All Actuaries Summit 2026

25 – 27 May 2026, Melbourne

Actuaries Institute.

Beyond our means? A distributional approach to assessing actuarial reserving in general insurance

Prepared by Calise Liu, Andrew Song, Alan Xian

Presented to the Actuaries Institute
2026 All-Actuaries Summit
25-27 May 2026

*This paper has been prepared for the Actuaries Institute 2026 All-Actuaries Summit.
The Institute's Council wishes it to be understood that opinions put forward herein are not necessarily those of the Institute and the Council is not responsible for those opinions.*

Actuaries Institute
ABN 69 000 423 656
Level 34, Australia Square, 264 George Street, Sydney NSW 2000
P +61 (0) 2 9239 6100 | actuaries.asn.au

Abstract

Fundamentally, actuaries are asked to put a number to uncertain future events. We spend a lot of effort quantifying this uncertainty to provide a range of reasonable outcomes. As part of standard valuation work, actuaries will present a central estimate alongside a metric for uncertainty, for example, a quantile of the loss distribution in the form of a risk margin or a risk adjustment.

A core part of the actuarial control cycle is the monitoring and assessment of the models against experience. In practice, this is focussed on the central estimate, with less emphasis placed on the distribution. “Actual vs Expected” analysis assesses the percentage variation against the central estimate, with a layer of actuarial judgement applied to justify the materiality of any differences (e.g. based on portfolio maturity, size of book, known events). This is likely sufficient in many cases but in situations where the modelling decisions/results are less clear cut, we wanted to explore the ability to leverage the distributional thinking already performed at past valuations to provide a better data-driven assessment to support judgement.

It turns out that this isn't as simple as just applying the overall risk margin to the central estimate and backing out some quantiles. This is because:

- The central estimate models are almost always different from the risk margin models
- Risk margins include consideration of systemic risk which may not be in the historical data
- Adjustments in central estimate models (e.g. blending or accounting for data quality/scarcity) means analytical expressions can be hard to derive explicitly
- Allocation - how do you allocate a risk margin calculated for the full projection period, to just the next reporting period?

In this paper, we describe an approach to calculate a distribution for each cell in the projection (or the combination of cells you are interested in). Most importantly, the additional time and effort required to do this analysis is minimal; much of it would already be part of your standard valuation pipeline.

Having access to these distributions opens up a number of areas where "distributional hindsight analysis" can be useful to help support actuarial decision making by adding statistical rigour (because you can't do statistical tests on X% above the mean). Examples of such decisions include:

1. When should we change our roll-forward modelling assumptions?
2. When should we revise our risk margins?
3. How outlying was the new event and does it require a post-balance date adjustment?
4. Should I analyse my new portfolio separately or should I just combine it with my current book?

Keywords: general insurance, reserving, risk margins, distributional hindsight analysis, statistical testing

Contents

1 Introduction

1.1 Background

Actuarial work is fundamentally concerned with quantifying uncertain outcomes. In a reserving context, the valuation of the insurance liabilities requires actuaries to distil the range of possible outcomes into a central estimate, together with an explicit allowance for uncertainty through a risk margin or risk adjustment[1] held in the statutory and financial accounts[2]. These reporting requirements mean that reserving actuaries regularly consider the distribution of the insurance liabilities beyond just the central estimate. However, we observe that this work is largely used only to satisfy external reporting requirements, despite the significant work performed as part of valuation processes in deriving the relevant quantiles.

Once the valuation is complete and a new period of experience has emerged, a core part of the actuarial control cycle is the monitoring and assessment of models against experience. The Actuaries Institute's Professional Standard 302, *Valuations of General Insurance Claims* (PS302), requires actuaries to perform "a comparison by Class of Business of the actual experience to the expected experience implied by the [previous] valuation basis". PS302 allows the actuary discretion in determining which aspects are relevant. While it requires disclosure of known or assumed reasons for deviations, it is not prescriptive about how the "Actual vs Expected" (AvE) comparison should be undertaken.

In practice, current AvE approaches typically focus on the central estimate, with analysis assessing percentage variation against that point estimate. The determination of whether the variation is material is informed by the context of the portfolio, line of business and broader environmental factors. The same percentage variation may be considered significant for one portfolio (e.g. a large, stable portfolio with homogenous risks), but not significant in another (e.g. an immature or rapidly growing portfolio). Actuarial judgement is applied to contextualise and justify the materiality of observed differences, taking into account factors such as portfolio maturity, size of book, and known events.

The assessment of AvE experience on a central estimate basis is likely sufficient in many cases. For example, there may be observed drivers of the experience that are clear in hindsight and have resulted in material deviations from the previous liability estimates. However, where the assessment is less clear cut (for example, where it is not clear whether experience stems from persistent systemic changes, or just due to once-off 'bad luck') , we propose that the distributional information set at the previous valuation, which was used to produce risk margin estimates, can be extended to provide additional objective evidence to support actuarial judgement at the subsequent valuation.

1.2 Statement of contributions

In this paper we develop a framework to perform "distributional hindsight analysis". The framework provides additional tools to support the reserving control cycle by introducing greater statistical rigour and consistency in the assessment of distributional assumptions and the interpretation of experience.

A key contribution of this framework is that it enables hindsight analysis to be performed in a way that is coherent with the previous valuation's risk margin. Constructing an empirical distribution based on observed experience is straightforward (and is typically already

[1] To avoid repetition, the term "risk margins" is used throughout this paper to refer to both risk margins and risk adjustments.

[2] Under AASB17 the risk adjustment is required to be reported as a quantile, even if the estimation approach uses a different basis (e.g. cost of capital approach).

performed in risk margin estimation processes). However, this will primarily capture independent risk and observed systemic risks, missing out on unobserved or future systemic risk components in a given risk margin. Importantly, our proposed methodology looks to explicitly incorporate the systemic risk components not observed in historical data into the distributional analysis.

In addition, risk margin analyses are typically concerned with the variability of the insurance liabilities in aggregate. Our framework provides an approach to supplement the aggregate risk margin outputs with estimates of the variability of outcomes over the next reporting period. This enables the actuary to derive the quantile for the observed experience based on prior modelling, providing additional evidence in assessing how extreme the experience was and providing supplementary empirical evidence to support modelling decisions.

A core objective in our development of this framework is practicality. The approach is designed to leverage existing risk margin estimation processes and the distributional information they generate, thus requiring minimal additional effort to implement.

1.3 Structure of this paper

The remainder of this paper sets out:

- A literature review of existing work on risk margins
- A summary of current industry practice
- Why distributional hindsight analysis is not currently performed as part of best practice
- Our proposed framework and methodology for distributional hindsight analysis, including the underlying theoretical assumptions
- Additional applications of distributional hindsight analysis to support actuarial and strategic decisions
- A worked example.

We also identify a number of possible extensions to this research.

2 Literature Review

There exists substantial actuarial literature that details different approaches to determining risk margins. For Australian and New Zealand practitioners, the 2008 paper by the Risk Margins Taskforce, set up by Actuaries Institute, proposed a risk margins framework which has been largely adopted by the actuarial industry. Under this framework, estimation error (i.e. parameter and model error) and process error is considered under three components:

1. Independent risk - comprising the random component of parameter and process uncertainty.
2. Internal systemic risk - comprising the systemic component of parameter and model uncertainty that is internal to the actuarial process.
3. External systemic risk - comprising the systemic component of process uncertainty that is external to the actuarial process.

Systemic risk in this context is defined as risks that are common across valuation classes or groups and includes risks that are not captured in the historical experience or actuarial models. As a result, the estimation of this risk is often more qualitative than the independent risk component.

There have been several expansions on this framework over time. Some actuaries have developed benchmarks for this framework for specific classes such as Riley and Watson (2009) which details the application of this framework for dust disease liabilities. APRA has also published industry reviews of general insurance risk margins, providing insight into the range of risk margins adopted in the industry across various classes of business.

In the academic sphere, the literature on stochastic loss reserving is vast. Seminal papers include Mack (1994) which proposed a distribution-free stochastic model that underlies the well-known Chain Ladder method. Renshaw and Verrall (1998), England and Verrall (2001) and Wüthrich & Merz (2008) establish statistical frameworks that are consistent with traditional actuarial techniques such as the Chain-Ladder and Bornhuetter-Ferguson methods, allowing quantification of the predictive distribution of reserves. The CAS monograph by Taylor and McGuire (2016) summarises current approaches to applying Generalised Linear Models (GLMs) for loss reserving, which is a common approach with many extensions in the literature.

Interested readers are referred to the detailed summary of commonly used stochastic reserving models produced by the Pragmatic Stochastic Reserving Working Party in 2016 and 2020.

Finally, our review of both industry and academic literature identified few papers discussing hindsight testing of valuation estimates through a distributional lens. An approach is outlined by Houltram (2003) which employs “hindsight re-estimates” (i.e. the amended amount that the actuary would have declared as the estimate of the outstanding claims liability if they had accounted for the newly emerged experience) to produce prediction errors. While this paper has commonalities with our proposed approach, it is not explicitly linked to the current practice framework used to determine risk margins (as the paper was written before that framework was established) and may require additional infrastructure to be set up in order to be implemented.

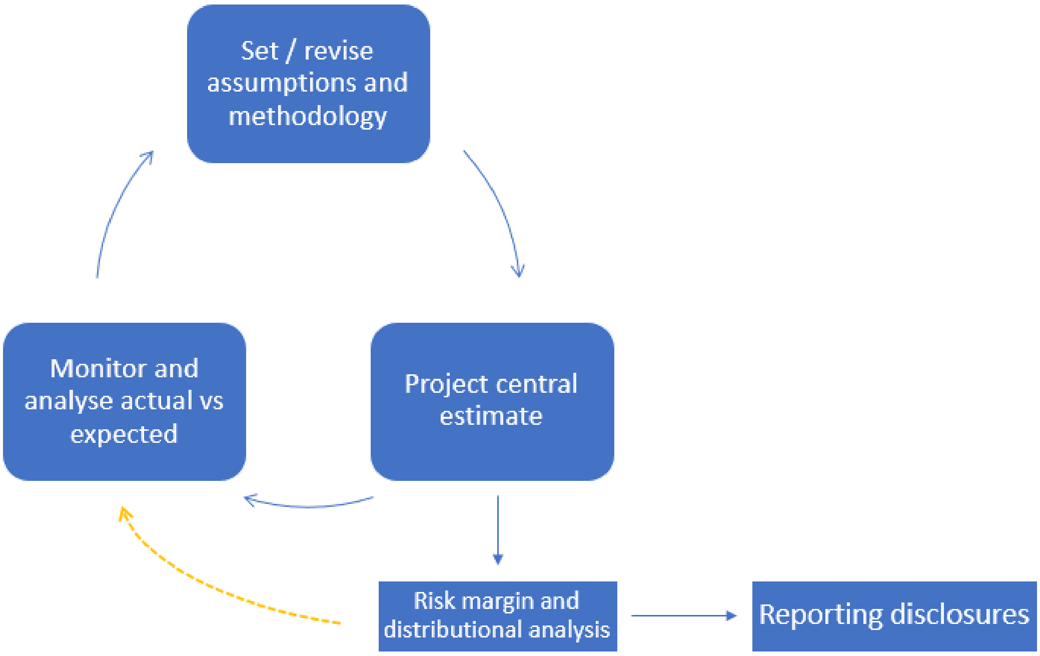
3 Current industry practice

To set the scene and garner greater appreciation of our objective to leverage existing risk margin estimation processes, we provide a short summary of current practices in Australian and New Zealand, which are well established.

3.1 Reserving control cycle

The actuarial control cycle underpins the reserving process, providing a structured framework for the ongoing monitoring and review of models as experience emerges.

Figure 3.1 – Reserving control cycle



A core monitoring tool within this framework is actual vs expected (AvE) analysis, which typically assesses outcomes relative to the central estimate. The significance of deviations from the central estimate is determined by actuarial judgement, taking into account the characteristics of the portfolio, the line of business and the broader environmental context.

As discussed in Section 1.1, the Actuaries Institute's Professional Standard 302, *Valuations of General Insurance Claims* (PS302) requires actuaries to perform AvE analysis at each valuation. The AvE results are used to evaluate model performance and inform whether assumptions require review. We note that the outcomes of the AvE analysis can support a range of reasonable conclusions, particularly when assessing whether observed movements represent emerging experience requiring assumption changes, or is consistent with normal variability.

3.2 Risk margins

Best practice risk margin estimation practices consider independent risk, internal systemic risk and external systemic risk components, using the following approach:

1. Independent risk is measured by investigating the variability in historical experience. We note that this implicitly includes some allowance for systemic risks to the extent

that it is present within the observed experience, although de-trending can be performed for known systemic events to remove their impact. Common assessment methods include bootstrapping residuals of the reserving triangle for a given model, stochastic chain ladder or the Mack-method.

2. Internal and external systemic risk are more difficult to empirically assess. Best practice measurement involves the use of weighted scorecard approaches and/or benchmarks.
3. The coefficients of variation (CoV) for independent and systemic risk are combined to produce an overall CoV using a sum-of-squares approach, assuming independence between these components.

A loss distribution is chosen and used with the CoV to estimate the amounts required to bring a central estimate to a certain probability of adequacy (often 75% in Australia and New Zealand).

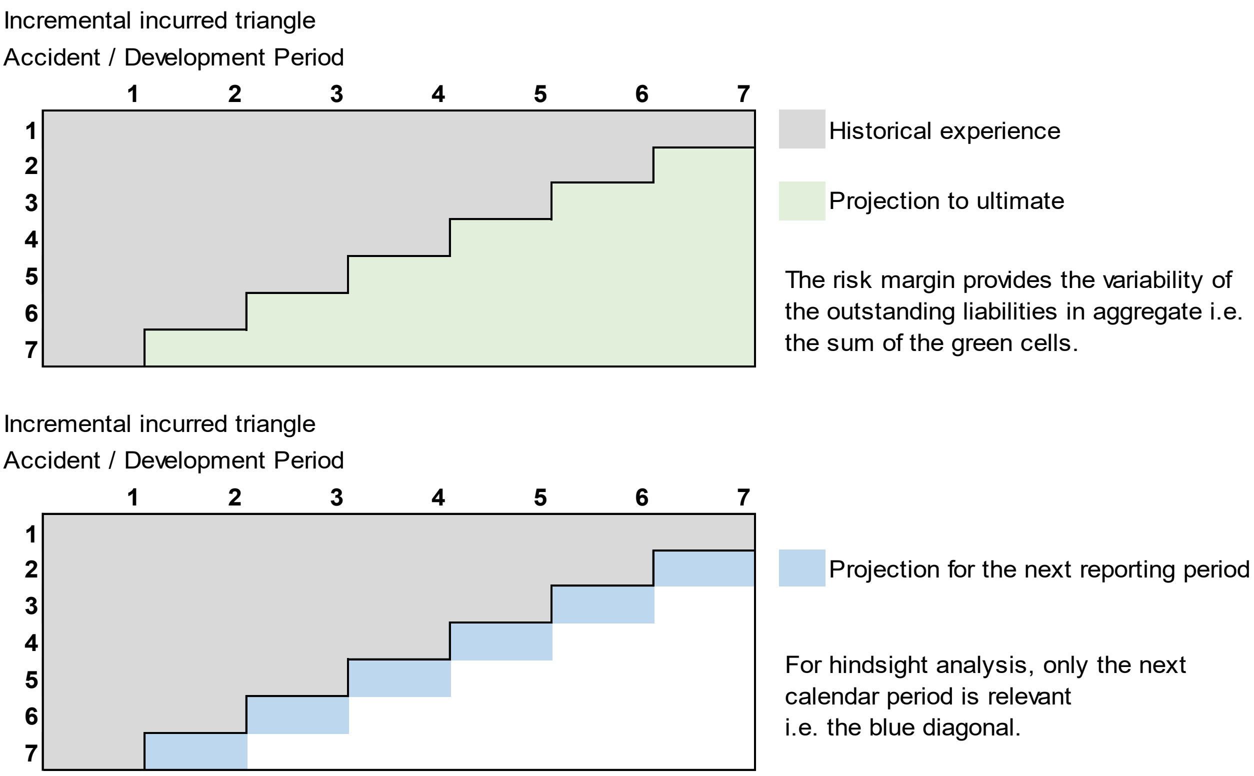
A comprehensive risk margin analysis is typically conducted in practice once every few years, or at earlier intervals when deemed appropriate (e.g. significant portfolio changes or changes in the external environment).

4 Challenges to distributional hindsight analysis

Readers may consider the use of risk margins to derive the implied quantile of the observation to be an obvious extension to AvE analysis. This raises the question of why, in practice, hindsight analysis has largely focused on the central estimate rather than the broader distribution.

Firstly, we observe that a quantitative analysis of the loss distribution is not always necessary to justify material deviations from previous estimates; known drivers can often be identified which explain much of the variation. Secondly, although standard reserving processes produce a distribution for the purposes of risk margin estimation, these are defined at an aggregate level and do not translate directly into distributions for individual reporting periods. More specifically for hindsight analysis, the relevant distribution is that of the next period's experience (i.e. the next diagonal in the lower-right of the triangle), and extracting this in a manner consistent with the adopted risk margin is not a straightforward transformation of the overall risk margin distribution. The difference between the distributions underlying risk margin estimation and those relevant for hindsight analysis is shown in the figure below.

Figure 4.1 – Risk margin vs hindsight analysis distribution



We considered the academic literature on reserving as a potential source of guidance, but found that it offered limited direct applicability in the context of distributional hindsight analysis. We outline the key challenges below.

Firstly, while there is an extensive body of academic literature on stochastic reserving models that generate full distributions, these models necessarily rely on features observable in the data. As a result, they predominantly capture independent risk, with limited recognition of systemic risk only where it has occurred in the past, and do not allow for features not yet observed in the data. Empirical studies by Leong et al (2014) and Meyers and Shi (2011) show that systemic risk is important and material in real world insurance portfolios. In practice, explicit consideration of systemic risks (e.g. reforms, economic/geopolitical shocks, pandemics) can form a significant component of reserving analysis and requires substantial actuarial judgement and effort.

This leads to the second challenge in performing distributional hindsight analysis: reserving models used in practice rarely conform to an exact theoretical model, and are almost always different to the risk margin models. In setting the central estimate, actuaries may adopt a blend of models that best reflect the key features of experience, as well as apply ad-hoc adjustments to allow for external factors or address data limitations. In comparison, the methods used to estimate the risk margins are primarily concerned with distributional properties rather than the accuracy of the central estimate itself, and are therefore typically applied without structural modelling adjustments. It may also be the case that the valuation models may be more granular than the risk margins analysis (e.g. the valuation is performed at finer claim segments or time periods than the risk margins analysis).

As a result, the central estimate adopted in the valuation can differ from the central estimate implied by the risk margin modelling. This inconsistency is managed in practice by deriving a risk margin loading from the risk margins model and applying it to the adopted central estimate, i.e. assuming alignment between the distributional properties of the two models. The central estimate and risk margin quantities ultimately adopted are therefore not anchored to a single, well-defined theoretical model, making the explicit derivation of analytical distributional quantities a challenging exercise.

One approach is to simply allocate the overall risk margin in proportion to the expected central estimate, which is a common approach when performing adhoc analysis on new experience. However, it is important to recognise that this is likely to understate the variability of individual reporting period outcomes, as it doesn't account for the diversification benefits that exist at the aggregate level.

5 Methodology

In this section, we detail our proposed approach to distributional hindsight analysis, addressing the challenges outlined in the previous section. This section is structured as follows:

- Section 5.1 describes the methodology currently used for risk margins estimation, based on the Stochastic Chain Ladder. This:
  - Produces a distribution for the aggregate outstanding claims liability that reflects both independent and systemic sources of risks.
  - Can generate distributions for individual projection periods, which capture features present in the historical data (but may not reflect all sources of systemic risk).
- Section 5.2 describes a proposed algorithm which allocates the overall variability from the risk margins analysis to individual projection periods. The assumptions and advantages underlying this approach are described in Sections 5.3 and 5.4 respectively.
- Section 5.5 describes two extensions to the proposed algorithm which allow modellers to explicitly introduce shocks or correlations into the allocation of the overall variability for individual projection periods.

Selected outputs from a worked example are shown in Section 7, with the workings provided via an Excel file in Appendix A.

5.1 Risk margin estimation using the Stochastic Chain Ladder

Our method leverages the commonly used Stochastic Chain Ladder (SCL) model. We apply the framework from Taylor (2000) in setting up the standard loss reserving as follows.

Let $X_{i,j}$ be the incremental claim amount in accident period i and development period j. Let $C_{i,j}$ be the cumulative claim amounts i.e. $C_{i,j} = \sum_{s=1}^{j} X_{i,s}$. The development factor f_j for $1 \leq j \leq n-1$ is estimated in the standard way as

$$\hat{f}_j = \frac{\sum_{s=1}^{n-j} C_{s,j+1}}{\sum_{s=1}^{n-j} C_{s,j}} ,$$

In the SCL model, it is assumed that these development factors f_j are log-normally distributed i.e.

$$f_j \sim LN(\mu_j, \sigma_j^2)$$

for some mean μ_j and variance σ_j^2. These factors are also assumed to be independent for different periods j. It is a well-known result that the chain-ladder estimator $\hat{f}_j$ above is unbiased.

In practice, the SCL is used to assess the level of independent risk before an overlay is applied with the systemic risk components, ultimately arriving at an overall implied risk margin. We outline this algorithm below:

1. Calculate the independent risk using the Stochastic Chain Ladder model.

   a. Make selections for the mean μ_j and standard deviation σ_j of each development factor f_j based on the observed experience in the upper left triangle.

   b. Using the selections above, simulate the lower right triangle a sufficiently large number of times to obtain a reasonable estimate for the overall independent risk coefficient of variation (CoV).

2. Perform the weighted scorecard or benchmarking approach to estimate the CoVs associated with the internal and external systemic risk.

3. Calculate the overall coefficient of variation including the independent and systemic risk components through a sum-of-squares approach i.e.

$$CoV_{Overall} = \sqrt{{CoV_{Indep}}^2 + {CoV_{Int\ Sys}}^2 + {CoV_{Ext\ Sys}}^2}$$

We then calculate the required quantile(s) by using the calculated mean and overall CoV of the outstanding claims along with an appropriate distribution for the outstanding claims liability (e.g. normal or log-normal) to produce a risk margin for the outstanding claims. Note that for the purposes of setting up the distributional hindsight analysis we are concerned with the undiversified risk margins at the segment granularity of the risk margins analysis. That is, we do not allow for diversification between classes of business, or between the outstanding claims and the premium liabilities.

The above approach provides a distribution of the entire outstanding claims liability. However as outlined in Section 4, for the purposes of distributional hindsight analysis, we instead require the distribution of either individual cells or the newest diagonal of the triangle.

5.2 Proposed approach

Our proposed algorithm for generating distributions in the lower right triangle utilises the calculations already completed in Steps 1 to 3. After these are completed, we instead perform the following steps:

4. Scale up all the standard deviations $\{\sigma_j\}$ selected in Step 1a by a positive constant k. Note the mean remains unchanged.

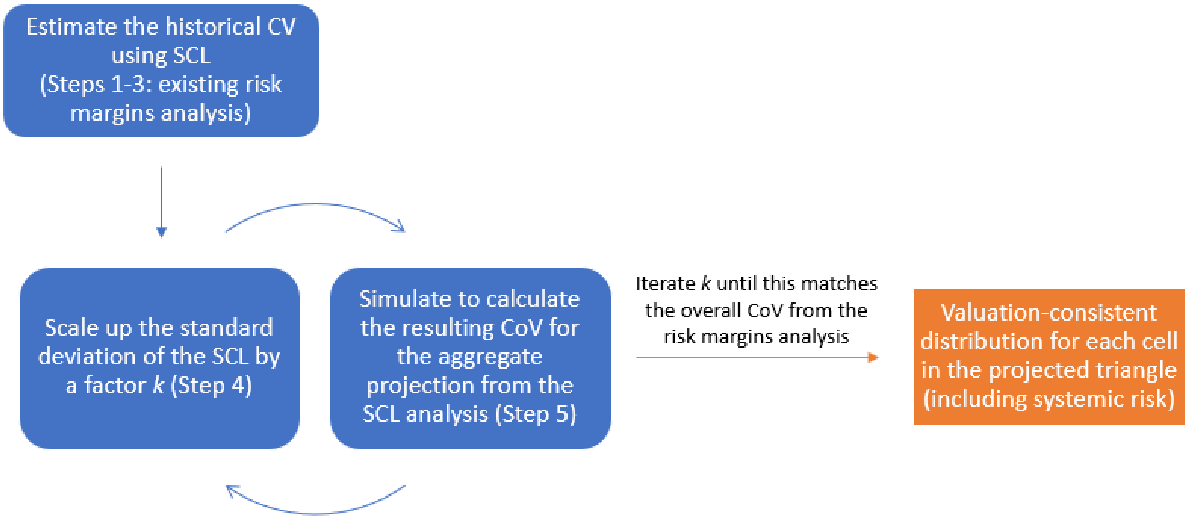
5. Proceed with Step 1b to simulate the lower right triangle and calculate the overall CoV. This should be higher than the original CoV_{Indep}.

6. Iterate Steps 4 and 5 by increasing k until the calculated CoV is equal to $CoV_{Overall}$. This means that the variability in the simulated data has now incorporated the additional variability from the systemic risk components.

7. From the simulations, empirically derive the distributions for the cells of interest (e.g. the latest diagonal or individual cells).

The distributional hindsight analysis framework is summarised in Figure 5.1.

Figure 5.1 – Distributional hindsight analysis framework



The outputs are distributions for individual cells, or groups of cells, in the lower right triangle that are consistent with the selected overall volatility including the impacts of systemic risk which may have been included through other approaches (e.g. balanced scorecard or benchmarking).

5.3 Assumptions underlying our proposed approach

The proposed approach essentially scales up the variability present in the historical data to the overall variability selected for both independent and systemic components, including the variability that may arise from features not yet observed in the data. As a result, the following assumptions are applied:

- Systemic risk is spread across all cells proportionately to the variability that exists in the historical data.
- Systemic risk does not affect the mean of development factors, only the standard deviations.

The implication of the first assumption is that the systemic risk calibrated using the balanced scorecard or benchmarking approach is spread across the data in the same structural manner as what has been observed historically. This may be appropriate in circumstances, for example where the historical data contains both independent and systemic risks. However, an actuary may have explicit views on how the variability of outcomes may vary across the projection period, which differ to the allocation of the independent risk. We propose extensions to the algorithm that enable shocks or correlations to be explicitly introduced into the allocation of overall variability across individual projection periods. These are discussed in Section 5.5.

The second assumption means that the additional systemic risk component will only impact the spread of projected outcomes, rather than the mean. This is appropriate in many cases, since any known regime-shifting impacts would generally already be accounted for in the selection of the means μ_j. However, this assumption can also be relaxed through the introduction of additional scaling parameters (i.e. $k_1, k_2, \cdots$, rather than k) which can be applied to the relevant development factor parameters. This increases the complexity of the calibration, and it may be preferable to perform this step in a more efficient programming environment such as R or Python. Once the calibration is complete, these parameters can be brought back into the Excel infrastructure used for the rest of the valuation.

5.4 Advantages to the proposed approach

Our approach has several advantages that enhance the usefulness and applicability of the outputs. Firstly, it imposes minimal additional overhead to an established valuation pipeline and can be completed quickly even in non-optimal environments for computation such as Excel.

Secondly, the approach produces a simulated distribution for each cell in the triangle, allowing the distribution of any combination of cells to be derived while automatically accounting for diversification and other distributional features. This flexibility can be particularly advantageous where only specific cells are relevant and computational resources are constrained, as only the required outputs (for example, the sum of the next diagonal in each simulation) need to be retained, substantially reducing storage requirements.

Finally, the calibration of the scaling factor k in the algorithm should be simple as it is monotonically increasing with the CoV in most cases. As a result, this task is easy to complete using standard optimisation procedures.

5.5 Extensions: Explicitly incorporating systemic risk

We propose an extension that explicitly models systemic risk within the projections, allowing the framework to accommodate a wider range of modelling circumstances. We decompose the potential impacts of systemic risk on the projections in the following way:

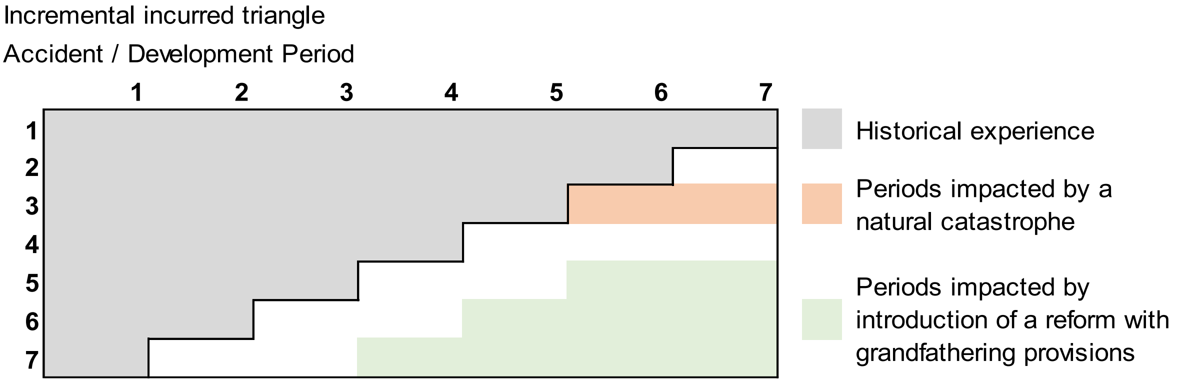
- A discrete, sudden shock occurring at a specific time point in the projection that causes a step-change (affecting some or all cells) from that point onwards. Examples of such shocks include natural catastrophes, legislative reforms, a landmark court ruling or test case outcome, or financial market crises.
- A persistent impact correlated across accident or calendar years which impacts all cells. Examples include super-imposed inflation, long-term changes in the legal environment, and/or cyclic progressions through the various cycles that impact insurance claims (for example, the insurance, business, litigation/tort, medical and reinsurance cycles).

While the base approach in Section 5.2 and the two extensions described in this section allow for the same overall variability, they differ in how this variability is allocated to individual projection periods. Note that an important assumption in our approach is that we assume that the overall risk margin selection is correct in aggregate. As a consequence, any scaling upwards of variability in specific projected cells of the lower triangle will result in a reduction of the variability elsewhere in order to maintain the same overall aggregate variability selected.

5.5.1 Extension A: Discrete systemic shocks

Discrete systemic shocks cause step-changes in the development factors. This extension is well suited to situations where known events are expected to materially alter the distribution of losses, for example past periods impacted by a natural catastrophe, or the introduction of a legislative reform with grandfathering provisions that will affect past accident periods. Figure 5.2 shows how these discrete systemic shocks can emerge in the projection triangle.

Figure 5.2 – Examples of discrete systemic shocks



To explicitly model the variability arising from specified discrete systemic shocks, we propose the addition of a separate shock model which can be optionally included in the algorithm when generating simulations of the bottom right triangle. This model will produce random shocks in the development factors in selected projection periods, in order to simulate these impacts. A simple parameterisation of the model is:

- A probability p indicating the probability that the shock occurs in selected periods.
- A set of parameters $\{s\}$ which represent the impact of the shock on the development factors.

The structure of the cells impacted by the shock, and the probability *p* of the shock occurring, are selected by the modeller based on historical experience or domain knowledge. Possible structures include impacts across an accident period, along a diagonal, or down a development period. The probability *p* is used to randomly simulate (with an outcome of 0 or 1) whether the selected cells are subject to the shock in a given simulation.

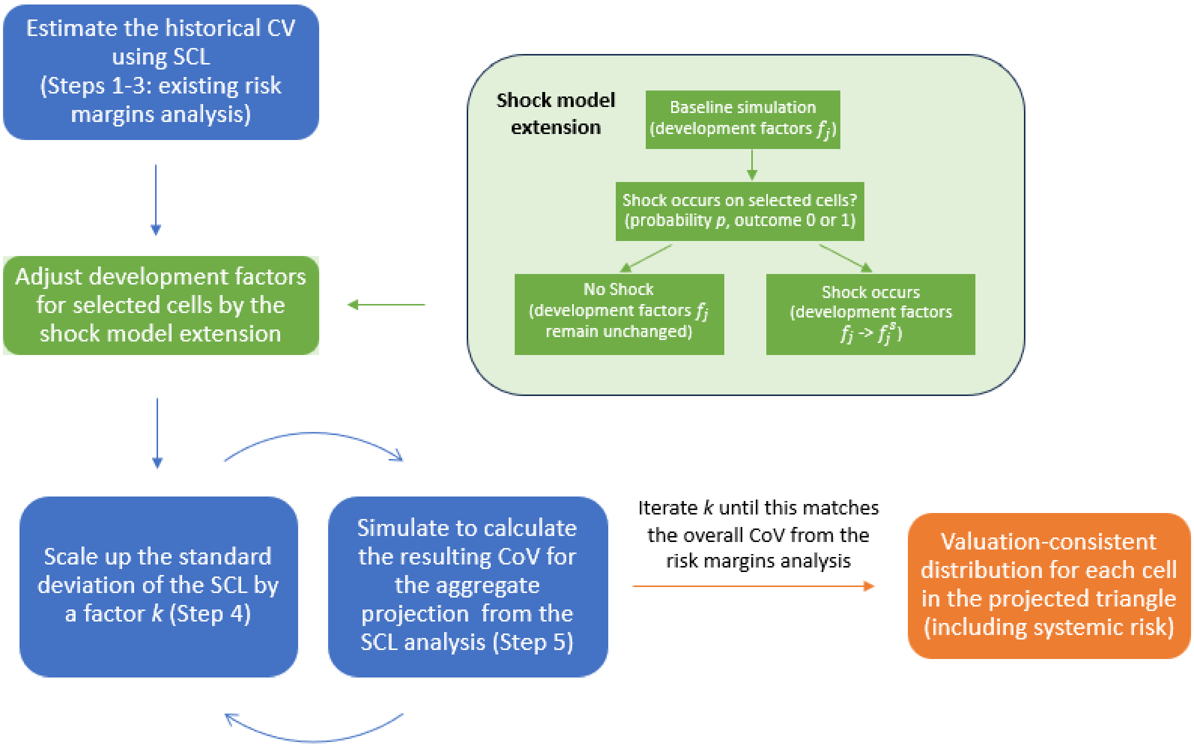
There are many ways to functionally represent the shocked development factor and the selected form will depend on the level of accuracy required by the actuary. Options include:

- A multiplicative relationship: $f_j^s = s \cdot f_j$, for example where a shock is expected to scale the development factor by a fixed multiple;
- A power relationship: $f_j^s = \left(f_j\right)^s$, for example where a shock brings forward multiple periods of development into a single period;
- A distributional relationship: e.g. $f_j^s \sim LN\left(s_\mu \cdot \mu_j, s_\sigma \cdot \sigma_j^2\right)$, allowing the shock to be reflected through changes in both the mean and variability of the development factor.

The parameters $\{s\}$ can be selected by observing the effects of known systemic shocks in the historical data or benchmarking. The simulated nature of our approach provides flexibility in the way that systemic shocks can emerge. Once the shock model has been selected, it can then be incorporated into the algorithm when conducting simulations (i.e. Steps 4 and 5). We expect that this will not add significant computational time, given the simplicity and independence of the shock model from the baseline algorithm.

A summary of the discrete systemic shock framework is below.

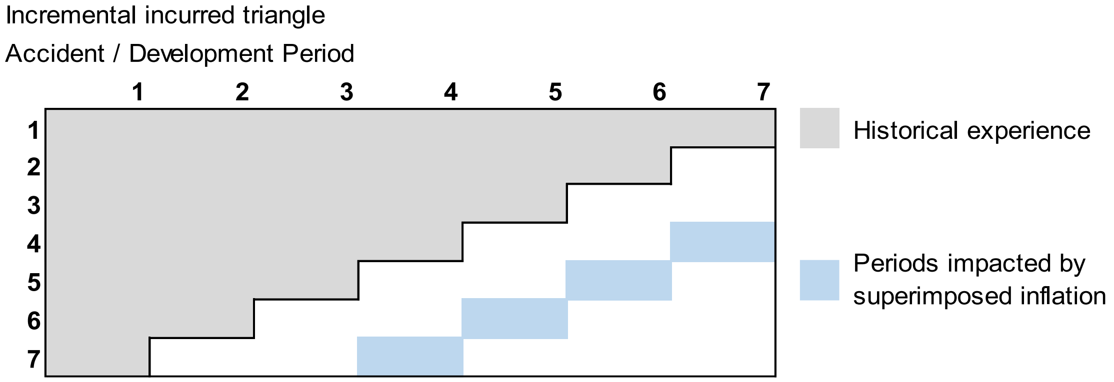
Figure 5.3 – Discrete systemic shock framework



5.5.2 Extension B: Persistent systemic impacts

Figure 5.4 shows how persistent systemic impacts, such as superimposed inflation, can emerge in the projection triangle.

Figure 5.4 – Examples of persistent systemic impacts



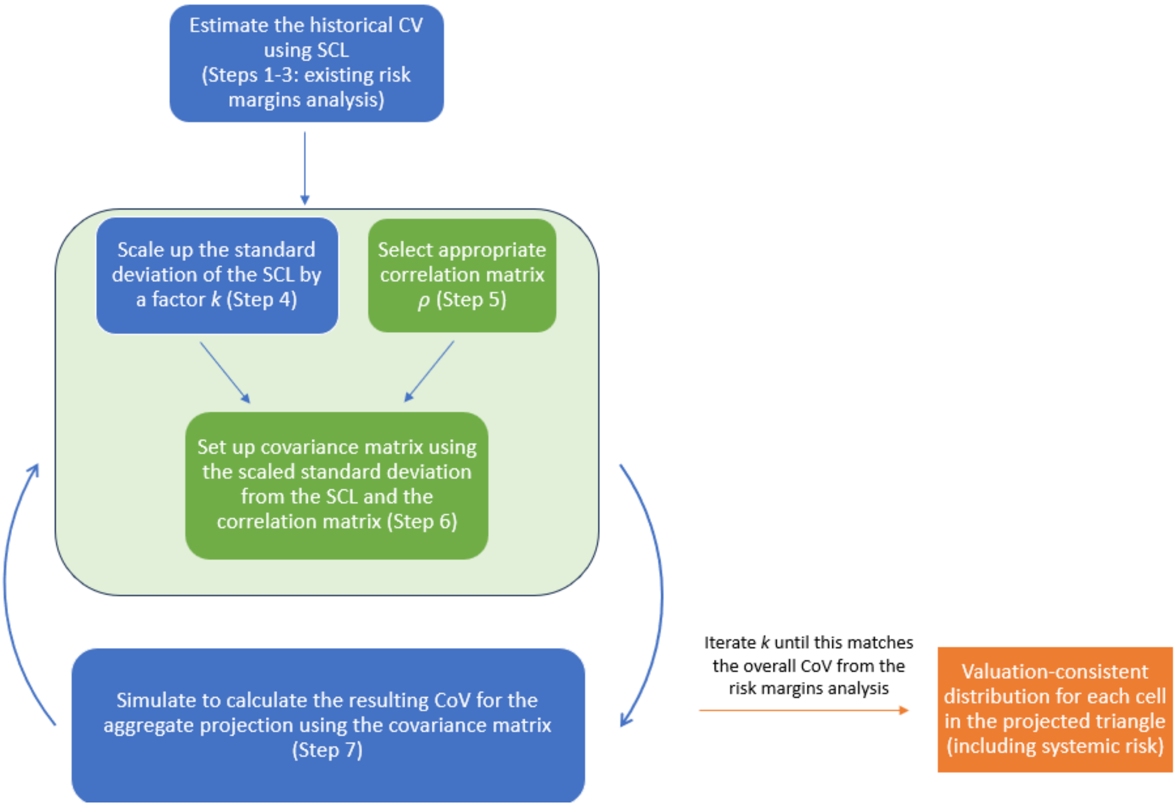
Systemic events can introduce correlation between cells in the projected triangle. To accommodate this correlation in the simulation, we can optionally apply a multivariate approach to the simulation of the projected development factors. Our suggested generalised framework applies a copula to impose a dependence structure on the simulation. There are a vast variety of copulas and many different dependence structures may be accommodated.

For the purposes of this paper, we will assume a Gaussian copula is applied due to the simplicity of the dependence structure as well as consistency with how the industry generally thinks about dependence through the Pearson correlation. Compared with the baseline algorithm outlined in Section 5.2, our extension modifies the steps to accommodate simulating multivariate normal variables:

4. Scale up all the standard deviations $\{\sigma_j\}$ selected in Step 1a by a positive constant k.
5. Select an appropriate correlation factor (or set of correlation factors) ρ which represents the correlation between different cells in the projected triangle.
6. Simulate the lower right triangle using a multivariate approach, using the scaled standard deviations (Step 4) and the selected correlation matrix (Step 5). The statistical details of this approach is as follows:
    a. For each cell in the projected triangle, set up:
        i. A mean vector $\vec{\mu}$ of length n, where n is the number of development factors/cells in the projected triangle.
        ii. A covariance matrix Σ of size $n \times n$ where:
            1. The diagonal entries are $\rho\sigma_r^2$, where σ_r is the standard deviation of the relevant development factor f_r.
            2. The off-diagonal entries are $\rho\sigma_r\sigma_s$, where σ_r and σ_s are the standard deviation of the relevant development factors f_r and f_s.
    b. Calculate the Cholesky decomposition of the covariance matrix Σ. Let this matrix be L.
    c. For each simulation path:
        i. Generate a vector $\vec{Z}$ of n independent standard normals.
        ii. Calculate the multivariate normal vector $\vec{X} = \vec{\mu} + L\vec{Z}$,
        iii. Calculate the multivariate log-normal vector $\vec{Y} = \exp(\vec{X})$
    d. Iterate Step 6c until you have sufficient simulations to provide a reasonable estimate for the overall CoV.
    e. Proceed with Step 1b to simulate the lower right triangle and calculate the overall CoV.
7. Iterate Steps 5 and 6 by increasing k until the calculated CoV is equal to $CoV_{Overall}$. This means that the variability in the simulated data has now incorporated the specified correlation structure.
8. From the simulations, empirically derive the distributions for the cells of interest (e.g. the latest diagonal or individual cells).

A summary of the persistent systemic impacts framework is below.

Figure 5.5 – Examples of persistent systemic impacts



6 Applications and practical considerations

We have identified a number of areas where distributional hindsight analysis can support actuarial and strategic decision making, broadly categorised as reserving control cycle applications and portfolio segmentation applications.

We also call out some practical considerations in the implementation and use of distributional hindsight analysis:

- The framework can produce a distribution for each cell in the bottom right triangle. For robustness, we recommend aggregating by reporting period to derive a calendar-year distribution, from which the quantile of the full year's experience can be calculated. That said, the framework retains flexibility to examine individual accident periods where this is informative, for example to examine the relative extremity of specific known events.
- A comprehensive risk margin review is typically conducted in practice once every few years. Between reviews, the risk margin loading is retained, which assumes that the risk margin and valuation models remain distributionally aligned even as valuation assumptions are updated over time. That is, industry practice assumes that while the mean of the distribution may shift between valuations, the standard deviation and higher moments remain unchanged. We adopt a similar approach, where we calculate the percentage deviation from the mean for each quantile of the loss distribution. This means that for the purposes of distributional hindsight analysis, risk margin reviews do not need to be performed more frequently than current practice.
- An important extension of distributional hindsight analysis is the ability to use formal statistical testing, to assess whether there is sufficient evidence that the assumed loss distribution is no longer appropriate. Relevant techniques involve the use of the probability integral transform (PIT), Kolmogorov–Smirnov (KS) and Anderson–Darling (AD) tests.

  These tests require a sufficient number of data points; we suggest that these can be obtained through back-testing of historical observations against the current assumed loss distribution, rather than waiting for multiple years of future experience to emerge.
- We view this analysis as an addition to the actuarial toolkit, providing greater rigour in areas that currently rely on actuarial judgement. While senior management within insurers are generally familiar with the concept of quantiles (particularly in the context of the insurance liabilities and capital requirements), they remain a technical metric. Actuaries should consider their audience and, where the communication of quantiles is considered appropriate, ensure that context is provided to support the interpretation of this metric.

In the remainder of this section we discuss some applications of this analysis in further detail.

6.1 Supplementing the reserving control cycle

Distributional hindsight analysis can strengthen the reserving control cycle discussed in Section 3.1, feeding into the monitoring of experience and assessment of models and assumptions.

The primary output of distributional hindsight analysis, the quantile of the observed outcome(s), provides a data-driven assessment of how outlying the experience is. This can supplement actuarial judgement when assessing the materiality of AvE deviations and, in turn, inform assessments of valuation assumptions.

We discuss the following reserving control cycle applications below:

- Enhancing current hindsight analysis practices with distributional information
- Assessing the appropriateness of:
    - Valuation assumptions
    - Rollforward valuation models
    - Risk margins
- Adjustment of risk margins for known or specified shock events.

We discuss these reserving control cycle applications below by working through a set of illustrative examples.

6.1.1 Distributional hindsight analysis to assess the 'outlyingness' of the experience

Consider an example where actual experience emerged at 180% of the expected mean, corresponding to the 57th percentile of the previously assumed loss distribution. The actuary has reported that:

> *"The actual experience emerged 80% higher than expected under the previous valuation. While the percentage variation is high, we note that significant variability is expected given the small size of the portfolio, the inherent volatility of the underlying risks, the immaturity of the business…"*

The implication of this commentary is that limited weight is placed on the experience despite its large proportional deviation, due to the characteristics of the portfolio and underlying risks.

Distributional hindsight analysis shows that the actual experience is within the range of expected variation, lying close to the centre of the loss distribution at the 57th percentile rather than in the tail. Despite the large percentage deviation from the central estimate, the experience is therefore not unusual when assessed relative to the underlying risk distribution, supporting the judgement that limited weight should be placed on the deviation for valuation purposes.

6.1.2 Assessing the appropriateness of valuation assumptions

In this example, we consider the potential valuation responses an actuary could make given the additional information provided by the distributional hindsight analysis.

> *"Actual experience emerged at 260% of the expected mean, corresponding to the 80th percentile of the previously assumed loss distribution."*

This outcome can be considered as a borderline case, close to what the APRA risk margin allows for, which is intended to be sufficient in three out of four years. While the distributional hindsight analysis provides additional context, the valuation response remains a matter of actuarial judgement. For example:

- If there are known environmental changes, or other reasons to believe that the result reflects more than random variation (that is, it is not simply "bad luck"), the actuary may choose to respond to the higher experience by adjusting valuation assumptions.

- Alternatively, if the outcome is considered to be within an acceptable range of variability, the actuary may assess this to be short-term randomness and retain existing assumptions.

In the latter case, this raises the question of how many years of adverse experience would be required before a response is warranted. We suggest that statistical testing can be helpful in informing this decision, as discussed in Section 6.

6.1.3 Assessing the appropriateness of rollforward valuation models

Rollforward valuation models (e.g. models used for monthly or quarterly updates) simplify the full reserving model into a smaller number of elements representing the key drivers of experience. In this context, the distributional hindsight analysis framework provides a structured way to assess whether emerging experience is being adequately captured by these models, and whether the assumptions underpinning the roll-forward model remain fit-for-purpose, or instead warrant a more comprehensive review.

6.1.4 Risk margin adequacy

There is currently no widely accepted best-practice approach for assessing the ongoing appropriateness of risk margins. In practice, risk margins are typically updated only every few years, unless exceptional circumstances prompt a review.

While the initial calibration of risk margins is usually based on a structured framework, it remains highly judgement-based. In particular, the balanced scorecard approaches require actuarial judgement to identify material risks, assign relative rankings, and translate those rankings into a coefficient of variation.

Distributional hindsight analysis provides a quantitative way of assessing whether an existing risk margin remains appropriate. In the previous example, where experience emerged at the 80th percentile of the assumed loss distribution, we discussed the interpretation of this outcome in the context of the central estimate.

An alternative explanation is that the central estimate (which is typically well understood through the actuarial reserving process) is reasonable, but the assumed uncertainty around it is understated. Distributional hindsight analysis can help separate questions about the mean from questions about tail behaviour. For example, repeated hindsight observations clustering in the upper and/or lower tail would suggest that the assumed distribution is too narrow, even if the mean remains broadly appropriate. Statistical testing can provide additional empirical support by assessing whether such outcomes are systematic. If so, this may suggest that the risk margin may not be achieving its intended level of sufficiency and therefore a comprehensive risk margin review may be required earlier than typically conducted.

6.1.5 Risk margin adjustments

In circumstances where a known or specified event is expected to materially change the loss distribution, the discrete shock framework described in Extension A can be used to incorporate the anticipated impact of the shock into the risk margins held.

For example, the emergence of a geopolitical conflict may introduce additional variability over one or more future reporting periods due to supply chain impacts or broader economic uncertainty. In such cases, the actuary applies the framework described in Section 5.5.1 to the affected projection periods, adjusting the corresponding loss distributions accordingly. For periods not impacted by the shock, the underlying loss distributions are assumed to remain unchanged.

The projections are then re-simulated to produce an updated aggregate loss distribution that reflects the increased variability in the affected periods, without altering the scaling factor k.

The resulting change in the aggregate distribution provides a quantitative and internally consistent basis for determining the required adjustment to the risk margin.

6.2 Portfolio segmentation

Distributional hindsight analysis can be used to conduct portfolio segmentation exercises, in the context of supporting actuarial modelling decisions as well as operational monitoring.

6.2.1 Actuarial modelling

While portfolios are often aggregated for modelling practicality, this can hide material differences in risk characteristics or divergent trends, and lead to mis-specification of the assumed loss distribution.

Segmentation may be appropriate where different parts of a portfolio are exposed to materially different risk drivers. For example, one segment may exhibit greater seasonality or volatility than another. Applying distributional hindsight analysis at a more granular level can reveal whether experience in specific segments is consistently under- or over-represented within the aggregate distribution, indicating that separate modelling may be warranted.

This approach can be used in situations such as a portfolio transfer, where it can assist the actuary in assessing whether the new portfolio should be analysed separately, or whether it can be credibly combined with the existing book.

Distributional hindsight analysis can also be applied to assess the appropriate time granularity for modelling, for example annual/annual versus annual/quarterly versus quarterly/quarterly triangles, with the key consideration being the trade-off being accuracy (ensuring that trends are not masked), and credibility (volume of data available).

6.2.2 Supporting operational monitoring

Distributional hindsight analysis can be used as part of general portfolio monitoring as an early warning mechanism to identify where emerging experience appears inconsistent with expectations. This can indicate areas where closer oversight is required. Once-off significant deviations should be investigated and understood. Persistent or systemic deviations may indicate underlying issues requiring targeted operational focus.

7 Worked example

In this section, we present a worked example of the distributional hindsight analysis undertaken on a hypothetical portfolio. We have incorporated correlated systemic risk impacts as described in Section 5.5.2 (Extension B). The full worked example in Excel is attached in Appendix A.

7.1 Stochastic Chain Ladder setup

Table 7.1 shows the initial steps required in setting up a standard Stochastic Chain Ladder (SCL), corresponding to Steps 1-3 described in Section 5. It also shows the scaling up of the standard deviation as described in Step 4 of Section 5.2. In this example, we use a scaling factor *k* of 2.7.

Table 7.1 – Standard SCL (Steps 1-3), plus scaling up of the standard deviation (Step 4)

All payments

Accident Year	Development Year 0	1	2	3	4	5	6	7
2018	2,699	1,941	1,087	1,940	310	231	492	0
2019	2,257	2,695	980	593	829	311	132	
2020	2,302	2,431	1,194	439	925	229		
2021	2,843	2,695	916	913	971			
2022	2,032	2,185	899	384				
2023	2,312	1,801	538					
2024	2,587	2,453						
2025	2,636							

Cumulative payments

Accident Year	Development Year 0	1	2	3	4	5	6	7
2018	2,699	4,641	5,727	7,668	7,978	8,208	8,701	8,701
2019	2,257	4,952	5,932	6,525	7,354	7,665	7,798	
2020	2,302	4,733	5,927	6,366	7,291	7,519		
2021	2,843	5,538	6,454	7,367	8,338			
2022	2,032	4,217	5,115	5,499				
2023	2,312	4,113	4,650					
2024	2,587	5,039						
2025	2,636							

Logged loss development factors

Accident Year	Development Year 0	1	2	3	4	5	6	7
2018		0.542	0.210	0.292	0.040	0.028	0.058	0.000
2019		0.786	0.181	0.095	0.120	0.041	0.017	
2020		0.721	0.225	0.071	0.136	0.031		
2021		0.667	0.153	0.132	0.124			
2022		0.730	0.193	0.072				
2023		0.576	0.123					
2024		0.667						
2025								
Selected Mean		0.670	0.181	0.133	0.105	0.034	0.038	
Selected Standard Deviation		0.086	0.038	0.092	0.044	0.007	0.029	

Scaled standard deviation

Selected scaling factor k	**2.7**					
Scaled Standard Deviation	0.234	0.102	0.249	0.119	0.019	0.079

7.2 Extension B – applying a selected covariance matrix

Table 7.2 shows the selected covariance matrix to be applied to the Stochastic Chain Ladder, based on the scaled standard deviation and a selected correlation matrix. This corresponds to Steps 5-6a in Extension B of our proposed approach, as described in Section 5.5.2.

In this example, we have assumed for demonstrative purposes that the development factors in the next reporting period (i.e. the diagonal) are correlated with a factor of 0.05. No correlation is assumed elsewhere in the projection triangle.

Table 7.2 – Set up covariance matrix (Steps 5-6a from Extension B)

B.5 Index *Used to specify which cells are correlated*

Accident Year	0	1	2	3	4	5	6	7
2018								
2019								22
2020							16	23
2021						11	17	24
2022					7	12	18	25
2023				4	8	13	19	26
2024			2	5	9	14	20	27
2025		1	3	6	10	15	21	28

B.6 Selected correlation 0.05

B.7 Selected correlation structure *Cells where correlation apply*

B.8 Index 1 2 4 7 11 16 22 *In this example: next diagonal only*

B.9 Correlation matrix

Index	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	**1.00**	0.05	-	0.05	-	-	0.05	-	-	-	0.05	-	-	-	-	0.05	-	-	-	-	-
2	0.05	**1.00**	-	0.05	-	-	0.05	-	-	-	0.05	-	-	-	-	0.05	-	-	-	-	-
3	-	-	**1.00**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4	0.05	0.05	-	**1.00**	-	-	0.05	-	-	-	0.05	-	-	-	-	0.05	-	-	-	-	-
5	-	-	-	-	**1.00**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
6	-	-	-	-	-	**1.00**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7	0.05	0.05	-	0.05	-	-	**1.00**	-	-	-	0.05	-	-	-	-	0.05	-	-	-	-	-
8	-	-	-	-	-	-	-	**1.00**	-	-	-	-	-	-	-	-	-	-	-	-	-
9	-	-	-	-	-	-	-	-	**1.00**	-	-	-	-	-	-	-	-	-	-	-	-
10	-	-	-	-	-	-	-	-	-	**1.00**	-	-	-	-	-	-	-	-	-	-	-
11	0.05	0.05	-	0.05	-	-	0.05	-	-	-	**1.00**	-	-	-	-	0.05	-	-	-	-	-
12	-	-	-	-	-	-	-	-	-	-	-	**1.00**	-	-	-	-	-	-	-	-	-
13	-	-	-	-	-	-	-	-	-	-	-	-	**1.00**	-	-	-	-	-	-	-	-
14	-	-	-	-	-	-	-	-	-	-	-	-	-	**1.00**	-	-	-	-	-	-	-
15	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**1.00**	-	-	-	-	-	-
16	0.05	0.05	-	0.05	-	-	0.05	-	-	-	0.05	-	-	-	-	**1.00**	-	-	-	-	-
17	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**1.00**	-	-	-	-
18	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**1.00**	-	-	-
19	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**1.00**	-	-
20	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**1.00**	-
21	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**1.00**

A Cholesky decomposition of the covariance matrix is required to induce the desired correlation structure in the simulated development factors. These development factors in turn are used to produce the future payments.

A sufficient number of simulations (20,000 in the example) are run to estimate the CoV on the total simulated outstanding claims. This is part of the standard approach in calculating risk margins using a stochastic chain ladder model.

Following this, the scaling factor *k* is adjusted, creating a new Cholesky decomposition. The simulations are rerun until the simulated CoV equals the overall CoV selected from the risk margins analysis.

We show one simulation of the development factors and future payments, under the assumed correlation structure and under assumed independence, in Table 7.3.

Table 7.3 – One simulation of the development factors and estimated future payments (Steps 6b-6e from Extension B)

Simulated development factors

Correlation applied

Accident Year	Development factors 0	1	2	3	4	5	6	7
2018								
2019								
2020							0.918	
2021						1.035	0.993	
2022					1.082	1.036	1.025	
2023				1.444	0.991	1.035	1.043	
2024			1.061	0.990	1.131	1.020	1.111	
2025		2.628	1.199	0.993	0.980	1.023	1.121	

No correlation applied

Accident Year	Development factors 0	1	2	3	4	5	6	7
2018								
2019								
2020							0.917	
2021						1.041	0.993	
2022					1.088	1.036	1.025	
2023				1.454	0.991	1.035	1.043	
2024			1.056	0.990	1.131	1.020	1.111	
2025		2.631	1.199	0.993	0.980	1.023	1.121	

Forecast payments

Correlation applied

Accident Year	Development factors 0	1	2	3	4	5	6	7
2018	2,699	4,641	5,727	7,668	7,978	8,208	8,701	8,701
2019	2,257	4,952	5,932	6,525	7,354	7,665	7,798	7,798
2020	2,302	4,733	5,927	6,366	7,291	7,519	6,903	6,903
2021	2,843	5,538	6,454	7,367	8,338	8,633	8,573	8,573
2022	2,032	4,217	5,115	5,499	5,953	6,167	6,322	6,322
2023	2,312	4,113	4,650	6,715	6,653	6,887	7,185	7,185
2024	2,587	5,039	5,348	5,295	5,990	6,110	6,790	6,790
2025	2,636	6,927	8,308	8,253	8,087	8,276	9,274	9,274

No correlation applied

Accident Year	Development factors 0	1	2	3	4	5	6	7
2018	2,699	4,641	5,727	7,668	7,978	8,208	8,701	8,701
2019	2,257	4,952	5,932	6,525	7,354	7,665	7,798	7,798
2020	2,302	4,733	5,927	6,366	7,291	7,519	6,899	6,899
2021	2,843	5,538	6,454	7,367	8,338	8,678	8,618	8,618
2022	2,032	4,217	5,115	5,499	5,983	6,198	6,354	6,354
2023	2,312	4,113	4,650	6,761	6,699	6,934	7,234	7,234
2024	2,587	5,039	5,321	5,268	5,959	6,079	6,756	6,756
2025	2,636	6,935	8,318	8,262	8,096	8,286	9,285	9,285

Projections

Correlation applied

Total OSC	11,365
Next reporting period	6,796

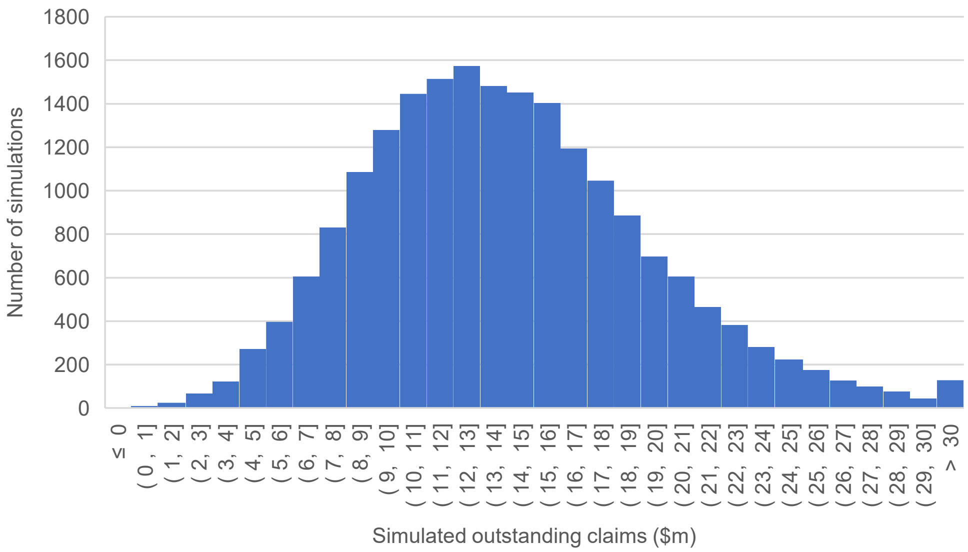
No correlation applied

Total OSC	11,462
Next reporting period	6,893

7.3 Search for *k*

In our worked example, the overall CoV selected for the risk margins is 38.0%. The algorithm produces the same overall CoV of 38.0% when scaling factor *k* = 2.7 is applied. The spread of the simulated outstanding claims is shown in Figure 7.1.

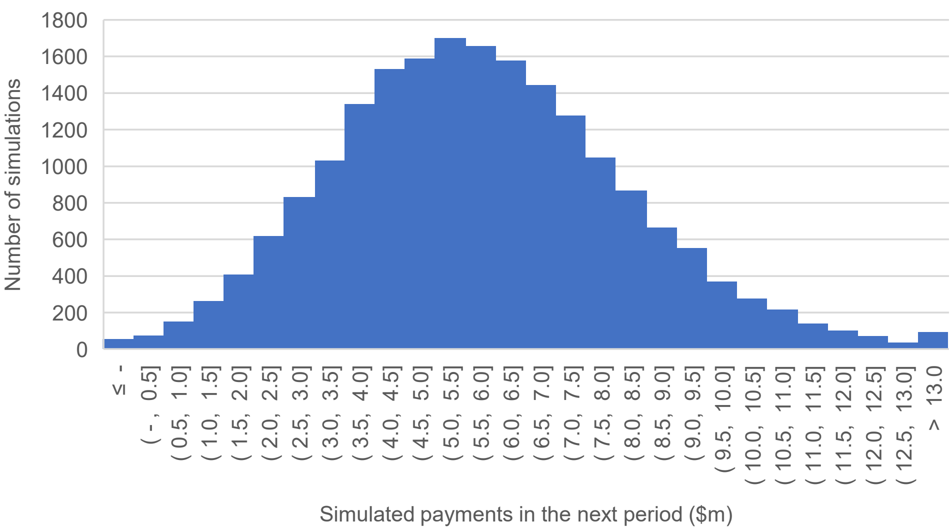
Figure 7.1 – Distribution of simulated outcomes for the OCL



Mean	14,001,975
Standard deviation	5,326,209
CoV	**38.0%**

We are also able to simulate the expected outcomes of the next reporting period (or any group of cells), by summing up the relevant cells of interest.

Figure 7.2 – Distribution of simulated outcomes for the next reporting period



Mean	5,756,864
Standard deviation	2,405,001
CoV	**41.8%**

7.3.1 Impact of specified correlation structure

We note that the inclusion of a positive correlation structure introduces additional variability into the loss distribution for the cells where the correlation is applied. This means that all else equal, a higher *k* is required in a scenario where nil correlation is specified, compared to where correlation is assumed, in order to reach the same desired overall CoV.

7.4 Hindsight analysis

Suppose actual experience for the new reporting period was $8.0m, compared to the $5.8m expected. Assuming a lognormal distribution, this would correspond to the 85th percentile, and may be characterised as a more extreme outcome.

Table 7.4 – Calculating a quantile for hindsight analysis

Loss distributions for the next reporting period

Mean		5,756,864
Standard deviation		2,405,001
CoV		42%
mu	-	0.08
sigma sqrd		0.16

	AvE Payments
Actual	8,000,000
Expected	5,756,864
% Difference	39%
Quantile	85%

Probability of Sufficiency	Next period
50%	-7.7%
60%	2.1%
70%	13.9%
75%	20.9%
80%	29.3%
85%	39.8%
90%	54.3%
91%	58.0%
92%	62.1%
93%	66.8%
94%	72.1%
95%	78.5%
96%	86.2%
97%	96.2%
98%	110.3%
99%	134.6%

In this example, the correlation factor of 0.05 assumed between cells in the new reporting period is small. A higher assumed correlation factor would allocate a greater proportion of total variability to the next reporting period, leading to a lower implied quantile.

As described in Section 6, if the risk margin loadings are retained in successive valuations, then it can be assumed that the relationship between the risk margin and the valuation models remain distributionally aligned (notwithstanding updates to the central estimate of liabilities). That is, in this example, AvE payments of 140% correspond to the 85th percentile in each successive valuation where the risk margin remains unchanged (i.e. where the overall CoV of 38% is maintained).

This holds in our framework where we assume that the development factors are lognormally distributed, as the ratio of the lognormal quantile to the lognormal mean is independent of the mean itself:

$$\frac{\text{LN Quantile } (p)}{\text{LN Mean}} = \exp(\mu + \sigma\Phi^{-1}(p) - \mu - 0.5\sigma) = \exp(\sigma(\Phi^{-1}(p) - 0.5)),$$

where Φ is the CDF of the standard normal distribution.

As a result, the same probability of sufficiency table shown above can be applied directly to the updated central estimates calculated at subsequent valuations in order to repeat the distributional analysis without requiring re-running the calibration and simulation process.

7.5 Statistical testing

The use of statistical testing can add further rigour to distributional hindsight analysis. However, as with most statistical applications, the power of the tests is reduced when sample sizes are small, a limitation that is likely to arise in practice. For example, in the worked example above, only a single quantile observation is generated per calendar period, which limits the ability to obtain statistically significant results in a timely and actionable manner.

There are a number of approaches that could alleviate this issue. One option is to apply back-testing techniques to historical periods, thereby generating additional observations without waiting for further experience to emerge.

Alternatively, it is possible to use the individual cell distributions to generate multiple observations per period, noting that this will change the hypothesis test to consider deviations at the individual cell level rather than for the overall emerged experience. This may be useful in circumstances where there are offsetting movements within the calendar period that may trigger rejection of the null hypothesis when assessed at an individual level, but not when the results are assessed in aggregate.

In the following example, we assume that our proposed approach has been applied to a monthly roll-forward model which has been maintained since the previous annual valuation.

The observed quantiles are as shown in Table 7.5.

Table 7.5 – Observed quantiles by month

Month	Quantile	Month	Quantile	Month	Quantile
July	41%	November	32%	March	44%
August	43%	December	75%	April	40%
September	44%	January	45%	May	43%
October	22%	February	42%	June	83%

We propose two statistical tests using these quantiles, noting that if the distributions are correctly specified then the quantiles should follow a Uniform (0,1) distribution:

1. The single-sample Kolmogorov-Smirnov test (or related tests such as the Anderson-Darling test)
2. An interval-censored Probability Integral Transform test using the chi-squared statistic.

7.5.1 Single-sample Kolmogorov-Smirnov (KS) test

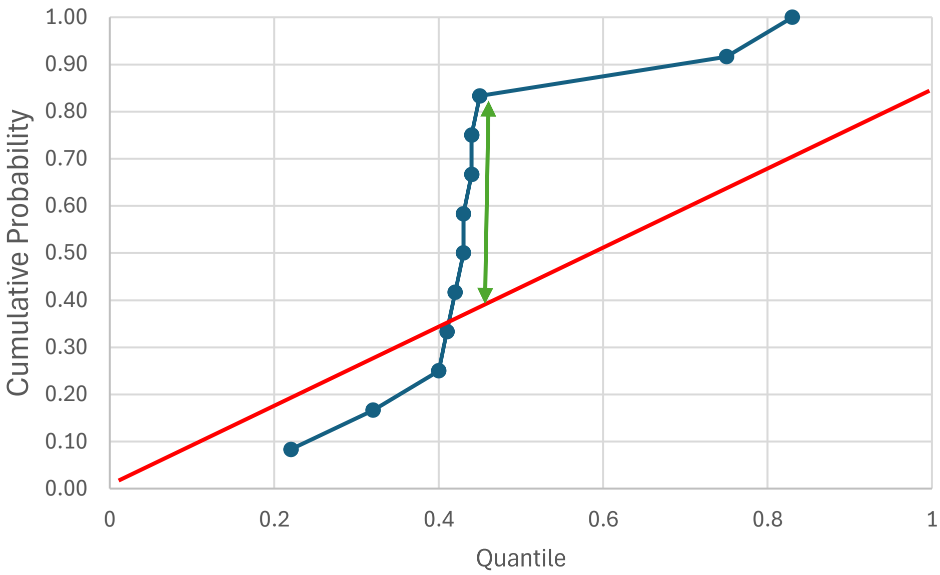
The KS test is a non-parametric goodness-of-fit test which is used to check whether a sample of data is consistent with a specified theoretical distribution. In this case, we are looking to check whether the observed quantiles come from a Uniform(0,1) distribution.

Let $F_n(x)$ be the empirical CDF from the observed n quantiles Q_i i.e.

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{(-\infty,x]}(Q_i).$$

Additionally, let $F(x)$ be the theoretical Uniform(0,1) distribution. The comparison of the two is shown in Figure 7.3 below (and is equivalent to a QQ plot).

Figure 7.3 – Empirical CDF vs Theoretical CDF (Uniform)



The test statistic D_n for the KS test is

$$D_n = \sup_x |F_n(x) - F(x)|,$$

where n is the number of observations, which intuitively is the largest vertical space between the empirical CDF (blue) and the theoretical CDF (red), as indicated by the green arrow.

It is known that D_n follows a Kolmogorov-Smirnov distribution for which critical values are available for different values of n. In our case, the critical value for D_{12} with 0.05 significance is 0.375 and we would reject the null hypothesis if our calculated D_n test statistic was greater than this amount.

From Figure 7.3, it appears that the empirical CDF does not match the theoretical CDF very well. In fact, the actual test statistic is 0.383 and so we conclude that there is sufficient evidence to reject the null hypothesis that the quantiles follow a Uniform(0,1) distribution. This would give some justification for changing the roll-forward assumptions to better account for the different emerging experience over year.

We note that this occurred because there was a substantial grouping of quantiles around the 0.4 to 0.5 range. This may be hard to detect in circumstances where the actuary is just reviewing the AvE experience with regards to the mean and in particular is only looking for significant deviations in each month.

7.5.2 Interval-censored Probability Integral Transform test

Another non-parametric test that can be applied which retains moderate power even with the small number of observations is the interval-censored PIT test. In this test, we define a contingency table by splitting the (0,1) interval into appropriate buckets. One approach is shown below, based on the example in Section 7.5.2 and Table 7.6.

Table 7.6 – Observed quantiles by month

Bucket range	Expected # of quantiles $\boldsymbol{E_j}$	Observed # of quantiles $\boldsymbol{O_j}$
[0,0.25)	3	1
[0.25,0.5)	3	9
[0.5,0.75)	3	0
[0.75,1)	3	2

The expected number of quantiles is calculated by taking the probability range in each bucket (25%) and multiplying it by the number of quantile observations (12).

The contingency table test statistic is

$$X = \sum_{j=1}^{K} \frac{\left(O_j - E_j\right)^2}{E_j} \sim \chi^2_{K-1},$$

where K is the number of buckets selected.

In our worked example, we obtain that $X = 16.67 \gg \chi^2_{3,0.95} = 7.81$ and so we have sufficient evidence to reject the null hypothesis that the observed quantiles follow a Uniform distribution and that our rollforward model is correctly specified. This outcome is consistent with the KS test result.

8 Further work

The framework developed in this paper is intended as a practical and implementable approach to distributional hindsight analysis, leveraging existing risk margin estimation processes. Notwithstanding this, there remain several areas where the work can be extended, which we describe below.

- Future research could develop practical rules of thumb for quantile assessments. Analyses conducted on large public or industry datasets may help identify thresholds at which realised quantiles are considered atypical, by line of business, and therefore warrant further investigation. This analysis could also assist with the calibration of systemic risk benchmarks across the industry.
- There is scope within the framework to extend the dependency structures to incorporate correlation between the outstanding claims liability and premium liability, as well as across lines of business. This enables analysis, particularly in dynamic financial analysis settings, on the assumed correlations underpinning diversification.
- While we have described in this paper some extensions to allow for systemic risks and shocks under the Stochastic Chain Ladder, further work could consider equivalent extensions for Bootstrap-based methods. This would broaden the applicability of this framework across other risk margin setting techniques. We refer interested readers to Houltram (2003) and Taylor and McGuire (2007).
- Sensitivity testing could be undertaken across the different approaches in allocating variability to the cells in the lower-right triangle. A structured comparison of the alternative methods may help industry practice progress to a more consistent and broadly accepted framework for measuring the distribution of and correlation between groups of cells within the triangle.
- Additional research could examine how the framework may be used to evaluate distributional assumptions. This includes assessing forecasting performance over multiple valuation cycles through probabilistic forecasting techniques and supporting integration within automated valuation approaches through facilitating model selection and benchmarking.
- There is scope to investigate more efficient calibration methods for k, including optimal search algorithms or closed-form solutions where available. Such developments may improve computational efficiency and practical implementation, particularly for insurers with a large number of portfolios.

Collectively, these areas represent natural extensions of the framework and may further strengthen its application within the reserving control cycle.

9 Acknowledgements

We thank our colleagues from Finity, Taylor Fry, UNSW Sydney and Macquarie University for their helpful discussions and feedback, which materially improved this paper. In particular we thank John Jeaitani for providing detailed peer review of this paper. Any errors or omissions in this paper remain the responsibility of the authors.

10 References

Actuaries Institute (2024) *Professional Standard 302 Valuations of General Insurance Claims*.

Australian Prudential Regulation Authority (2015) *General insurance risk margins: Industry review report as at 30 September 2013*.

Carrato, A., McGuire, G., and Scarth, R. (2016) *A Practitioner's Introduction to Stochastic Reserving*, Institute and Faculty of Actuaries.

England, P. D., and Verrall, R. J. (2001) *A flexible framework for stochastic claims reserving*, Proceedings of the Casualty Actuarial Society Vol. 88, No. 1, pp.1–38.

Houltram, A. (2003) *Reserving Judgement Considerations Relevant to Insurance Liability Assessment Under GPS210,* Proceedings of the 13th General Insurance Seminar.

Houltram, A. (2005) *An Empirical Approach to Insurance Liability Prediction Error Assessment with Applications to APRA Risk Margin Determination,* Proceedings of the 15th General Insurance Seminar.

Leong, J., Wang, S., and Chen, H. (2014) *Back-Testing the ODP Bootstrap of the Paid Chain-Ladder Model with Actual Historical Claims Data*, Variance Vol. 8, Issue 2, pp.182-203.

Mack, T. (1994) *Measuring the variability of chain ladder reserve estimates*, Casualty Actuarial Society Forum Vol. 1, No. 1, pp. 101-182.

Marshall, K., Collings, S., Hodson, M., O'Dowd, C. (2008) *A Framework for Assessing Risk Margins,* Presented at the 16th General Insurance Seminar.

Meyers, G., and Shi, P. (2011) *The Retrospective Testing of Stochastic Loss Reserve Methods*, *CAS E-Forum*.

Renshaw, A. E., and Verrall, R. J. (1998) *A stochastic model underlying the chain-ladder technique,* British Actuarial Journal, Vol. 4 Issue 4, pp.903–923.

Riley, B. and Watson, B. (2009) *Asbestos Liabilities & the New Risk Margins Framework*, Presented to the Institute of Actuaries of Australia 2009 Accident Compensation Seminar.

Scarth, R., Jain, S., and Roberto, R. (2020) *A practitioner's introduction to stochastic reserving.*, Institute and Faculty of Actuaries.

Taylor, G., and McGuire, G. (2007) *A synchronous bootstrap to account for dependencies between lines of business in the estimation of loss reserve prediction error*, North American Actuarial Journal Vol. 11 Issue 3, pp.70–88.

Taylor, G., and McGuire, G. (2016) *Stochastic loss reserving using generalized linear models*, CAS Monograph No. 3, pp.1-112.

Merz, M., and Wüthrich, M.V. (2008) Modelling the Claim Development Result for Solvency Purposes, CAS E-Forum Fall 2008, pp.542-68.

Appendix A: Excel worked example

The worked example discussed in Section 7 of distributional hindsight analysis are provided in an accompanying Excel file. The example is illustrative only and is based on a hypothetical portfolio. We have incorporated correlation in the next reporting period (i.e. Extension B is applied), to demonstrate the mechanics of the calculations.