
2005	252%	-57%	-148%	-72%		
2006	10%	163%	99%			
2007	14%	11%				
2008	-43%					

Skewness of the 1st period: 208%

Negative Skewness in a Developemnt Period
ACE, North American Workers' Compensation
Residuals from the Incurred Data

AY	1	2	3	4	5	6	7	8
2000	73%	239%	-104%	-79%	-44%	-109%	-44%	96%
2001	66%	-28%	123%	167%	-81%	-98%	143%	-104%
2002	116%	-86%	-187%	-117%	-45%	103%	-88%	
2003	-127%	-111%	81%	105%	-74%	89%		
2004	-201%	-39%	25%	-35%	185%			
2005	83%	20%	39%	-4%				
2006	47%	4%	19%					
2007	27%	21%						
2008	19%							

Skewness of the 1st period:
-142%

In the origin period direction, development behaviours could vary between cohorts in statistically significant ways. Where narratives and explanations could be found, they are usually in relation to unusual loss events. The World Trade Centre losses from 2001, natural catastrophe losses, the recent subprime / credit crunch events, are such examples. They could give rise to significantly high developments for the year, or significantly *variable* developments when we deal with incurred claim triangles.

The Axis property insurance paid claim residuals from the 2005 AY have a mean of 91%. Note that under the original Mack bootstrap assumptions, we would expect residuals to have a mean of zero.

High Developments along an Origin Period
Axis, Property Insurance
Residuals from the Paid Data

AY	1	2	3	4	5	6
2002	56%	-95%	-75%	-90%	-94%	-127%
2003	63%	-18%	87%	-96%	137%	63%
2004	126%	-181%	-169%	-99%	-49%	
2005	121%	67%	61%	113%		
2006	-58%	82%	67%			
2007	-126%	81%				
2008	-115%					

Mean of the 2005 AY:
91%

The same could be said of the calendar period dimension. Examples of potential narratives and explanations may be significant changes in case reserving philosophies, or emergence of issues in a long-tailed liability book such as changes in legal interpretation or inflation regime changes.

A mean of 122% is observed from the Axis marine insurance incurred 2008 residuals. The XL casualty insurance incurred residuals give a mean of -85% in the 2005 year. Note that the first of the 2008 calendar year residuals starts from the 2007 accident year.

High Developments along a Calendar Period
Axis, Marine Insurance
Residuals from the Incurred Data

AY	1	2	3	4	5	6
2002	-12%	61%	-16%	-58%	-131%	102%
2003	-33%	-31%	172%	156%	112%	-98%
2004	37%	-94%	-31%	88%	9%	
2005	-101%	-86%	40%	-68%		
2006	57%	163%	-132%			
2007	226%	111%				
2008	-55%					

Mean of the 2008 CY: 122%

Low Developments along a Calendar Period
XLI, Casualty Insurance
Residuals from the Incurred Data

AY	1	2	3	4	5	6
2000	120%	-169%	-92%	46%	-119%	-65%
2001	165%	103%	198%	-100%	-14%	118%
2002	-151%	57%	-122%	-22%	165%	-136%
2003	45%	-33%	-74%	198%	-25%	57%
2004	-51%	-19%	30%	-91%	87%	
2005	-46%	-36%	27%	-5%		
2006	46%	11%	9%			
2007	-121%	186%				
2008	-44%					

Mean of the 2005 CY: -85%

What should we do with the exceptions in assessing estimation error?†

Before we launch into the details of the proposed extensions to the Mack bootstrap model for assessing *estimation error*, we ought to evaluate the various options to deal with exceptions that contradict the Mack independence assumptions.

a) **A first option is to acknowledge them but do nothing.** This may be taken if the practitioner believes that the failure of the independence assumptions lead to only immaterial distortions to the risk assessments, giving more weight to simplicity. However, as we shall discuss in the subsections below, the estimation error could be materially underestimated in some cases.

b) **We could take out the exceptions in the modelling of estimation error.** With this option, the exceptions would be regarded as unhelpful in obtaining distributions around the mean parameters, f_j, for projection into the future. They would have more to do with volatility, and hence should be more properly considered in the forecast error. This is not an unreasonable line of thought. However, ideally, one would want the means to be taken as the average of all possible outcomes, and not only on those outcomes that are not volatile: excluding these exceptions could significantly underestimate the volatility of f_j. Moreover, data are precious in reserving risk modelling: excluding residuals could result in losing too much data for resampling.

c) **The Mack bootstrap could be abandoned for this class in preference for another model.** The Mack model assumes independence: if this assumption does not hold, then the model is not valid. This could be theoretically more appealing. With development factors that apply to all accident years, it is not natural for the chain ladder to deal with non-trivial calendar period dynamics. There is a wide variety of triangular approaches in the literature (see, for example, (Wüthrich & Merz, 2008), (Barnett & Zehnwirth, 2000), (Shapland & Leong, 2010), (Martínez Miranda, Nielsen, & Verrall, 2011)). On top of this, non-triangular approaches are being seriously discussed (for a flavour of such methods, see (Orr, A simple multi-state reserving model, 2007) and Chapter 10 of (Wüthrich & Merz, 2008)). The practitioner taking on this approach would likely require different data requirements. For triangular methods, they would also need to deal with cases where there are negative developments. This could also lead them to consider moving away from a simpler approach of dealing with the first two moments which their colleagues would be used to. They could also be considering maintainability of the processes of the different methods if there are many classes of business to consider. All or some of these could make this approach less desirable in a practical context.

d) **We could adjust the data to eliminate the exceptions.** The exceptions are taken seriously and recognised as undermining the validity of the Mack bootstrap model. For each exception, the practitioner would investigate the reasons for the exceptions, and adjust the data accordingly so that the exceptions no longer exist. This could be a reasonable approach if the exceptions could be identified and the exceptions are assured to be "one-offs" (e.g. an origin period exception in 2005 coming from an unusual claims handling process that could not have happened in any other origin periods, due to other years having a different event experience). However, this may not be desirable or possible for all exceptions for all triangles: either because the exceptions are genuinely part of the usual claims development process, or because it could be very time consuming to do.

e) **The Mack bootstrap could be extended so that, in the extended model, the exceptions would no longer appear exceptional.** The extension could be employed to model exceptions that should be considered as part of the usual claim dynamics. For it to be useful, this option needs to avoid the trappings of option (c): the extended model should be similar to the Mack bootstrap. This approach attempts to bring practicality into play, building on a model which is already popular and understood. In the remainder of this section, we shall describe techniques of extending the model. The key issue with this approach is that key statistical properties could be lost. For example, some of the proposed techniques in this paper would force the means of the ultimates to deviate from the chain ladder projections. The practitioner will need to assess the significance of such issues when using the extensions.

The brief discussions above point to the pros and cons with the different options. In the remainder of this section, we shall assume that the practitioner has decided on trying out option (e). We next consider these proposed extensions at a high level, before entering into more detailed discussions on the techniques.

High level view of the proposed extensions†

The example exceptions discussed earlier attempt to link statistically significant features with how claims develop in reality. It is useful to identify these features so as to determine how to proceed in the modelling. On the one hand, it may be the case that claims from unusual events should be isolated and modelled outside the Mack bootstrap. This is likely to be more the case with origin period features. On the other hand, the practitioner may want to leave some of these features in – deeming them part of the usual claim process. In this case, some extension to the Mack model may be required. This section deals with how one may want to construct such extensions.

From now on, we call these features *exceptions*. It is important for our discussion here to recognise that features are only exceptional relative to a model **M**. In our case, one way to do this is to use hypothesis testing – with the null hypothesis being that the observed residuals come from **M**, and the alternative hypothesis being that they do not come from **M**. Where a feature has us rejecting the null hypothesis, we shall call that feature an exception (relative to model **M**).

The general line of attack is to iteratively perform hypothesis testing and extending the model to give a series of models **M(0)**, **M(1)**, ..., with each designed to accommodate the exceptions found in the previous model. The starting point, **M(0)**, is the original Mack bootstrap. From the point of view of understanding the drivers of the statistical process and of communication of modelled results, it is useful for the extended model to still be closely related to the basic structure of independence. We may also see this as an instance of the general aim of statistical modelling: we want to capture essential features, and, at the same time, to avoid overfitting the model.

While the exceptions are identified using hypothesis testing, the extensions are achieved through how resampling of the residuals is performed. One way is what we call *sieve resampling*, in which residuals from an exceptional part of the triangle are constrained to be resampled within that part. This could be used, for example, to accommodate significantly positively skewed residuals in the first development periods of a triangle. The extension thus produced would be a model that reflects significantly different distributional behaviours between different parts of the triangle.

Another way is *exception resampling*. The independent bootstrap would be the norm. However, some simulations would have the exceptional set of residuals (e.g. those from an exceptional calendar period) being applied to a part of the triangle (e.g. to another calendar period). In the example of the calendar period, the extended model attempts to reflect the possibility of *any* calendar period being exceptional. The observed calendar period, that was exceptional in relation to the pre-extended model, would be much more usual in the extended model.

A useful key is that the extended resampling techniques channel their influence to the simulated reserve distributions through the distributions of the development factors, f_j. Exception resampling influences the *dependencies* between the f_j's, while sieve resampling in the development period dimension influences the *shapes* of the f_j's. An immediate consequence of changing the f_j's in this way is that the mean of the projected ultimates can deviate from the chain ladder projections. We do not see this as too much of a problem, since reserving is nowadays rarely done only with the chain ladder method. Scaling is commonly performed so that the Mack bootstrap mean matches with the reserving actuary's best estimate reserves. The practitioner will need to check that the amount of additional scaling required through the use of the extension is tolerable.

Identification of exceptional features

As mentioned above, we define exceptional features (relative to model **M**) to be those over which we reject the null hypothesis that the observed residuals come from **M**. One way to do this is to repeatedly resample residuals assuming **M**, giving rise to a distribution of a test statistic, T. The observed test statistic, t, could then be read off the distribution for one- or two-tailed tests.

If we continue with the example using the Axis marine insurance incurred data, then we see that the 2008 calendar period is exceptional relative to **M(0)**, the original Mack model of independent residuals.

High Developments along a Calendar Period

Axis, Marine Insurance

Residuals from the Incurred Data

AY	1	2	3	4	5	6
2002	-12%	61%	-16%	-58%	-131%	102%
2003	-33%	-31%	172%	156%	112%	-98%
2004	37%	-94%	-31%	88%	9%	
2005	-101%	-86%	40%	-68%		
2006	57%	163%	-132%			
2007	226%	111%				
2008	-55%					

Mean of the 2008 CY: 122%

We now follow the following steps:

i. Define a test statistic. In our case, we consider the 2008 calendar period residuals. A natural statistic is the mean of the 2008 calendar period residuals. Let us call this *T*.

ii. Calculate the observed test statistic, *t* = Average(226%, 163%, …, 102%) = 122%. (As an aside, notice that the first residual from the 2007 AY cohort is actually a 2008 residual. That residual says how claims develop from year end 2007 to year end 2008.)

iii. Produce a large number of resampling of the above triangle of residuals, according to **M(0)**. For each resample, *h* = 1, …, calculate the test statistic $\widetilde{t_h}$. Three resamples might be as follows, with $\widetilde{t_1} = -12\%$, $\widetilde{t_2} = 45\%$ and $\widetilde{t_3} = -36\%$.

Resampled residuals from Axis Marine Insurance Incurred Data (simulation 1)

AY	1	2	3	4	5	6	7	8
2002	172%	88%	112%	61%	-31%	-101%		
2003	-55%	-132%	172%	57%	-98%	-31%		
2004	226%	-58%	102%	226%	-31%			
2005	-101%	88%	-94%	-132%				
2006	57%	9%	226%					
2007	-12%	-16%						
2008	112%							
2009								

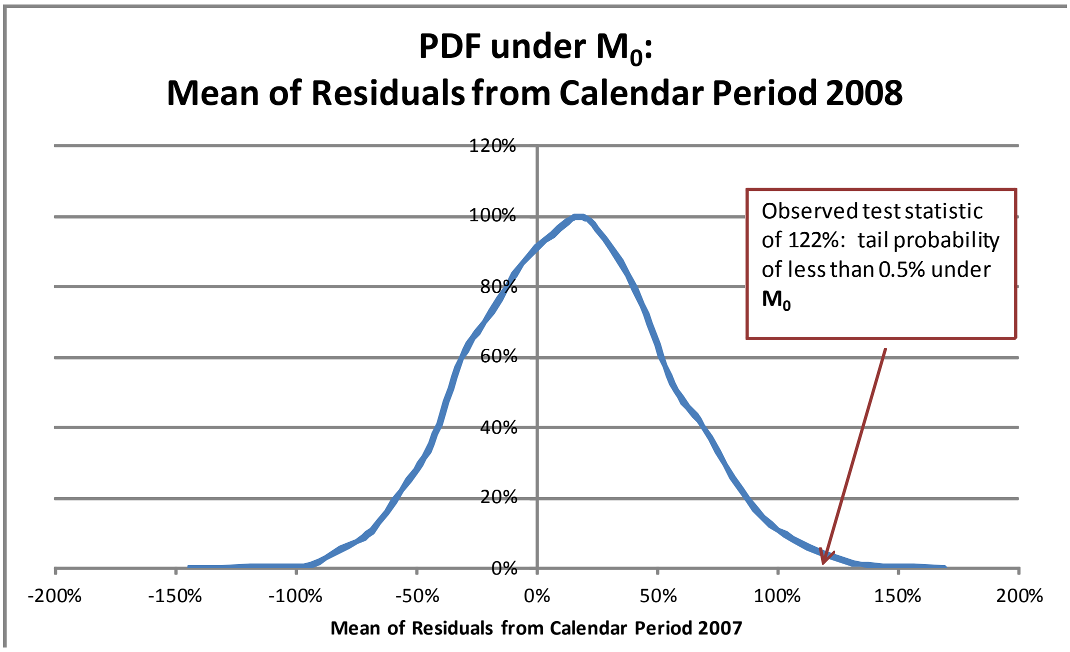
Resampled residuals from Axis Marine Insurance Incurred Data (simulation 2)

AY	1	2	3	4	5	6	7	8
2002	-16%	-33%	-132%	-31%	9%	-68%		
2003	112%	156%	-94%	-12%	-98%	-12%		
2004	37%	163%	-68%	163%	-68%			
2005	156%	-94%	226%	112%				
2006	-31%	57%	102%					
2007	-12%	-131%						
2008	-86%							
2009								

Resampled residuals from Axis Marine Insurance Incurred Data (simulation 3)

AY	1	2	3	4	5	6	7	8
2002	-132%	-98%	9%	57%	-94%	9%		
2003	61%	88%	-31%	57%	9%	57%		
2004	-101%	-58%	-31%	-86%	111%			
2005	40%	-55%	-55%	-98%				
2006	40%	37%	37%					
2007	-132%	226%						
2008	102%							
2009								

iv. Use the resampled test statistics $\widetilde{t_h}$ to approximate a distribution of *T* under **M(0)**. Measure the observed test statistic, *t*, against the distribution.



v. If the tail probability is less than a pre-determined level (traditionally, 5% is often taken to be a boundary that separates out what is random and what is not), then we reject the null hypothesis that the residuals come from **M(0)**. In our case, the tail probability is less than 0.5% for a one-tailed test (or 1% for a two-tailed test). As the tail probability is below 5%, we reject the null hypothesis at the 5% level, and declare that the 2008 calendar period is exceptional in relation to **M(0)**.

The two-tailed test is arguably preferable to the one-tailed test. Using the two-tailed test seems more reasonable, if we are interested in the whole reserve distribution. Identification of calendar periods with exceptionally high or exceptionally low residuals would be useful. However, it is possible, especially purely for solvency purposes, that the practitioner might want to focus on clusters of high residuals. Even here, the identification of clusters of low residuals is useful: since the residuals are centred around zero, a clustering of low residuals must be balanced by a clustering – albeit in a wider area – of high residuals. Of course, in our example above, our conclusion would be the same whether we were performing the one-tailed or two-tailed test.

It is relatively straightforward to program the above to systematically test for features in all three dimensions – or, indeed, for many well-defined statistics. Means, standard deviations and skewness can be used as test statistics. Using this methodology, all examples of exceptional features mentioned previously could be discerned.

We note that Step iii in the above procedure is the same as Step b in the Mack bootstrap. Indeed, since all the proposed extensions (see below) to accommodate the exceptions are versions of resampling, the above procedure could be used to identify exceptional features relative to **M(0)** with the proposed extensions. This point helps to drive the iterative extension of the models described in the "high level" section above.

Moreover, the similarity between Step iii and Step b for **M(0)** and its extensions is helpful in that similar computer codes could be deployed for hypothesis testing and for actual implementation of the extended models.

We end with a note on Type II error of the hypothesis tests. It is understood that there is a 5% chance (in the above example) that a non-exceptional feature is wrongly identified as one. However, it is not clear how to quantify the chance of Type II error – that of non-identification of an exceptional feature. It seems intuitive that this Type II error is less likely for features with more residuals, or where the exceptional features have materially different test statistics. As is the case in this general area of modelling, having large – and applicable – triangle is valuable. While there is advantage in keeping the selected model with as few extensions as possible, not least to avoid overfitting, the practitioner may also use their experience of modelling with other datasets to supplement exceptional features that may not be observed in the triangle. We shall not discuss this point further, and leave it as an open area for future research. (Please also see the Further Remarks subsection below.)

Sieve resampling

The idea here is that we partition the triangle into subsets, and constrain the residuals to resample back into their own subsets.

While in theory it can be used for *any* partitions of the triangle, thus far, we have found them most useful to accommodate significant distributional differences in the development period dimension. It could potentially be useful in the origin period dimension, when there are fixed events that drive the claim developments of particular years. However, in significant cases, the practitioner would usually isolate claims from these events for separate and more transparent analyses: there would then be less need for sieve resampling.

In the case of the ACE North American workers' compensation incurred data, we note that the residuals from the first development period have a significantly non-zero skewness, with a *p*-value of 2%. The key question for sieve resampling is how to partition the triangle into subsets. It is useful to perform hypothesis testing not just on the first development period, but on the first *n*. For the ACE data, we have the following:

p-values of hypothesis testing
Data: ACE North American workers' compensation incurred

Test statistic: Skewness

	Development Period Regions				
Model	1st only	1st to 2nd	1st to 3rd	1st to 4th	1st to 5th
M(0)	2%	84%	39%	54%	98%

This shows that a reasonable partition would consist of two subsets: one with the first development period only, and the other with all the other development periods. If we, say, took the first two development periods, then we would no longer have an exceptional feature, and it would seem more spurious to perform any special resampling at all.

Staying on the subject of selecting a partition, it is interesting also to consider the *p*-values of the remainders of the triangle – that is: all but the first development period, all but the first two periods, and so on.

p- values of hypothesis testing					
Data: ACE North American workers' compensation incurred					
Test statistic: Skewness					
	Development Periods excluding:				
Model	1st only	1st to 2nd	1st to 3rd	1st to 4th	1st to 5th
M(0)	22%	76%	21%	17%	76%

There is no reciprocity. Indeed there is no reason why one would expect significant skewness to exist after the first development period, just because there is significant skewness in the first period. The Mack bootstrap residuals are not related to skewness, although by definition, they have zero mean and unit standard deviation.

There is no material requirement, therefore, for the second subset (the region outside the first development period) to have sieve resampling to simulate its "true" distribution. The sieve resampling comes from the idea that each residual should appear in the triangle with the same frequencies as all other residuals in a set of simulations: that the *overall distribution* (represented by all the residuals) should be the same in the observed world as in the simulated worlds.

Performing the said sieve resampling to construct **M(1)** accommodates the negative skewness (*p*-value is now 80%), and gives the following distribution of reserves. The impact is much less pronounced than seen in exception resampling (see subsections below). A key point is that the exception resampling examples in this paper all increase correlations between the development factors, f_j, driving up estimation error. However, the sieve resampling on development period regions maintains the lack of correlations between the factors. The change here relies mainly (if not only) on the skewness of the first development period, and so only impacts the reserve projection on the last accident year – and only on one development factor.

All AY IBNR Reserve uncertainty Due to Estimation Error (USD 000) Data: ACE NA Workers' Comp Incurred *See text for definition of M(1)*			
	M(0)	**M(1)**	**Difference**
Mean	869,156	861,679	-0.9%
SD	125,026	123,699	-1.1%
Percentiles			
75th	952,615	944,340	-0.9%
90th	1,030,600	1,019,541	-1.1%
99.5th	1,211,111	1,208,322	-0.2%

Performing the same sieve resampling on the Arch third party claims made paid data also yields similar results: estimation error decreases by around 2%. The volatility is further suppressed through the sieve resampling by disallowing the high residual (254%) observed in the first period to be resampled to other periods. Where volatility may be previously driven by this high residual for any of the origin periods save the last one, this is no longer possible.

One-step exception resampling of one exceptional feature: theoretical considerations

The exceptional feature, $\mathcal{E}$, can be thought of carrying two pieces of information, ($\mathcal{L}$, $\mathcal{S}$). One is the location, $\mathcal{L}$, in the triangle. The location is a subset of the triangle – e.g. calendar period, or two adjacent development periods. The other is the structure, $\mathcal{S}$, which relates the residuals in $\mathcal{L}$ with one another. If $\mathcal{L}$ is a calendar period, then examples of $\mathcal{S}$ are "the residuals in $\mathcal{L}$ do not relate to one another" and "the residuals in $\mathcal{L}$ that are in adjacent origin periods are related to one another". The first example is suggesting that $\mathcal{E}$ is exceptional due to a collective feature of the residuals in $\mathcal{L}$, irrespective to their positions relative to one another. Such feature could be the mean or standard deviation of the residuals in $\mathcal{L}$. We will call such a structure *simple*. The second example would be saying that $\mathcal{E}$ derives its exceptionality from how each pair of adjacent residuals in the calendar period $\mathcal{L}$ relate to one another. Such a relationship could be significantly high correlations between the pairs. (The structure, $\mathcal{S}$, could be more rigorously defined to be a set of ordered sets.)

Exception resampling allows exceptional features, $\mathcal{E}$, to occur in a set of target locations, **L**. We assume, for now, that **L** is a partition of the triangle (so that its elements are pairwise disjoint subsets of the triangle, and the union is equal to the triangle). This assumption will be discussed at the end of this subsection, and is not unreasonable, especially when the structure, $\mathcal{S}$, is simple.

If $\mathcal{L}$ is a particular calendar period in the triangle, then **L** could be the set of all calendar periods in the triangle. In this case, the exception resampling could then allow any calendar period in the triangle – even those that are not the same as $\mathcal{L}$ – to have the same exceptional features (e.g. exceptionally high mean or standard deviations, depending on $\mathcal{E}$). In this example, we are suggesting that the claim development dynamics can give rise to clusters of high residuals (or highly variable residuals) along any calendar period – and the observed $\mathcal{L}$ is an example.

We now consider how exception resampling could be implemented under a Monte Carlo set up. Given an exceptional feature, $\mathcal{E}$ = ($\mathcal{L}$, $\mathcal{S}$), with $\mathcal{S}$ being the simple structure, with target locations **L**, an exception resampling could take the following steps in a simulation:

i. For each $\mathcal{L}^*$ in **L**, determine whether it is exceptional: this could be a Bernoulli trial, independent of other members of **L**, with probability *p*

ii. If $\mathcal{L}^*$ is not exceptional, then residuals from outside of $\mathcal{L}$ are resampled into $\mathcal{L}^*$

iii. If $\mathcal{L}^*$ is exceptional, then residuals from $\mathcal{L}$ are resampled into $\mathcal{L}^*$

The probability *p* could be calibrated so that each residual has the same (unconditional) probability of being simulated as any other in any position of the triangle. Clearly, *p* would then need to be $|\mathcal{L}| \big/ \left(\frac{1}{2}(I-1)I-1\right)$, where $|\mathcal{L}|$ is the number of residuals contained in $\mathcal{L}$, and $\frac{1}{2}(I-1)I-1$ is the total number of residuals in the triangle for resampling. (Note that the total number of residuals for resampling is *not* $\frac{1}{2}(I-1)I$ because there is no residual available in development period $I-1$ due to the $\widehat{\sigma^2_{I-1}}$ being undefined.)

We conclude this subsection with a technical note on **L** being a partition of the triangle. Since the singletons are allowed to be elements of **L**, any points of the triangle not in the union of **L** could just be included into a new target as singletons. If the elements of **L** were not disjoint, then we would run into difficulties when deciding whether intersections should follow $\mathcal{E}$ or not in any given simulation. Until some natural rules could be employed to make such decisions, it is easier to require all elements of **L** to be disjoint.

One-step exception resampling of one exceptional feature: examples

We now demonstrate the theoretical discussions with examples from the calendar period, the origin period and then the development period dimensions.

In the Axis marine insurance incurred example, the 2008 ***calendar year*** was identified as exceptional, through its residuals having significantly high mean in relation to **M(0)**. We can therefore set $\mathcal{S}$ to be the simple structure. Assuming any other calendar period could have had similar feature, we define the target set, **L**, to be the set of all calendar periods. Recall that stochastic development factors are obtained after resampling of residuals (see the Mack Bootstrap subsection above). If these stochastic development factors are used to project the triangle, with no further variability imposed, then we would have a measure of the estimation error. The table below shows such distribution for the IBNR: **M(0)** stands for the standard Mack bootstrap, and **M(1)** is **M(0)** but with the aforementioned exception resampling. The number of simulations was 10,000.

All AY IBNR Reserve uncertainty Due to Estimation Error (USD 000) Data: Axis Marine Insurance Incurred *See text for definition of M(1)*			
	M(0)	**M(1)**	**Difference**
Mean	16,910	17,356	2.6%
SD	25,060	35,563	41.9%
Percentiles			
75th	33,580	40,690	21.2%
90th	50,178	67,120	33.8%
99.5th	82,679	116,359	40.7%

We note some observations:

- The reserves derived here are very different from those published by Axis, where the IBNR is booked at over $150m across all accident years. This should not distract us from the current discussion focussed on exception resampling, remembering that it is not best practice to set reserves by using the chain ladder model blindingly, without regards to nuances in the data or commercial considerations.
- The volatility has increased dramatically. An increase is not unexpected. With the calendar period exception resampling, **M(1)** now has non-zero correlations between the different stochastic development factors. We note **M(0)** has zero correlations between the stochastic development factors.
- The size of the increase is large. Intuitively, one expects this from the findings in the hypothesis testing on the 2008 year (see the "Identification of exceptional features" subsection above), with very low *p*-value.
- There is a small change in the mean reserves. This is again due to there being correlations between the stochastic development factors. Recall that the chain ladder projection relies on products of development factors, and that the mean of a product is not necessarily the product of the means if the random variables are not independent. (See also Statement 6 in the discussion of the Mack model earlier in the paper.)

We have discussed calendar period exception resampling above as an example. This could be modified to give ***origin period*** exception resampling, although one should usually be cautious about one-off unusual claim developments. With the Axis property insurance paid as an example, we noticed an exception in the 2005 accident year. The traditional way of isolating unusual claim developments for analysis outside the Mack framework is a good and transparent approach. However, if the practitioner decides that the observed 2005 origin period exception is not a one-off, but could have happened in other years, then performing an exception resampling on that year could be a way to incorporate the feature. An increase of around 2% in the estimation error is seen with exception resampling.

All AY reserve uncertainty Due to Estimation Error (USD 000) Data: Axis Property Insurance Paid *See text for definition of M(1)*			
	M(0)	**M(1)**	**Difference**
Mean	470,387	472,511	0.5%
SD	285,798	291,253	1.9%
Percentiles			
75th	719,187	721,550	0.3%
90th	835,023	841,163	0.7%
99.5th	995,999	1,036,105	4.0%

If we only test on mean and standard deviations, by the definition of $\widehat{f_J}$ and $\widehat{\sigma_J^2}$, one should not obtain single ***development periods*** as an exception with the simple structure. However, there may be a case for exception resampling pairs of adjacent development periods, with $\mathcal{S}$ being the pairs of residuals on the same origin period. The Axis liability reinsurance incurred example had significant negative correlation between the second and third development years. The practitioner now has to decide whether this is a feature that could have happened in any adjacent pair of development periods, or if it is contained in the observed second and third periods. If it is the latter, then the sieve resampling method on pairs of residuals would be helpful (see above). If it is the former, then a way forward would be to use exception resampling. Doing so – now on *pairs* of residuals (namely, (52%, -15%), (-138%, 137%), (163%, -139%) and (-27%, 41%): see the *Examples of exceptional features* subsection above) with targets that are *pairs* of adjacent development periods – would give a decrease of estimation error of around 7%. A similar thought process could be used for the Arch third party occurrence incurred residuals, where we observed a high correlation between the third and fourth periods. If one performs exception resampling to accommodate this observation, then there would be an increase of around 11% in estimation error.

All UWY IBNR Reserve uncertainty Due to Estimation Error (USD 000) Data: Axis Liability Reinsurance Incurred *See text for definition of M(1)*			
	M(0)	**M(1)**	**Difference**
Mean	293,454	293,106	-0.1%
SD	31,120	29,027	-6.7%
Percentiles			
75th	313,900	311,942	-0.6%
90th	334,122	330,648	-1.0%
99.5th	378,884	376,520	-0.6%

All AY IBNR Reserve uncertainty Due to Estimation Error (USD 000) Data: Arch 3rd Party Occ Incurred *See text for definition of M(1)*			
	M(0)	**M(1)**	**Difference**
Mean	722,956	723,122	0.0%
SD	60,943	67,827	11.3%
Percentiles			
75th	764,670	768,101	0.4%
90th	802,245	811,979	1.2%
99.5th	883,359	899,154	1.8%

*One-step simultaneous exception resampling**

The previous subsection has described an approach to performing exception resampling of one exceptional feature (namely, $\mathcal{E}$, onto **L**). If **L** is a partition of the triangle, it is quite easy to extend the above to simultaneously perform exception resampling from more than one exceptional features, $\mathcal{E}_h$ = ($\mathcal{L}_h$, $\mathcal{S}_h$), for h = 1, 2, ... The extension is that for each $\mathcal{L}^*$ in **L** in a simulation, the Monte Carlo algorithm would determine whether it is exceptional, and if it is, with which $\mathcal{E}_h$. This could be done with probability p_h for each $\mathcal{E}_h$, again, independently of other members of **L**. If we further require that each residual has the same probability of being simulated as any other in any position of the triangle, then it would be helpful to impose that the different $\mathcal{L}_h$'s are disjoint. In this case, p_h is $|\mathcal{L}_h| \Big/ \left(\frac{1}{2}(I-1)I-1\right)$.

An example would be two or more calendar period being exceptionally resampled onto all the calendar periods. We illustrate this by considering the XL casualty insurance incurred data. Here the residuals from the 2005 year have a significantly low mean of -85%, and those from the 2006 year have a significantly low standard deviation of 25%. Note that, under **M(0)**, we would expect means of residuals to be around 0% and standard deviations of six residuals to be around 1/sqrt(6-1) = 45%.

Low Developments along a Calendar Period; low Variability along another

XLI, Casualty Insurance

Residuals from the Incurred Data

AY	1	2	3	4	5	6	7	8
2000	120%	-169%	-92%	46%	-119%	-65%	-106%	98%
2001	165%	103%	198%	-100%	-14%	118%	133%	-102%
2002	-151%	57%	-122%	-22%	165%	-136%	-35%	
2003	45%	-33%	-74%	198%	-25%	57%		
2004	-51%	-19%	30%	-91%	87%			
2005	-46%	-36%	27%	-5%				
2006	46%	11%	9%					
2007	-121%	186%						
2008	-44%							

CY	2005	2006
Mean	-85%	-40%
SD	41%	25%

Let **M(1)** be the model produced from performing exception resampling with the 2005 calendar period on all calendar periods, based on **M(0)**. Under **M(1)**, it can be seen that the 2006 calendar period is also an exception, in that its standard deviation is significantly low with *p*-value of around 2%. This clustering of low volatility residuals would imply a balancing clustering of high volatility residuals outside it. We may also, then, want to perform exception resampling with the 2006 calendar period as well as with the 2005 year. This we shall call **M(2)**. We tabulate statistics of the IBNR reserve distributions from the extended models below.

All AY IBNR Reserve uncertainty Due to Estimation Error (USD 000) Data: XL Casualty Insurance Incurred *See text for definition of M(1) and M(2)*				Differences	
	M(0)	**M(1)**	**M(2)**	**M(1):M(0)**	**M(2):M(1)**
Mean	1,048,807	1,051,043	1,052,919	0.2%	0.2%
SD	285,075	312,350	328,777	9.6%	5.3%
Percentiles					
75th	1,240,258	1,265,652	1,275,012	2.0%	0.7%
90th	1,426,201	1,463,335	1,497,918	2.6%	2.4%
99.5th	1,820,165	1,871,175	1,922,043	2.8%	2.7%

The difference between **M(2)** outputs and those of **M(1)** is smaller than that between **M(1)** and **M(0)**. This is intuitively not surprising. Clustering of high residuals, characterised by exceptions with high / low means, should have more of an impact in driving the correlations between the f_j's than clustering of highly *volatile* residuals, characterised by exceptions with high / low standard deviations.

*Small exceptions and parametric bootstrapping**

The discussions above are centred on non-parametric bootstrap. The distributions assumed for the residuals is the empirical observed distributions. However, one can also perform resampling by assuming that the residuals follow a family of parametric distributions (e.g. the normal distribution). This can be especially useful when $|\mathcal{L}|$ is small, for two reasons.

Firstly, due to the method of sampling *with* replacement, the accommodation of exceptional features with small locations could be hampered: an early calendar period that has exceptionally high standard deviation could require the whole range of residuals to be resampled into it to achieve a similar level of standard deviation.

This is the case for the 2002 calendar year for the XL casualty insurance incurred triangle. Under **M(2)** (see above), this calendar year still has a *p*-value of under 2% (under a two-tailed test on the standard deviation).

Low Developments on a Calendar Period; Low Variability on another; High Variability on a third
XLI, Casualty Insurance
Residuals from the Incurred Data

AY	1	2	3	4	5	6	7	8
2000	120%	-169%	-92%	46%	-119%	-65%	-106%	98%
2001	165%	103%	198%	-100%	-14%	118%	133%	-102%
2002	-151%	57%	-122%	-22%	165%	-136%	-35%	
2003	45%	-33%	-74%	198%	-25%	57%		
2004	-51%	-19%	30%	-91%	87%			
2005	-46%	-36%	27%	-5%				
2006	46%	11%	9%					
2007	-121%	186%						
2008	-44%							

CY	2005	2006	2002
Mean	-85%	-40%	-2%
SD	41%	25%	237%

A first try of accommodating the 2002 year would be **M(3)**, where exception resampling would be performed (not yet with parametric bootstrapping). Below is a table of *p*-values from the means and standard deviations of the residuals relative to each of **M(0)**, **M(1)**, **M(2)** and **M(3)**. It shows that under **M(3)**, the 2002 year is still not accommodated, with no improvement in the *p*-value.

p-values of hypothesis testing
Data: XL Casualty Insurance Incurred
See text for definition of M(1), M(2) and M(3)

(i) Test Statistic: Mean

CY	M(0)	M(1)	M(2)	M(3)
2002	98%	97%	94%	94%
2005	4%	13%	13%	14%
2006	31%	40%	49%	49%

(ii) Test Statistic: Standard Deviation

CY	M(0)	M(1)	M(2)	M(3)
2002	1%	1%	2%	1%
2005	7%	22%	48%	50%
2006	0%	2%	20%	20%

Secondly, significance is not the issue when deciding how to accommodate it. A better question is how confident we are in saying that these two residuals can be resampled onto larger calendar periods – and thus ordaining that the exception distribution consists of two points, each with a 50% weight. The focus is on how to accommodate the exception. This is where parametric bootstrapping of the exception is also useful, by providing a fuller distribution.

The procedure would be to first calibrate a reference distribution for each exception. Then in the simulation, whenever a resampling is required from an exception, these corresponding reference distributions would be called upon to return a residual rather than the observed set of residuals.

For the XL casualty insurance incurred triangle, one could calibrate a normal distribution for each of the 2002, 2005 and 2006 calendar periods. The 2002 mean and standard deviation parameters would be (-2%, 237%), for example. Let us set **M(4)** to be the same as **M(3)**, but with parametric bootstrapping using the normal distribution. It can be seen that the 2002 year has a higher *p*-value relative to **M(4)** in standard deviation – higher than 5%. We have the following results for the IBNR distributions. As one would expect, the variability increases, as the exceptions no longer have only discrete points to simulate from.

All AY IBNR Reserve uncertainty					
Due to Estimation Error (USD 000)					
Data: XL Casualty Insurance Incurred					
See text for definition of M(2), M(3) and M(4)				**Differences**	
	M(2)	**M(3)**	**M(4)**	**M(3):M(2)**	**M(4):M(3)**
Mean	1,052,919	1,054,034	1,052,691	0.1%	-0.1%
SD	328,777	329,457	349,382	0.2%	6.0%
Percentiles					
75th	1,275,012	1,278,864	1,279,293	0.3%	0.0%
90th	1,497,918	1,490,349	1,511,435	-0.5%	1.4%
99.5th	1,922,043	1,941,363	2,035,217	1.0%	4.8%

Occasionally, the parametric resampling approach could have us producing unrealistically large negative / positive residuals. Capping of residuals could be helpful: such as demanding that residuals be within ±3 standard deviations, say (a value of 3 corresponds to a cumulative probability of more than 99.8% in a standard normal distribution).

We note that it is reasonable for a practitioner not to accommodate every observable exception using exception resampling. After all, *heuristically speaking*, at the 5% level, we would expect to see an exception for every twenty (independent) tests we perform! In the light of the points raised in this subsection, even if there is the help of parametric bootstrapping, the uncertainty may be so large that exceptions with small $|\mathcal{L}|$ would be left unaccommodated.

For further reference, Section 7.1.2 of (Wüthrich & Merz, 2008) and Section 4.9 of (Shapland & Leong, 2010) have brief discussions on the parametric bootstrap.

*Superimpositions of exceptionally resampled residuals**

As mentioned above, simultaneous exception resampling could run into difficulties if the target location set, **L**, has overlapping elements. This could arise if a triangle of residuals has exceptional calendar periods and exceptional origin periods, for example. What makes the matter more complicated is that the practitioner would likely want more than one target location sets. A calendar period exception would naturally resample onto a set of calendar periods, while an origin period exception onto a set of origin periods.

In such circumstance, a series of exception resampling could be superimposed serially. More generally, suppose we have a series of exceptional features, $\mathbf{E} = \{\mathcal{E}_{g,h} = (\mathcal{L}_{g,h}, \mathcal{S}_{g,h}): g = 1, \dots, G; h = 1, \dots, H_g\}$, with similar ones grouped together. There are G groups, with $\mathcal{E}_g = \{\mathcal{E}_{g,h}: h = 1, \dots, H_g\}$. Let $\mathbf{L}_g$ be partitions of the triangle, for each $g = 1, \dots, G$, acting as targets for $\mathcal{E}_g$.

i. Perform (simultaneous) exception resampling with $\mathcal{E}_1$ onto $\mathbf{L}_1$. For each simulation, s, let $R_{1,s}$ be the set of resampled residuals and their positions in the triangle (this could be more rigorously defined as a set of triplets – the first dimension being origin periods, second development periods, and third the resampled residual value).
ii. Given we have $R_{g,s}$ for $g<G$, we now construct $R_{g+1,s}$. For each member, $\mathcal{L}^*$, of $\mathbf{L}_{g+1}$, determine whether it is to be resampled from an exceptional feature in $\mathcal{E}_{g+1}$ (see the discussions above on one-step exception resampling). If not, then $R_{g+1,s}(i,j) = R_{g,s}(i,j)$ for each (i,j) in $\mathcal{L}^*$. Otherwise, if it is determined that $\mathcal{E}$ (in $\mathcal{E}_{g+1}$) is to be resampled onto $\mathcal{L}^*$, then resample residuals from $\mathcal{E}$ onto $\mathcal{L}^*$, as per usual.
iii. $R_{G,s}$ would be the final resampled residuals for simulation s.

In this way, the exceptional features are superimposed onto a triangle of residuals from the previous iteration. The procedure can also involve sieve resampling in the sequence. At this point, it is useful to make the following observations.

- The effects of resampling performed earlier in the sequence can get diluted by later resampling. Therefore, the practitioner will need to determine which exceptional features are more important. A rule of thumb could be to perform development-period-related resampling first, as these can be considered as part of the normal course of events, and would be helpful to be embedded early as a basis. Another could be that calendar periods could be performed late in the sequence, to emphasise the shock nature of the exceptions associated with calendar periods.
- The procedure above could cause some residuals – particularly, those on the intersections of exceptional features – to be sampled more often than others. For non-parametric resampling, it is possible to adjust for this in step ii above. One way would be to demand that only residuals not in an exceptional feature of $\mathcal{E}_{g+1}$ can resample onto $\mathcal{L}^*$ that is determined not to be exceptional (in the sense of $\mathcal{E}_{g+1}$) for the simulation. If (i,j) is in $\mathcal{L}^*$ that is not exceptional (again, in the sense of $\mathcal{E}_{g+1}$) in the simulation, whenever $R_{g,s}(i,j)$ is a residual in an exceptional feature of $\mathcal{E}_{g+1}$, then one could perform a resample just for (i,j) to make sure that $R_{g+1,s}(i,j)$ is a residual not in any exceptions in $\mathcal{E}_{g+1}$.

We now continue with the Arch third party occurrence incurred example. Recall from a previous discussion that **M(1)** is constructed to accommodate the positive correlations seen between the third and fourth development periods. While this is successful in accommodating the exception, the practitioner may want to accommodate the 2004 origin period, which has a significantly high mean under both **M(0)** and **M(1)**.

Positive Correlations between Successive Development Periods;
High Developments on an Origin Period
Arch, 3rd Party Occurrence Insurance
Residuals from the Incurred Data

AY		1	2	3	4	5	6
	2002	93%	37%	111%	119%	-32%	128%
	2003	-15%	188%	-87%	-130%	-122%	-59%
	2004	**155%**	**70%**	**127%**	**92%**	**118%**	
	2005	-23%	-14%	34%	-18%		
	2006	-32%	-94%	-114%			
	2007	115%	-97%				
	2008	-149%					

Mean of the 2004 AY: 112%

A way of doing so would be to proceed as described abstractly earlier in this subsection. In our context, we would superimpose the exception resampling with the 2004 origin period onto the resampling already done under **M(1)**. We call this **M(2)**.

While **M(2)** does incorporate the 2004 origin period exception, it has an exception in the 2005 calendar year with a significantly high mean.

Positive Correlations between Successive Development Periods;
High Developments on an Origin Period and on a Calendar Period
Arch, 3rd Party Occurrence Insurance
Residuals from the Incurred Data

AY		1	2	3	4	5	6
	2002	93%	37%	**111%**	119%	-32%	128%
	2003	-15%	**188%**	-87%	-130%	-122%	-59%
	2004	**155%**	**70%**	**127%**	**92%**	**118%**	
	2005	-23%	-14%	34%	-18%		
	2006	-32%	-94%	-114%			
	2007	115%	-97%				
	2008	-149%					

Mean of the 2004 CY: 151%

We can now construct **M(3)** by further superimposing the exception resampling with the 2005 calendar year onto **M(2)**. This new model accommodates all three exceptions seen under **M(0)** in the development period, origin period and calendar period dimensions. The following tables present the progression of the *p*-values of the exceptions under the different iterations of the modelling, and statistics from the IBNR reserve distributions.

p- values of hypothesis testing				
Data: Arch 3rd Party Occ Incurred				
See text for definition of M(1), M(2) and M(3)				
(i) Test Statistic: Correlation Coefficient				
DP	**M(0)**	**M(1)**	**M(2)**	**M(3)**
3rd vs 4th	2%	21%	16%	7%
(ii) Test Statistic: Mean				
	M(0)	**M(1)**	**M(2)**	**M(3)**
2004 AY	2%	4%	19%	8%
2005 CY	1%	1%	1%	9%

All AY IBNR Reserve uncertainty							
Due to Estimation Error (USD 000)							
Data: Arch 3rd Party Occ Incurred							
See text for definition of M(1), M(2) and M(3)						**Differences**	
	M(0)	**M(1)**	**M(2)**	**M(3)**	**M(1):M(0)**	**M(2):M(1)**	**M(3):M(2)**
Mean	722,956	723,122	724,114	725,581	0.0%	0.1%	0.2%
SD	60,943	67,827	81,277	96,710	11.3%	19.8%	19.0%
Percentiles							
75th	764,670	768,101	777,828	788,493	0.4%	1.3%	1.4%
90th	802,245	811,979	832,016	857,280	1.2%	2.5%	3.0%
99.5th	883,359	899,154	946,678	1,008,285	1.8%	5.3%	6.5%

It is interesting to observe that the *p*-values decrease after first accommodation. The positive correlations between the third and fourth development periods are accommodated with **M(1)** with a high *p*-value. However, further superimpositions in **M(2)** and **M(3)** dilute this effect – so that by **M(3)**, the *p*-value gets close to our 5% threshold again. Similar comments could be made for the 2004 AY.

With regards to volatility, **M(3)** has added around 60% to the estimation error from **M(0)**. This is not a trivial gap, and represents a more extreme case (among the collection of global loss triangles considered) of how the volatility could *potentially* be underestimated by assuming the independence assumptions of the Mack bootstrap.

Sieve resampling could also be in the superimposition "chain". In the subsection on sieve resampling, we have already constructed **M(1)** for the ACE North American workers' compensation incurred data, constraining the resampling on the first development period and the remainder. Under this model, there is significantly high correlation between the third and fourth development periods (*p*-value at 2%). Performing exception resampling to accommodate this exception, and then superimposing on **M(1)** would give a further impact of around 11% increase in estimation error.

Further remarks

Exception and sieve resampling are presented as possible extensions to the Mack bootstrap. They attempt to add just enough to the machinery already available to the practitioner – the original Mack bootstrap – to accommodate instances where its independence assumptions fail as part of its usual claim development dynamics. This approach contrasts with employing a new model "from scratch", which also has its time and place.

An advantage of using these resampling on the Mack bootstrap is that it is relatively intuitive to the practitioner who is already familiar with Mack bootstrapping. Its intuitiveness is derived from confronting direct the perennial failure of the independence assumptions in most triangles. If required, the simulations could be analysed to help discussions of possibilities of large estimation error. How often do we see exceptions being simulated in, say, the worst 10 percent of estimation error? Which calendar periods are driving such simulations? Informed by answers of such questions, the practitioner can weave together narratives to further the company's understanding of claim developments.

A disadvantage is that the analytic formulae for estimation error no longer hold. Mack's formulae for the estimation and forecast errors are fundamentally based on the independence assumptions. Practitioners appreciate such formulae as they give immediate indications of risk in spreadsheets, without triggering runs of Monte Carlo simulations. The author leaves this as an open question to derive such formulae for the various exception and sieve resampling.

As indicated in the *Definitions and terminology* section above, the formula $\widehat{\sigma_j^2} = \frac{1}{I-j-1}\sum_{i\leq I-j} C_{i,j}\left(\lambda_{i,j} - \widehat{f_j}\right)^2$ gives an unbiased estimator for the variance parameter σ_j^2 under the independence assumptions. It is unbiased due to the independence of the developments through the development period j. In our simulation testing, we do not observe the means of the variance parameters of the pseudo data to significantly change with calendar period or origin period exception resampling (data: Arch third party occurrence incurred). This is expected, as exception resampling with calendar or origin periods maintains independence of residuals down the columns. A small change is observed – up to -2% – when we used development period exception resampling on pairs of residuals. A difference is expected here, since there is implied dependencies down the column, although not material. Using parametric resampling on small exception periods can increase simulated variance parameters: up to +10% impact is observed when using parametric resampling on the small 2002 year on XL casualty insurance incurred data.

We leave this as a future research area for formulas to indicate impacts on the variance parameters of various different resampling methods. For now, the reader may want to perform these impact analyses, and where the difference becomes intolerable, adjust the variance parameters appropriately. In assessing the impacts, it is useful to remember that a +10% increase in the variance parameter represents only around +5% change in the estimation error.

There are other ways to proceed from here for further refinements of the methodology. One way is for a bank of exceptional features to be set up. As we mentioned, companies are more and more open to publishing claim development triangles. Regulators make available triangles from statutory returns to the market. The collection of exceptional features identified in the data of one's company could be supplemented by those seen in other appropriate triangles, giving a richer set of exceptions to simulate from. A difficulty is the determination of the weighting given to the exceptional features from different triangles. Another issue is how to adjust for the different claim processing bases and procedures of different companies.

Yet another way, somewhat related to the above, would be to consider adjustments of *frequencies* of the exceptions. Much of this section assumes the exceptional features are at the return periods that are consistent with the size of the observed exceptional features. However, the return periods may well be very different – our observed may be a random occurrence (or non-occurrence!) of the exceptions. In the extreme case, a frequency of zero might be employed if the observed exception was deemed a one-off special case that would not be repeated in the future.

The proposed extensions work well with shocks. Trends are more difficult to deal with under such extension. We have not dealt with trends here as they are more of a problem with larger triangles (e.g. those going back to the 1970's or 1980's). Data from the more remote past are typically less applicable in the context of the fast changing London market environment. Significant changes in coverage, terms and conditions, and claim handling procedures, contribute to this, and the practitioner is likely to ignore these earlier data. Nevertheless, a next step could be to develop an enhancement to the exception resampling approach to take into account of trends – although we suspect that a different approach would be more appropriate (see the above subsection *What should we do with the exceptions in assessing estimation error?*).

Example exceptional features mentioned in this section are with single calendar periods, single origin periods, development regions and pairs of development periods. With the amount of relevant data typically available to London Market actuaries, these seem a sensible list to start with. With more data, one could investigate other structures. An example is *pairs* of calendar periods, in which issues may take up to two calendar years to be resolved. Another is to consider calendar periods confined to the later origin periods – to take care of situations where, for example, policies have fixed discovery periods. Narratives behind candidate structures for testing are important to avoid modelling random observations. To an extreme, it is also possible – although probably not desirable – to randomly mine for exceptional regions. In the Axis marine insurance incurred triangle, the largest negative residuals are in positions (2006, 3), (2002, 5) and (2005, 1): unless more external information is available to tell us otherwise, there is no reason to declare this set of observations as significant, as no natural connection exists to connect the three points in the triangle.

Estimation error can be relatively difficult to communicate on its own. There are benefits in being able to do so. Not only could the information help identify the source of uncertainty in the ultimate distributions, it could also help with model validation. Estimation error is to do with how far (estimated) best estimates, $\widehat{C_{i,I}}$, could deviate from the actual mean $E\left(C_{i,I}|D\right)$. Bootstrapping gives us a possibility of creating different triangles (the "pseudo data") to derive different development pattern estimates – and hence a distribution of possible $\widehat{C_{i,I}}$ – that could have been derived under the same claim development dynamics. The reserving actuary deals with best estimate ranges regularly (for example, please refer to (Gibson & others, 2011) for a discussion): but it is unusual to place probabilities on these ranges, perhaps because it is not intuitive to place probabilities on what *could have* happened (as is required of us here).

Attempts to use a frequency approach in communicating estimation error can have limited success. Statements like "out of one hundred independent actuaries, given the same data, we would expect ten to give an answer above $Xm" seem to help. But they can be too far from what is being modelled – the example statement can give the impression that we do include model error, which is not true. They can also be too unrealistic to draw analogies with. Considering the same statement, another question might be that, given their similar training and background, do actuaries really derive projections independently? A third difficulty is that the model (extended or otherwise) assumes the actuaries *blindly* apply the chain ladder. At this point, we leave as an open challenge to the reader to come up with better ways to communicate estimation error.

Finally, we note that the formulation of sieve and exception resampling presented here seem to be sparse in the literature, even though the author knows that sieve resampling is widely used (see Section 4.6 of (Shapland & Leong, 2010)). A search on the wider literature suggests sieve resampling is used in other statistical contexts – see, for example, (Bühlmann, 1997). Other resampling techniques can be seen in (Taylor & McGuire, 2005) and Section 7.6 of (Wüthrich & Merz, 2008).

The independence assumptions and forecast error

Both intuitively and statistically, the independence assumptions in the original Mack Bootstrap model are not tenable for projecting forecast error. For long-tailed liability, specific issues could emerge over a calendar period, stretching back several origin periods. These issues may range from external systemic events such as inflation shocks, discovery of previously latent perils, changes in legal interpretations, to internal factors such as changes in claim handling processes. Statistically, many triangles do not fulfil the strong independence assumptions, as required by the original Mack model. We have discussed this in the identification of exceptional features previously.

At this point, it is worth mentioning that we are talking about independence *conditional on* the parameters (i.e. the development patterns). Since the chain ladder uses the same development patterns to project all origin periods, albeit offset to the right development periods, estimation error becomes a driver across origin periods to the overall risk. For instance, simulations with high estimation error would tend to give higher projections in the mean across all origin periods, and vice-versa. The ultimate distributions of long-tailed classes – taking into account both estimation and forecast errors – can give moderate Spearman's rank correlation coefficients between adjacent origin periods in a Mack bootstrap with independent forecast. The following table shows those simulated from the XL casualty insurance incurred data.

Rank correlation coefficients of IBNR Distributions between accident years
Data: XL Casualty insurance Incurred
Model: Original Mack Bootstrap

	2001	2002	2003	2004	2005	2006	2007	2008	2009
2001	100%	2%	4%	2%	2%	2%	3%	1%	2%
2002	2%	100%	16%	16%	16%	15%	14%	9%	6%
2003	4%	16%	100%	16%	18%	15%	14%	8%	7%
2004	2%	16%	16%	100%	18%	12%	13%	7%	5%
2005	2%	16%	18%	18%	100%	17%	15%	9%	5%
2006	2%	15%	15%	12%	17%	100%	17%	9%	7%
2007	3%	14%	14%	13%	15%	17%	100%	10%	8%
2008	1%	9%	8%	7%	9%	9%	10%	100%	8%
2009	2%	6%	7%	5%	5%	7%	8%	8%	100%

We propose an approach to extend the Mack bootstrap to forecast into the future, with primacy given to dependencies within diagonals in the bottom halves of the triangles – i.e. within future calendar periods. Secondary dependencies are applied between the diagonals. The extension preserves the use of the gamma distributions commonly used for modelling each future cell, $C_{i,j+1}$ of the triangle, the use of simulated $\widehat{f_j}\ C_{i,j}$ as the mean and the use of the variance parameters $\widehat{\sigma_j^2}\ C_{i,j}$ as the variance.
The section is divided into three subsections.

- **Calendar period drivers**. A statistical way of incorporating calendar period driver is presented. There is a step-by-step example, as well as indications of impacts on the results from a claims triangle. The example is one which we have already considered in the section above on extended resampling techniques.

- **Secondary dependencies**. Correlations between calendar periods and within calendar periods are discussed. The example in the previous subsection is carried forward here.

- **Further remarks**. The section ends with some further remarks on the strengths and limitation of the calendar period driver approach; ways to generalise to model different structures; possible next steps and research topics.

Calendar period drivers

We now describe the first part of the proposed extension. The aim of this part is to model calendar period shocks for individual future calendar periods, in the framework of using the gamma distributions (or other appropriate distributions) for each cell in the cumulative triangle.

We can exploit a very common Monte Carlo method to generate distributions: first simulate from a uniform (0,1) distribution, and then use the inverse cumulative distribution function (CDF) of the distribution to obtain simulations from the distribution. For example, if a future calendar period (i.e. a diagonal in the bottom half of the triangle), k, is simulated to be "exceptional" in a certain sense, then we would ask that simulation to tend to return "exceptional" uniforms $u_{i,j}$ (where $i+j-1 = k$). When the inverse gamma CDFs are then applied to the $u_{i,j}$, we would then tend to have "exceptional" values for that simulation in the calendar period, k.

One way to define "exceptional" is to look for calendar periods in the observed residuals that are exceptional relative to **M(0)**. The model **M(0)** is used here due to our wish to extend **M(0)**, the one that has the independence assumptions. Suppose we have one exceptional calendar period, k, that we want to model as a shock in the forecast. Let R_k be the set of observed residuals in calendar period k. Let R_k^c be the set of observed residuals outside the calendar period k. Then the following procedure could be followed to forecast this shock into the future.

i. Find statistical properties of the R_k. Mean, standard deviation and perhaps skewness are examples. Do the same for R_k^c.
ii. Use these statistical properties to fit distributions, using usual procedures such as the method of moments, for the residuals. For example, one could use the normal or translated gamma. Let F_k and F_k^c be the CDFs of the distributions corresponding to R_k and R_k^c, respectively.
iii. Find an overall distribution F which is a weighted mix of F_k and F_k^c. This could be done using a simulation approach. For example, one could simulate 1,000 simulations from F_k, and then independently, simulate 1,000 simulations from F_k^c. Let p_k be the weight for F_k and p_k^c be the weight for F_k^c. Then 1,000 simulations of F could be obtained by randomly picking from F_k and F_k^c according to the weights p_k and p_k^c.
iv. Let V_k: (0,1) → (0,1) be a map, so that the $V_k(u)$ is the percentile under the mixed distribution of the uth percentile under F_k. That is to say, $V_k(u) = F\left(F_k^{-1}(u)\right)$. Similarly, define the map V_k^c: (0,1) → (0,1) such that $V_k^c(u) = F\left(F_k^{c-1}(u)\right)$. Under a simulation approach from Step 3 (in the Mack bootstrap subsection above), the maps could be estimated through simulation by reading off the different percentiles of the 1,000 simulated F_k and F_k^c against the simulated F.
v. We are now ready to project the bottom half of the triangle, iteratively along the origin period dimension (similarly to the original Mack bootstrap). To define the iteration, we assume that we have simulated $C_{i,j}$ for all (i,j) in calendar period k^* and are wanting to simulate $C_{i,j+1}$ (i.e. the cumulative amounts for the next calendar period, k^*+1).
vi. For the calendar period k^*+1, determine whether it is exceptional like k or not, using weights p_k and p_k^c.
vii. For each cell, $C_{i,j+1}$, derive the gamma parameters α and β as per usual, using the method of moments, with $\widehat{f_j}.C_{i,j}$ as the mean and $\widehat{\sigma_j^2}.C_{i,j}$ as the variance. Let the inverse gamma distribution thus parameterised be $G_{i,j+1}^{-1}$.
viii. Let $u_{i,j+1}$ be from the uniform (0,1) distribution for each $(i,j+1)$ in calendar period k^*+1.

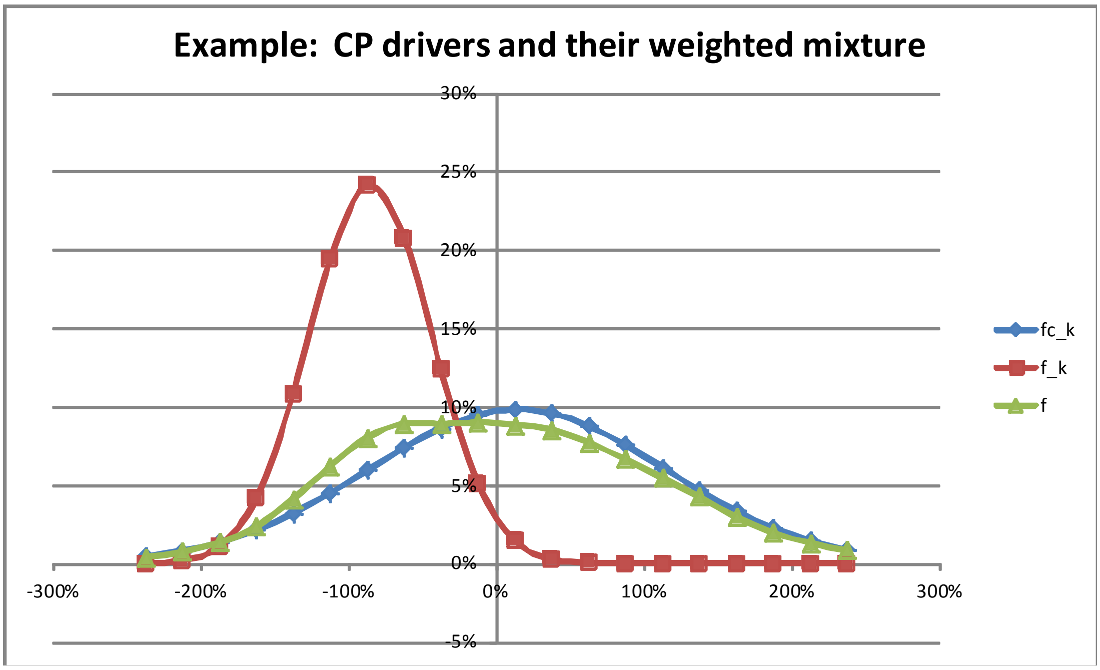
ix. If k^*+1 is simulated to be exceptional like k, then set $C_{i,j+1} = G_{i,j+1}^{-1}(V_k(u_{i,j+1}))$. If k^*+1 is simulated not to be exceptional, then set $C_{i,j+1} = G_{i,j+1}^{-1}(V_k^c(u_{i,j+1}))$.

In summary, Steps i to iv produce a mechanism under which the exceptional feature is defined to be a distribution on the unit interval (0,1). Steps v to ix is the usual Mack bootstrap projection into the future, but with an extra step that determines the presence of exceptional features in individual calendar periods. Steps iii and vi make use of the same weights: this is important so that, over all simulations, we would get the whole gamma distribution without distortions in probabilities.

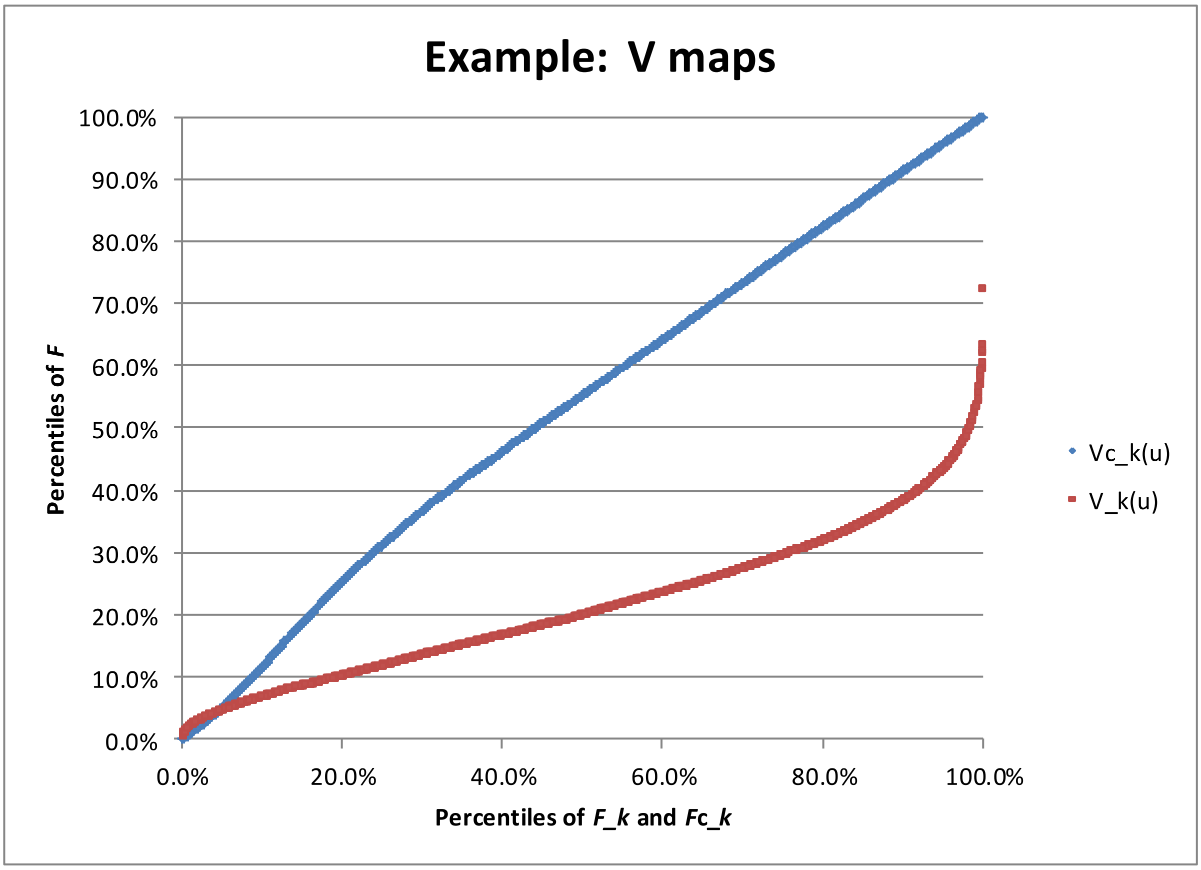
We now demonstrate how the above steps play out in an example, using the XL casualty insurance incurred triangle, paying particular attention to steps i to iv to produce the V maps. The 2005 calendar year is our exceptional period, R_k.

i. The mean and standard deviation from R_k are -85% and 41%, respectively. Those for R_k^c are 14% and 101%.
ii. We make use of the normal distributions, with the above means and standard deviations, for F_k and F_k^c.
iii. There are 44 observed residuals in total, and five are in R_k. Set p_k to be 5/44 and p_k^c to be 39/44. The following table shows the first 10 simulations: the first two columns are normal distributions for the driver (from F_k) and for the complement (from F_k^c). The fourth column is the mixture, F, of the first two, determined by whether it takes its value from F_k or from F_k^c. The table is followed by a graphical representation.

Simulation	Fc_k	F_k	Driver?	***F***
1	27%	-53%	No	***27%***
2	39%	-95%	No	***39%***
3	47%	-111%	Yes	***-111%***
4	-65%	-68%	No	***-65%***
5	23%	6%	No	***23%***
6	-76%	-86%	No	***-76%***
7	-53%	-28%	No	***-53%***
8	104%	-133%	No	***104%***
9	-197%	-62%	No	***-197%***
10	5%	-84%	No	***5%***



iv. The *V* maps could now be built, by considering where each simulated value of F_k and of $F_k^{\,c}$ sits in F. The following chart is a picture of these maps.



Steps v to ix are straightforward, with the usual calibration of the gamma distribution, supplemented by conditionally transformed percentiles.

The above steps can easily be adapted to simulate for more than one exceptional calendar period as drivers into the future. In the XL casualty insurance incurred data, the 2006 year was also identified as an exception. Indeed, all historic calendar periods can potentially be incorporated as drivers in this way. Using all historic calendar periods may not be statistically parsimonious: we include this here to help the discussion in the following sections, on applying correlations between calendar periods.

The following table gives the impact of including (i) the 2005, (ii) the 2005 and 2006, and (iii) the 2004, 2005, …, 2009 calendar years as drivers as described above on *purely* the forecast error. As one would expect, the changes in volatility can be significant. Including all years would increase the forecast error by 29%. (The 2004, …, 2009 test run is not *all* the calendar periods in the triangle. It is hard to calibrate the earlier calendar periods, as there are only a few residuals in each of them.)

All AY IBNR Reserve uncertainty
Due to Forecast Error ONLY (USD 000)
Data: XL Casualty Insurance Incurred

	0	**1**	**2**	**3**				
	Original	**with 2005**	**with 05/06**	**with 04-09**			**Differences**	
	Mack	**CY Driver**	**CY Drivers**	**CY Drivers**		**1:0**	**2:1**	**3:2**
Mean	1,048,526	1,048,003	1,047,123	1,046,523		0.0%	-0.1%	-0.1%
SD	322,866	363,079	374,729	415,192		12.5%	3.2%	10.8%
Percentiles								
75th	1,255,961	1,293,729	1,296,548	1,318,965		3.0%	0.2%	1.7%
90th	1,472,228	1,515,636	1,531,180	1,598,364		2.9%	1.0%	4.4%
99.5th	1,933,570	2,015,232	2,060,083	2,176,476		4.2%	2.2%	5.6%

The table below summarises the prediction error (i.e. estimation error together with forecast error) under different bootstrapping and calendar period driver assumptions. For estimation error estimation, recall that **M(0)** stands for the original Mack bootstrap. The model **M(1)** represents calendar period exception resampling with the 2005 year, with parametric bootstrapping. The impacts on the prediction error from the original Mack bootstrap are also shown. These are large: not allowing for any calendar period drivers or exception resampling could be underestimating risk by 20%.

All AY IBNR Reserve Prediction Error (USD 000)
Data: XL Casualty Insurance Incurred

Estimation Error Models	***Forecast Error Models (see above for definition)*** 0	1	2	3	***Differeces vs Original Mack Bootstrap*** 0	1	2	3
None	322,866	363,079	374,729	415,192	-25%	-15%	-13%	-3%
M(0)	**428,543**	462,257	472,505	504,967	**0%**	8%	10%	18%
M(1)	454,242	485,591	493,175	514,301	6%	13%	15%	20%

A final remark concerns the distribution chosen in Step ii for each calendar period driver. The author has tested using the translated gamma instead of the normal. Doing this would also incorporate the third moment. However, the impact to the prediction error is small for this dataset. The force of the driver is more expected to come from its mean levels and then uncertainty around the means (in the form of the standard deviations). We would not expect the third moment to be of significance, as the V maps squeeze the distributions to the unit interval, suppressing the impacts from tails.

Secondary dependencies

The practitioner may have reasons to impose further dependencies between the origin periods inside a future calendar period, k. An example may be that adjacent origin periods may be associated closely. This can be easily performed by generating the $u_{i,j+1}$ in Step viii with, say, a Gaussian copula with appropriate correlation coefficients between the different pairs of origin periods. The calibration of such correlation coefficients is likely to be informed heavily by qualitative reasoning, although limited quantitative evidence could be obtained by computing sample correlation coefficients from the observed triangle of different pairs of origin periods.

It could also be desirable to demand dependencies be imposed between the calendar periods. There could be "runs" of high emergence of claims, for example. This extra dependency could be applied to a statistic (see Step 1 above) from the exception residual distributions, depending on the context. Applying dependency to the means would allow a severe development shock to have a higher chance of being followed by another severe development shock. Applying it to the standard deviations would be saying that highly variable developments in a calendar period would have a higher chance of being followed by another year of highly variable developments. Again, the practitioner may rely heavily on qualitative reasoning for the correlation coefficients.

Note that the above way of correlating between calendar periods imposes a copula on mass points (there are only a limited number of calendar period drivers to simulate from). This could give residual progressions that are too stepped. One way to get around this would be to enlist *all* observed calendar periods – except for, say, the earliest ones in one's triangle due to the lack of observed data there – for projection into the future. Using all calendar periods as drivers may be a good compromise than using just one or two: trading parsimonious modelling for a smoother outcome. In the example below, we shall demonstrate this impact.

The following tabulates the prediction error from a variety of options of extended the Mack forecasting, based on the XL casualty insurance incurred data. The extended Mack bootstrap model **M(1)** is used for estimation error. The rows of the table represent the calendar periods to be counted as drivers (see the subsection immediately above). The columns represent secondary dependencies mentioned in this subsection: 10% on adjacent origin periods within each calendar period, and then with 10% on the means of adjacent calendar period drivers.

All AY IBNR Reserve Prediction Error (USD 000)							
Data: XL Casualty Insurance Incurred							
Estimation Error Model: **M(1)** (see text for definition)							
Calendar Period Drivers	***Secondary Dependencies (see text for definition)***					**Impacts**	
		None	10% within CP OP correlation	also with 10% CP Correlation		10% within CP OP correlation	10% CP Correlation
None		454,242	464,466	464,466		2.3%	0.0%
2005		485,591	495,883	499,491		2.1%	0.7%
2005 & 2006		493,175	505,018	507,579		2.4%	0.5%
2004, 2005, …, 2009		514,301	528,012	539,590		2.7%	2.2%

We make the following observations in relation to the above results.

- Taking one calendar period at a time, the lag-1 auto-correlations between the residuals range from -59% to +23%. Given the small number of residuals in each calendar period, the subjective figure of 10% is reasonable for the correlation coefficient between adjacent origin periods in a calendar period.
- The impacts to the prediction error are not large: around 2%.
- The figure of 10% for correlating between adjacent calendar periods is also subjectively chosen. It seems tolerably prudent against *just one* possible observation: the lag-1 auto-correlation of the means of the calendar period residuals is -38%.
- Clearly, when no calendar period driver is used, there is no driver means to apply correlations to – so impact is nil in this case.
- Impacts, as expected, from the calendar period driver correlations increase as we use more drivers: up to 7% when we use six drivers, from 2004 to 2009.

Another way to do this may be to correlate a statistic (e.g. the means) from the simulated residuals or uniform distributions. This should give claim evolutions that are much smoother from one calendar year to the next. However, the rationale behind this approach is trickier to communicate: instead of imposing dependencies between drivers, we are imposing dependencies between the *random outcomes* of drivers. Computationally this could also be quite complicated to programme and simulate – as the simulation of next year's driver (i.e. Step 6) would require the knowledge of the simulated outcomes from all potential drivers.

Further remarks

The calendar period driver approach, together with imposing dependencies, is statistical, driven by percentiles. The key advantage of this approach is that it adapts what we already have to further model important structures such as calendar period shocks. To some extent, it allows investigation into how particular parts of the reserve distribution come about, through tracing back the simulations. It is also possible to output monetary impacts for wider discussions.

However, it also means that there is limited direct control over severities. Moreover, using the *V* maps puts a further distance between the driver assumptions – which are based on statistics of residuals – and the output monetary amounts. It would be a helpful next step to research on how better one could implement the drivers.

Even with using all the calendar periods in a ten-year triangle, there is likely to be other calendar period behaviour that is not taken into account. A bank of calendar period drivers would be useful taken from triangles from other similar classes of business. It would be useful to have research done on how one can place credibility weighting on these other reference drivers.

Earlier in the paper, we have presented various ways to perform resampling to assess estimation error. The exception resampling method with calendar period locations is most consistent with the calendar period driver approach described in this section. We have presented the calendar period driver approach here as the calendar period emergence of risk is one that the industry is particularly interested, with regards to forthcoming regulatory requirements. However, using a similar framework, other forecast projection methods could be devised to correspond with other specific exception resampling definitions. For example, if the horizontal pairwise correlations of residuals are judged particularly important for the class, instead of calendar period drivers, we could have development period driver-pairs. Each pair would have particular correlation coefficients as seen in the data triangle. The counterpart to sieve resampling is likely to come from the use of more carefully distinguished distributions in the iterative forecast projection.

We note that for origin period dynamics, there are usually different and better ways to deal with them. For example, we can take into account information derived from analyses done by the reserving actuary. Scaling the development factors f_j for each origin period is one way of doing so. Adjusting the variance parameters $\widehat{\sigma_j^2}$ for different business mixes or reinsurance for each origin period is another.

As with exception resampling, the means of the resulting ultimates would no longer be the chain ladder projected ultimates. We make the same comments here as for exception resampling. We assume that the reader engages in widespread, albeit small-scale, scaling, so that the outputs of the stochastic model has pre-defined figures as the means of the distributions. They should check that the additional scaling required after the extension techniques (around 3% in the case of our simulation on the XL casualty insurance incurred data) is tolerable and would not materially distort risk assessments.

With regards to variance parameters, correlating calendar period drivers positively could increase the effective variance, as one would expect. For the XL casualty insurance incurred triangle, the increase is small: around 2%. Incorporating calendar period drivers themselves, or imposing dependencies between origin periods inside a calendar period, has no impact: independence down the columns are maintained. Appropriate adjustments to the variance parameters would be required if impact becomes large. Further research could be undertaken to arrive at a formula for impact on the variance parameters on the various different approaches.

Typically, copulae are imposed on the overall reserve distributions between classes. This has the advantage of simplicity in an area where calibration is particularly difficult due to the lack of data. However, this overall approach gives the practitioner only indirect control over how the risk emerges for several classes over the next calendar period. With the calendar period driver approach, it is possible to improve on this, through imposing copulas on the calendar period driver simulations. (The reader may also refer to a similar approach proposed in (Cairns, 2010), Slide 37 and following.)

We make a final note with regards to (blind) applications of the chain ladder method in claims reserving. Assuming underlying claim development dynamics, it is possible to simulate claim developments from the first development period to ultimate. A triangle could be cut out from each simulation, and the chain ladder method could be applied to the simulated triangle for an estimation of the ultimate. (Please see Section 9 of (ROC / GIRO Working Party, 2007) for a similar testing procedure in the context of examining stochastic reserving methods.) For example, assuming the statistics, $\widehat{f_j}$ and $\widehat{\sigma_j^2}$, from the XL casualty insurance incurred triangle, we could start with the observed first development period incurred amounts and simulate claim developments assuming (i) no calendar period drivers and (ii) the 2005 calendar period driver that could occur in any calendar period. For the second case, the machinery introduced in this section helps to incorporate such a calendar period driver. The table below shows the mean and standard deviations of the simulated results.

Statistics of "future" amounts ($m)
Data: XL casualty insurance incurred

	Simulated		Chain Ladder		Difference		SD Comparison
	Mean	SD	Mean	SD	Mean	SD	Difference vs Simulated
M(0)	1,077	339	1,077	353	0	462	36%
M(0) with 2005 calendar period driver	1,073	382	1,079	390	7	510	33%

The following observations could be made.

- For the assumed claim development dynamics, performing the chain ladder at year-end 2009 would on average give little difference (around 0.7%) against what would actually occur in the simulated run-offs.
- The standard deviations of the differences are high compared with those of the simulated future amounts: 36% and 33% for the two dynamics. Intuitively, the chain ladder would project extreme historic developments into the future.
- The two standard deviation comparisons (36% and 33%) being near to one another suggests that the (blind) application of the chain ladder does not fare worse as a claims reserving method in the second dynamics, relative to the general uncertainty underlying the dynamics.

The third observation is surprising, given that the chain ladder does not deal naturally with calendar period shocks. We hypothecate here that this is due to the way that we assume the calendar period driver was not a "one-off", but something that could happen equally likely in the future as in the past.

This observation is important for applying the estimation error concepts discussed in this paper to the triangle. If the chain ladder future amounts fare much worse in a world with exceptions, then there would be further evidence to steer away from using the chain ladder method in reserving. In such a case, it could become difficult to justify the use of the (stochastic) f_j's as the sole conduit for estimation error, as we have done in this paper.

We leave to the interested reader to investigate these observations further. A particular question is: under what dynamics does the (blind) chain ladder fare worse than under the original Mack dynamics? How can we adjust for *intelligent* applications of the chain ladder?

Conclusions

The paper has presented practical techniques to extend the Mack bootstrap, when there are reasons to weaken the independence assumptions in the method. It has also discussed how one may be able to identify structures – the *exceptions* – in the residuals which break the independence assumptions. Calendar periods, origin periods and development period regions have been considered.

The estimation error is assessed via resampling in the Mack bootstrap framework. Two resampling techniques have been discussed to extend the framework: they aim to build extended models to accommodate the exceptions. These are the sieve resampling – which is used in the industry – and the exception resampling. The example does not show sieve resampling to be impactful on estimation error. Exception resampling could increase the estimation error significantly, through driving dependencies between the development factors.

The forecast error is measured via an iterative simulation approach, typically using the gamma distribution. The incorporation of calendar period drivers into this framework has been discussed. Ways to model further dependencies have been indicated.

We urge the practitioners to continue the debate of the role played by the independence assumptions in the Mack bootstrapping framework. This debate would be important as the assumptions can help to underestimate risk. Recently published loss development triangles have been used to demonstrate the techniques at all key points in this paper: in particular, the XL casualty insurance incurred data have been used across both sections. For this dataset, there could be an increase in the prediction error of around 17% with using some or all of the extended techniques discussed in this paper:

- around +6% for using extended Mack bootstrapping in the form of calendar period exception resampling with the 2005 year;
- around +7% for using calendar period driver 2005 in the forecasting; and
- around +3% for using secondary dependencies on the forecasting, as discussed in this subsection.

Finally, there are open questions that could be helpfully investigated and further improvements that could be usefully made. These are indicated throughout the paper, and it is the author's hope that further research could be undertaken by academia and the industry on these points.

Bibliography

ACE Limited. (2010). *ACE Limited 2009 Global Loss Triangles.* Retrieved July 11, 2011, from http://phx.corporate-ir.net/External.File?item=UGFyZW50SUQ9Mzk3NjA3fENoaWxkSUQ9NDAzOTk0fFR5cGU9MQ==&t=1

Arch Capital Group Ltd. (2010, September 1). *2009 Loss Development Triangles.* Retrieved July 11, 2011, from http://www.archcapgroup.com/docs/2009%20Global%20Loss%20Triangles.pdf

Axis Capital Holdings Limited. (2010, May 11). *2009 Loss Development Triangles.* Retrieved July 11, 2011, from http://www.axiscapital.com/document-centre/pdf/2010/financial/AXIS%20Capital%202009%20Loss%20Development%20Triangles.pdf

Barnett, G., & Zehnwirth, B. (2000). Best estimates for reserves. *Proceedings of the Casualty Actuarial Society, LXXXVII*(2), 245-321.

Bühlmann, P. (1997). Sieve bootstrap for time series. *Bernoulli, 3*(2), 123-148.

Busse, M., Mueller, U., & Dacorogna, M. (2010). Robust estimation of reserve risk. *Astin Bulletin, 40*(2), 453-490.

Cairns, M. (2010, October 12-15). *Statistical quality standards? Correlations? Help!* Retrieved July 20, 2011, from The Actuarial Profession (GIRO 2010): http://www.actuaries.org.uk/sites/all/files/documents/pdf/workshop-d06-martin-cairns.pdf

England, P., & Verrall, R. (2002). Stochastic claims reserving in general insurance. *British Actuarial Journal, 8*, 443-544.

England, P., & Verrall, R. (2006). Predictive distributions of outstanding liabilities in general insurance. *Annals of Actuarial Science, 1*(II), 221-270.

Gibson, E., & others. (2011). A meta-study of the General Insurance Reserving Issues Taskforce and Reserving Oversight Committee Research in this area between 2004 and 2009. *British Actuarial Journal, 16*(1), 63-80.

Mack, T. (1993). Distribution-free calculation of the standard error of chain ladder reserve estimates. *Astin Bulletin*, *23*(2), 213-225.

Martínez Miranda, M. D., Nielsen, J. P., & Verrall, R. (2011). Double chain ladder. *ASTIN Colloquium*.

Orr, J. (2007). *A simple multi-state reserving model.* Retrieved August 4, 2011, from International Actuarial Association (ASTIN Colloquium, 2007): http://www.actuaries.org/ASTIN/Colloquia/Orlando/Papers/Orr.pdf

Orr, J., & Hawes, W. (2010, October). *One year volatility & risk margins - the regulatory view.* Retrieved August 4, 2011, from The Actuarial Profession (GIRO 2010 presentation): http://www.actuaries.org.uk/sites/all/files/documents/pdf/workshop-a08-james-orr-wendy-hawes.pdf

ROC / GIRO Working Party. (2007, July 24). *Best estimate and reserve uncertainty.* (L. Gibson, Ed.) Retrieved July 24, 2011, from The Actuarial Profession, U.K.: http://test.actuaries.org.uk/sites/all/files/documents/pdf/bhprizegibson.pdf

Shapland, M., & Leong, J. (. (2010). Bootstrap modeling: beyond the basics. *CAS E-Forum, 2010 Fall*.

Taylor, G., & McGuire, G. (2005). Synchronous Bootstrapping of Seemingly Unrelated Regressions. *Astin Colloquium*.

Wüthrich, M. V., & Merz, M. (2008). *Stochastic claims reserving methods in insurance.* Wiley.
XL Capital Ltd. (2010, May 5). *2009 Global Loss Triangles.* Retrieved October 18, 2010, from http://investor.xlgroup.com/phoenix.zhtml?c=73041&p=irol-financialreports

Appendix: Example data used in the paper

The following published loss development triangles have been used to demonstrate the techniques presented in the paper. The figures correspond to the page numbers where they have been used.

For ease of reference, we now include here the triangles. The estimates, $\widehat{f_J}$ and $\widehat{\sigma_J^2}$ are also included, as are the adjusted residuals $r_{i,j}$. Note that $\widehat{\sigma_{I-1}^2}$ cannot be defined by the Statement 5: we have taken a usual convention to set it equal to the minimum of $\widehat{\sigma_{I-2}^2}$ and $\widehat{\sigma_{I-3}^2}$ (see also (Mack, 1993) for an alternative). Where relevant, the specification of the extended Mack bootstrap models **M(*n*)** is listed.

Data: ACE North American workers' compensation incurred

Cumulative Triangle

	Dev Year									
AY	1	2	3	4	5	6	7	8	9	10
2000	38,829	80,723	132,149	136,011	136,643	144,122	149,306	155,008	161,436	162,306
2001	29,184	61,263	70,150	88,634	113,500	118,297	122,552	131,052	131,716	
2002	55,661	123,705	133,160	127,622	123,103	129,715	140,501	144,843		
2003	100,257	119,012	123,441	146,927	173,148	181,572	195,540			
2004	160,035	170,907	197,240	224,017	236,341	261,263				
2005	136,205	260,162	318,024	362,556	391,019					
2006	167,886	299,460	361,339	407,641						
2007	155,156	268,417	328,479							
2008	148,450	253,437								
2009	129,684									
f-hat	165%	120%	112%	108%	107%	106%	104%	102%	101%	
sigma^2-hat	15,084	3,042	1,136	1,520	130	91	54	94	54	

Residuals

	Dev Year								
AY	1	2	3	4	5	6	7	8	9
2000	73%	239%	-104%	-79%	-44%	-109%	-44%	96%	
2001	66%	-28%	123%	167%	-81%	-98%	143%	-104%	
2002	116%	-86%	-187%	-117%	-45%	103%	-88%		
2003	-127%	-111%	81%	105%	-74%	89%			
2004	-201%	-39%	25%	-35%	185%				
2005	83%	20%	39%	-4%					
2006	47%	4%	19%						
2007	27%	21%							
2008	19%								

Extended Mack Bootstrap Models

M(0)	Original Mack bootstrap: residuals are independent and identically distributed
M(1)	Sieve resampling, partitioning out the 1st development period
M(2)	Pairwise exception resampling on the 3rd and 4th development periods, superimposed onto M(1)

Data: Arch Third Party Claims Made Paid

Cumulative Triangle

AY	Dev Year 1	2	3	4	5	6	7	8
2002	2	297	618	525	471	478	469	2,002
2003	1,074	6,761	9,117	11,442	16,172	19,024	22,469	
2004	6,777	23,422	42,521	61,868	70,153	72,965		
2005	640	27,972	52,188	67,299	73,347			
2006	6,229	37,250	84,099	122,292				
2007	7,020	43,080	85,970					
2008	13,283	54,006						
2009	6,956							

	1	2	3	4	5	6	7
f-hat	550%	198%	140%	113%	107%	118%	427%
sigma^2-hat	5,421,704	41,633	10,981	11,109	3,875	592	592

Residuals

AY	Dev Year 1	2	3	4	5	6	7
2002	48%	5%	-82%	-34%	-12%	-140%	
2003	7%	-156%	-81%	184%	156%	22%	
2004	-44%	-75%	72%	-1%	-74%		
2005	252%	-57%	-148%	-72%			
2006	10%	163%	99%				
2007	14%	11%					
2008	-43%						

Extended Mack Bootstrap Models

M(0)	Original Mack bootstrap: residuals are independent and identically distributed
M(1)	Sieve resampling, partitioning out the 1st development period

Data: Arch Third Party Occurrence Incurred

Cumulative Triangle

AY	Dev Year 1	2	3	4	5	6	7	8
2002	1,167	5,377	10,145	16,817	23,801	27,489	31,331	30,715
2003	13,369	33,969	69,915	96,286	114,318	129,862	145,136	
2004	11,392	42,002	76,902	117,726	153,553	187,200		
2005	20,546	51,662	87,578	128,206	160,254			
2006	22,147	54,874	86,696	118,412				
2007	23,313	74,085	118,210					
2008	34,009	69,636						
2009	21,972							

	1	2	3	4	5	6	7
f-hat	263%	172%	144%	126%	118%	112%	98%
sigma^2-hat	193,504	43,014	15,161	12,209	7,450	353	353

Residuals

AY	Dev Year 1	2	3	4	5	6	7
2002	93%	37%	111%	119%	-32%	128%	
2003	-15%	188%	-87%	-130%	-122%	-59%	
2004	155%	70%	127%	92%	118%		
2005	-23%	-14%	34%	-18%			
2006	-32%	-94%	-114%				
2007	115%	-97%					
2008	-149%						

Extended Mack Bootstrap Models

M(0)	Original Mack bootstrap: residuals are independent and identically distributed
M(1)	Pairwise exception resampling on the 3rd and 4th development periods
M(2)	Exception resampling on the 2004 AY, superimposed onto M(1)
M(3)	Exception resampling on the 2005 calendar year, superimposed onto M(2)

Data: Axis Liability Reinsurance

Cumulative Triangle

UWY	Dev Year 1	2	3	4	5	6	7
2003	252	4,626	6,541	7,362	8,780	9,899	14,279
2004	5,290	16,791	19,230	25,196	27,390	30,290	
2005	7,376	23,607	34,388	35,671	40,478		
2006	12,899	31,034	39,681	47,118			
2007	17,758	37,132	47,463				
2008	21,838	40,483					
2009	18,206						

	1	2	3	4	5	6	7
f-hat	235%	130%	116%	112%	111%	144%	100%
sigma^2-hat	16,037	268	329	37	3	0	0

Residuals

UWY	Dev Year 1	2	3	4	5
2003	220%	52%	-15%	120%	123%
2004	52%	-138%	137%	-117%	-70%
2005	63%	163%	-139%	44%	
2006	6%	-27%	41%		
2007	-30%	-30%			
2008	-63%				

Extended Mack Bootstrap Models

M(0)	Original Mack bootstrap: residuals are independent and identically distributed
M(1)	Pairwise exception resampling on the 2nd and 3rd development periods

Data: Axis Marine Insurance Incurred

Cumulative Triangle

	Dev Year							
AY	1	2	3	4	5	6	7	8
2002	23,087	29,866	35,051	34,675	33,947	33,393	33,515	33,225
2003	20,644	25,605	26,341	34,063	35,853	36,344	35,452	
2004	79,663	109,129	109,535	108,057	109,784	109,857		
2005	354,142	446,611	466,813	479,460	475,957			
2006	57,558	81,091	99,884	89,932				
2007	64,850	106,533	124,645					
2008	77,653	97,184						
2009	60,176							

	1	2	3	4	5	6	7
f-hat	132%	108%	101%	100%	100%	99%	99%
sigma^2-hat	1,524	847	881	54	8	14	8

Residuals

	Dev Year					
AY	1	2	3	4	5	6
2002	-12%	61%	-16%	-58%	-131%	102%
2003	-33%	-31%	172%	156%	112%	-98%
2004	37%	-94%	-31%	88%	9%	
2005	-101%	-86%	40%	-68%		
2006	57%	163%	-132%			
2007	226%	111%				
2008	-55%					

Extended Mack Bootstrap Models

M(0)	Original Mack bootstrap: residuals are independent and identically distributed
M(1)	Exception resampling with the 2008 calendar year

Data: Axis Property Insurance Paid

Cumulative Triangle

AY	Dev Year 1	2	3	4	5	6	7	8
2002	75	191	222	275	442	775	864	864
2003	7,151	53,898	73,817	89,525	93,657	93,769	93,819	
2004	50,694	243,313	342,798	357,918	364,078	377,126		
2005	146,865	553,457	776,139	895,229	1,016,912			
2006	50,559	119,826	140,925	152,083				
2007	66,988	136,254	168,461					
2008	127,544	235,995						
2009	48,806							

	1	2	3	4	5	6	7
f-hat	299%	136%	112%	110%	103%	100%	100%
sigma^2-hat	108,201	1,543	929	1,312	161	10	10

Residuals

AY	Dev Year 1	2	3	4	5	6
2002	-1%	-8%	6%	27%	147%	141%
2003	126%	8%	92%	-50%	-83%	-13%
2004	134%	71%	-165%	-155%	37%	
2005	99%	94%	106%	113%		
2006	-45%	-175%	-57%			
2007	-81%	-124%				
2008	-133%					

Extended Mack Bootstrap Models

M(0)	Original Mack bootstrap: residuals are independent and identically distributed
M(1)	Exception resampling with the 2005 accident year

Data: XL Casualty Insurance Incurred

Cumulative Triangle

AY	Dev Year 1	2	3	4	5	6	7	8	9	10
2000	372,664	1,025,600	1,071,250	1,180,111	1,284,152	1,253,885	1,230,820	1,238,218	1,357,193	1,372,758
2001	174,900	555,832	820,393	1,089,914	1,116,841	1,113,948	1,121,906	1,138,232	1,120,433	
2002	185,833	294,177	420,096	432,125	455,133	480,740	461,852	465,418		
2003	133,758	347,168	416,819	451,506	543,818	539,790	540,887			
2004	142,767	291,898	359,316	432,865	435,109	448,392				
2005	179,852	378,188	453,031	542,234	578,000					
2006	127,954	333,830	436,690	514,356						
2007	197,417	346,041	597,923							
2008	160,099	336,203								
2009	148,036									
f-hat	233%	128%	117%	107%	100%	99%	101%	104%	101%	
sigma^2-hat	51,297	22,827	6,320	2,546	649	302	22	7,403	22	

Residuals

AY	Dev Year 1	2	3	4	5	6	7	8	9
2000	120%	-169%	-92%	46%	-119%	-65%	-106%	98%	
2001	165%	103%	198%	-100%	-14%	118%	133%	-102%	
2002	-151%	57%	-122%	-22%	165%	-136%	-35%		
2003	45%	-33%	-74%	198%	-25%	57%			
2004	-51%	-19%	30%	-91%	87%				
2005	-46%	-36%	27%	-5%					
2006	46%	11%	9%						
2007	-121%	186%							
2008	-44%								

Extended Mack Bootstrap Models

M(0)	Original Mack bootstrap: residuals are independent and identically distributed
M(1)	Exception resampling with the 2005 calendar year
M(2)	Exception resampling with the 2005 and 2006 calendar years
M(3)	Exception resampling with the 2002, 2005 and 2006 calendar years
M(4)	Exception resampling with the 2002, 2005 and 2006 calendar years, with parametric bootstrapping