Incorporating Model Error into the Actuary's Estimate of Uncertainty

Abstract

Current approaches to measuring uncertainty in an unpaid claim estimate often focus on parameter risk and process risk but do not account for model risk. This paper introduces simulation-based approaches to incorporating model error into an actuary's estimate of uncertainty. The first approach, called Weighted Sampling, aims to incorporate model error into the uncertainty of a single prediction. The next two approaches, called Rank Tying and Model Tying, aim to incorporate model error in the uncertainty associated with aggregating across multiple predictions. Examples are shown throughout the paper and issues to consider when applying these approaches are also discussed.

Keywords

Model uncertainty, model risk, model error, parameter risk, process risk, model variance, parameter variance, process variance, mean squared error, unpaid claim estimate, uncertainty, reserve variability, bias, simulation, scaling, weighted sampling, rank tying, model tying.

Table of Contents

1 Introduction 3
1.1 Background 4
2 Scaling 5
3 Mean Squared Error 6
3.1 Process Variance 7
3.2 Parameter Variance 7
3.3 Squared Bias 8
3.4 Estimating the MSE – Single Model 8
3.5 Estimating the MSE – Multiple Models 9
4 Model Error 11
4.1 User Error 11
4.2 Historical Error 12
5 Incorporating Model Error 12
5.1 Weighted Sampling 12
5.2 Considerations 15
5.2.1 Simulations 15
5.2.2 Individual Model Distributions 15
5.2.3 Lumpiness 16
5.2.4 Assigning Weights to Models 19
5.2.5 Effect on MSE 19
6 Aggregating Variability 19
6.1 Weighted Sampling Revisited 20
6.2 Dependencies 22
6.2.1 Origin Period Dependency – Process Error 22
6.2.2 Origin Period Dependency - Parameter Error 22
6.2.3 Origin Period Dependency - Model Error 22
6.3 Rank Tying 23
6.4 Model Tying 25
6.5 Aggregation Considerations 28
6.5.1 Broken Strings 28
6.5.2 Increasing Complexity 29
6.5.3 Effects on MSE 31
7 Summary 32

1 Introduction

One of the core practices performed by property and casualty actuaries is the estimation of unpaid claims, which according to Actuarial Standard of Practice Number 43 (ASOP 43), *Property/Casualty Unpaid Claim Estimates*, is defined as:

> Unpaid Claim Estimate – The actuary's estimate of the obligation for future payment resulting from claims due to past events.

Estimates by their nature are subject to uncertainty and our profession has strived to communicate the uncertainty inherent in unpaid claim estimates to the users of our services. In the past, communications were mostly verbal in the sense that they warned the user of the risk that the actual outcome may vary, perhaps materially, from any estimate, but were rarely accompanied by a quantification of the magnitude of this uncertainty. More recently, actuaries have developed approaches to measure uncertainty and have included this information in their communications.

ASOP 43 suggests that there are three sources of uncertainty in an unpaid claim estimate.

> Section 3.6.8 Uncertainty – "When the actuary is measuring uncertainty, the actuary should consider the types and sources of uncertainty being measured and choose the methods, models and assumptions that are appropriate for the measurement of such uncertainty…Such types and sources of uncertainty surrounding unpaid claim estimates may include uncertainty due to model risk, parameter risk, and process risk." (emphasis added)

ASOP 43 defines each risk as follows:

> 2.7 Model Risk – "The risk that the methods are not appropriate to the circumstances or the models are not representative of the specified phenomenon."
>
> 2.8 Parameter Risk – "The risk that the parameters used in the methods or models are not representative of future outcomes."
>
> 2.10 Process Risk – "The risk associated with the projection of future contingencies that are inherently variable, even when the parameters are known with certainty."

Common approaches to measuring uncertainty, such as the Bootstrapping approach described by England and Verrall (1999, 2002 and 2006) and England (2001) and the distribution-free methodology described by Thomas Mack (1993), are based on the premise that a single model in isolation is representative of the unpaid claims process, and as a result, uncertainty is measured only for parameter and process risk. We believe that circumstances exist in current practice where model risk is evident in the uncertainty surrounding an unpaid claim estimate, and as a result, this paper introduces methodologies to incorporate its impact. These methodologies leverage existing approaches that measure parameter and process risk by supplementing their results with the inclusion of model risk. Examples are shown throughout the paper that, to the extent practical, are based on a single case study which is discussed in more detail in Appendix A.

1.1 Background

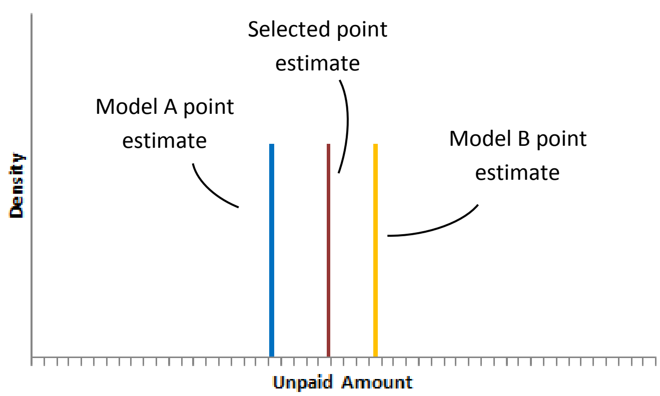
The genesis of this paper and the methodologies presented herein are the result of a dilemma that the authors observed when estimating uncertainty associated with an unpaid claim estimate. This dilemma is perhaps best explained through a hypothetical example.

Consider a hypothetical situation where an actuary uses two actuarial projection methodologies (i.e. models) to estimate unpaid claims for a book of business: Model A and Model B, which both produce a point estimate. Based on the actuary's expertise and professional judgment, the actuary selects the central estimate (colloquially referred to as a "best estimate") to be the straight average of the two point estimates. In other words:

$$Central\ Estimate = \frac{(Model\ A\ Point\ Estimate + Model\ B\ Point\ Estimate)}{2}$$

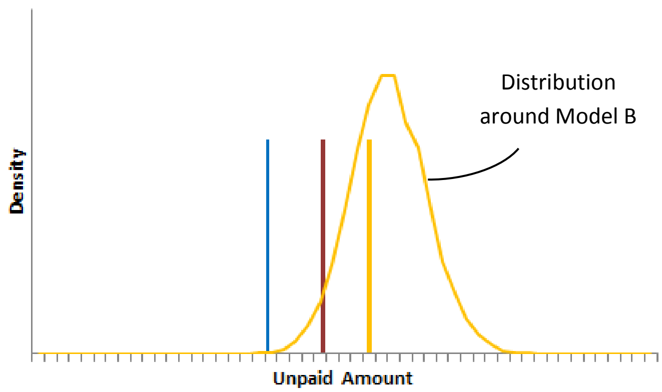
Graphically, these point estimates are shown in Figure 1.

Figure 1. Actuarial central estimate



In order to convey uncertainty in this example, the actuary uses Model B as the basis for estimating uncertainty and observes the following distribution in Figure 2.

Figure 2. Distribution around Model B



If it is assumed that the distribution in Figure 2 is intended to represent the range of uncertainty in the actuary's estimate, then a couple of observations raise concern:

- The actuarial central estimate is not centrally located within the distribution; and
- The distribution implies that the point estimate from Model A is an unlikely outcome, which conflicts with the actuary's professional judgment to equally weight the point estimates from Model A with Model B in selecting a central estimate.

This example is not unique in that it is common for an actuary to estimate unpaid claims with more than one model and it is rare for different models to produce point estimates that are equivalent. Furthermore, current approaches to estimating uncertainty tend to model uncertainty within the context of a single model, which often is not equivalent to the actuary's selected central estimate.

2 Scaling

One approach to dealing with this dilemma is to shift the distribution about Model B so that the mean of the distribution is set equal to the actuary's selected central estimate. This approach, referred herein as scaling, can be done additively, which maintains the same variance, or multiplicatively, which maintains the same coefficient of variation, where:

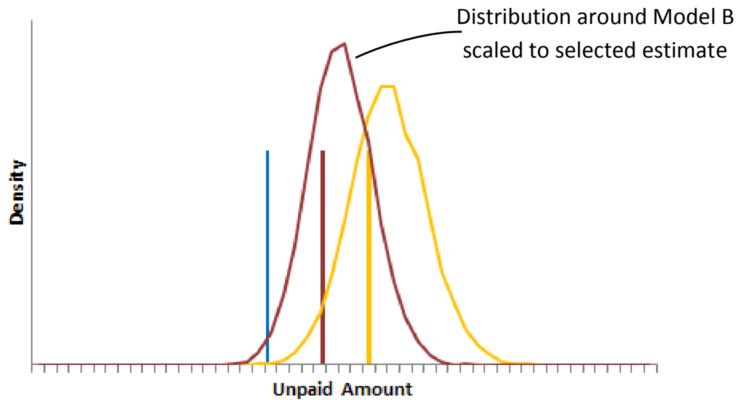
For each point, x_i, within a distribution with mean equal to $\bar{x}$, the corresponding scaled points, x'_i, in the distribution are equal to:

$$Additive\ Scaling: x'_i = x_i + [central\ estimate - \bar{x}]$$

$$Multiplicative\ Scaling: x'_i = x_i \frac{[central\ estimate]}{\bar{x}}$$

Scaling a distribution can be a suitable approach when the magnitude of scaling is immaterial, however, this approach tends to produce unsatisfactory results as the magnitude of the difference between the point estimates increase. For example, consider the hypothetical results before and after scaling multiplicatively to the actuarial central estimate in Figure 3.

Figure 3. Scaling



In this situation, the mean of the implied distribution after scaling reconciles with the actuarial central estimate, however, the point estimate from Model A continues to appear as an outlier. While this example may be an exaggeration, it highlights a dilemma that an actuary faces when the indications from various models diverge.

3 Mean Squared Error

In order to address this dilemma it may be helpful to explore uncertainty in an estimate from a mathematical perspective. [Authors note: The mathematical terms and formulas in this section are used only for the purpose of establishing a theoretical foundation for uncertainty and its relationship with model error. The approaches introduced afterward for incorporating model error do not rely on these formulas and this section of the paper, however, these formulas are believed to be useful for understanding the basic concepts of uncertainty.]

Uncertainty, as used in the context of this paper, implies that the actual outcome may turn out to be different from our estimate (i.e. prediction). In statistics, the Mean Squared Error (MSE) measures this difference. Consider an outcome as a random variable, y and a prediction, $\hat{y}$. The mean squared error is:

$$E[(y - \hat{y})^2]$$

Expanding this term through additive properties yields:

$$E[(y - \hat{y})^2] = E[(y - \hat{y} + (E[y] - E[y]) + (E[\hat{y}] - E[\hat{y}]))^2]$$

Reordering yields

$$E[(y - \hat{y})^2] = E\left[\left((y - E[y]) - (\hat{y} - E[\hat{y}]) + E[y] - E[\hat{y}]\right)^2\right]$$

If we assume y and $\hat{y}$ are independent, then the formula reduces to

$$E[(y - \hat{y})^2] = E[(y - E[y])^2] + E[(\hat{y} - E[\hat{y}])^2] + (E[y] - E[\hat{y}])^2$$

Appendix B derives this formula in more detail. This equation as it is currently structured highlights a key relationship: the mean squared error equals the sum of process variance, parameter variance and squared bias, where:

$$\boldsymbol{Process\ Variance = Var(y) = E\left[(y - E[y])^2\right]};$$

$$\boldsymbol{Parameter\ Variance = Var\ (\hat{y}) = E\left[(\hat{y} - E[\hat{y}])^2\right]; and}$$

$$\boldsymbol{Squared\ Bias = (Bias(y, \hat{y}))^2 = (E[y] - E[\hat{y}])^2}.$$

These terms are discussed further below.

3.1 Process Variance

$\boldsymbol{Var(y) = E\left[(y - E[y])^2\right]}$

The formula for process variance uses the terms y and $E[y]$. The variable y is the actual outcome we are trying to predict, which is presumed to be a random variable that is generated from a distribution with mean equal to $E[y]$. In other words, process variance measures the variance of actual outcomes.

Insurance is believed to be a stochastic process (or nearly stochastic in the sense that the sheer number of conditions which contribute to an actual outcome makes it appear random simply because we are unable to account for all of that information) and the variability inherent in a single outcome occurring is measured by process variance. Consider the flipping of a coin where the probability of a "head" occurring is equal to the probability of a "tail." Despite this knowledge of the underlying probabilities, we are still unable to accurately predict the outcome from a single flip of the coin because there is an element of randomness to any single observation. The estimation of unpaid claims in insurance is similar in that the actual outcome to which we are predicting is a single observation that is one of many probable outcomes which could occur.

3.2 Parameter Variance

$\boldsymbol{Var\ (\hat{y}) = E[(\hat{y} - E[\hat{y}])^2]}$

The formula for parameter variance uses the terms $\hat{y}$ and $E[\hat{y}]$ where the variable $\hat{y}$ is the prediction. Actuaries make predictions of unpaid claims through the application of projection methodologies that attempt to model the overall insurance process using parameters that are estimated from a data sample. Generally speaking, not every point within the distribution of probable predictions from a model is a suitable candidate for an actuarial prediction. Our goal as actuaries is to parameterize the model such that the resulting prediction, $\hat{y}$, is central to the distribution, however, this prediction may not be equal to the true underlying mean of the model, $E[\hat{y}]$, because of our uncertainty in estimating the model's parameters from the data sample. Parameter variance is also called estimation variance because this term of the MSE measures the uncertainty in the estimation of the model parameters.

3.3 Squared Bias

$$(\boldsymbol{Bias}(\boldsymbol{y}, \widehat{\boldsymbol{y}}))^2 = (\boldsymbol{E}[\boldsymbol{y}] - \boldsymbol{E}[\widehat{\boldsymbol{y}}])^2$$

In statistics, a prediction, $\hat{y}$, is considered unbiased if the expected value of the prediction is equal to the expected value of the outcome, y, to which we are trying to predict. Otherwise, statistical bias exists and is measured through this term of the mean squared error. Squared bias is relevant when attempting to estimate the parameters of the MSE, which is beyond the scope of this paper. Some methods of estimation, such as maximum likelihood techniques, may produce biased estimates and will require squared bias to be incorporated into the MSE but for simplicity of discussion we will assume squared bias is equal to zero and we will not address it further in this paper when discussing the MSE.

3.4 Estimating the MSE – Single Model

Although the formula for the mean squared error provides theoretical insights into the components of uncertainty in a prediction, it remains a quandary to apply in an actuarial context since it requires us to be able to measure statistical properties (namely mean and variance) of outcomes that could occur, which are unknown. In many industries, the statistical properties of actual outcomes can be derived by observing a sufficiently large number of trials, but unfortunately, the unpaid claim process is not a repeatable exercise.

One way actuaries have dealt with this predicament is by estimating uncertainty on the condition that a particular actuarial projection methodology (i.e. model) in isolation is representative of the random variable, y. In other words, if the unknown distribution of probable outcomes, $f(y)$, is defined by the distribution of probable predictions from Model A, represented as $f_A(y)$, such that:

$$f(y) = f_A(y)$$

then

$$[MSE|f(y) = f_A(y)] = E[(y_A - \hat{y}_A)^2]$$

where,

y_A is the actual outcome, y, generated from Model A, and

$\hat{y}_A$ is the prediction, $\hat{y}$, from Model A.

Under this conditional assumption, process variance can be defined as the distribution of probable outcomes generated from Model A and parameter variance can be defined as the variance in actuarial estimates generated from Model A.

An interesting observation is that the distribution of uncertainty corresponding to the MSE represents a range that is at least as wide and most likely wider than the range of probable outcomes (i.e. process variance) since it must also incorporate the uncertainty associated with the actuary's estimate of the model's parameters (i.e. parameter variance). In other words the distribution of uncertainty, such as the one shown for Model B in Figure 2, represents the actuary's estimate of potential outcomes conditional on the particular model (i.e. process variance) and the data sample used to estimate the model's parameters (i.e. parameter variance).

3.5 Estimating the MSE – Multiple Models

In isolation, a distribution derived from a single model has intuitive appeal since it represents the only information available. In practice, however, it is uncommon for an actuary's analysis of unpaid claims to be comprised of evaluating only a single model in isolation. ASPOP 43 states:

> Section 3.6.1 Methods and Models – "The actuary should consider the use of multiple methods or models appropriate to the purpose, nature and scope of the assignment and the characteristics of the claims, unless in the actuary's professional judgment, reliance upon a single method or model is reasonable given the circumstances. If for any material component of the unpaid claim estimate the actuary does not use multiple methods or models, the actuary should disclose and discuss the rationale for this decision in the actuarial communication."

Therefore, if multiple models are utilized by the actuary to estimate unpaid claims it seems prudent that the measure of uncertainty recognize the additional knowledge gained from the application of more than one model. As previously hypothesized in Section 1.1, if an actuary uses two models to estimate unpaid claims for a book of business, Model A and Model B with corresponding distributions of probable predictions that could be used to define the distribution of outcomes, $f_A(y)$ and $f_B(y)$ respectively, then two alternatives for estimating the MSE are:

$$[MSE|f(y) = f_A(y)] = E[(y_A - \hat{y}_A)^2]$$

$$[MSE|f(y) = f_B(y)] = E[(y_B - \hat{y}_B)^2]$$

However, it is very likely that

$$f_A(y) \neq f_B(y)$$

and hence the actuary is left with two conflicting solutions for the MSE in this example. If both models are believed to be reasonable representations of $f(y)$, then it may not be appropriate to assume that only one is representative of $f(y)$ because of the ramification it implies with the other model.

$$If\ f(y) = f_A(y), then\ f(y) \neq f_B(y)$$

And likewise

$$If\ f(y) = f_B(y), then\ f(y) \neq f_A(y)$$

Perhaps both models are reasonable representations of $f(y)$ but each model suffers from some unknown function of inaccuracy that we will characterize as model error, such that

$$Model\ Error\ of\ Model\ A = \xi_A = f(y) - f_A(y)$$

$$Model\ Error\ of\ Model\ B = \xi_B = f(y) - f_B(y)$$

Then the introduction of model error can be used to explain the inconsistency between models:

$$f(y) = f_A(y) + \xi_A = f_B(y) + \xi_B$$

Unfortunately, we revert to the predicament of defining uncertainty with unknown terms since model error is unknown. If we use Model A and its corresponding model error to define the distribution, $f(y)$, then:

$$[MSE|f(y) = f_A(y) + \xi_A] = E[(y_A - \hat{y}_A)^2] + \Xi_A$$

is equal to

$$[MSE|f(y) = f_B(y) + \xi_B] = E[(y_B - \hat{y}_B)^2] + \Xi_B$$

where Ξ_A represents the unknown inaccuracy in the MSE as a result of model error in Model A (i.e. ξ_A) and Ξ_B represents the unknown inaccuracy in the MSE as a result of model error in Model B (i.e ξ_B).

If the distribution of uncertainty reflects the uncertainty in outcomes defined by a particular model (i.e. process variance) and the uncertainty associated with estimating that model's parameters (parameter variance) it seems reasonable to incorporate the additional uncertainty associated with the potential error in the underlying model (i.e. model error). Otherwise, the actuary's estimate of uncertainty may be incomplete.

Model error and its corresponding impact on the MSE are both unknown, however, as a general rule the actuary strives to minimize model error. Nevertheless, some model error may remain because it is not possible or practical to identify and correct for it. In the context of selecting a central point estimate, the actuary must choose a single number and oftentimes that number will be based on a weighted average of the reasonable indications from multiple models rather than being set equal to the estimate from any single model. The philosophy underlying this approach, which is akin to hedging one's bet, is that a weighted average of models results in a corresponding unknown model error that is **preferred** to relying on the unknown model error of any single model.

This same philosophy is proposed as our approach to incorporating model error into the actuary's distribution of uncertainty. Revisiting our previous hypothetical that an actuary uses two models to estimate unpaid claims for a book of business, Model A and Model B, and after minimizing model error in Model A and Model B to the extent appropriate the actuary uses expertise and professional judgment to assign weight to the point estimates from these models in accordance with their perceived value as a reasonable predictor such that:

$$Central\ Estimate = w\hat{y}_A + (1 - w)\hat{y}_B$$

where

$$0 \leq w \leq 1;$$

$$\hat{y}_A = the\ prediction\ from\ Model\ A;and$$

$$\hat{y}_B = the\ prediction\ from\ Model\ B$$

Then, the MSE and corresponding distribution of uncertainty expressed as a weighted average of predictions from Model A and Model B where each model is separately considered in isolation as representative of the random variable, y,

$$[MSE|f(y) = wf_A(y) + (1-w)f_B(y)]$$

is preferred to the MSE and corresponding distribution conditional only on Model A

$$[MSE|f(y) = f_A(y)]$$

or the MSE and corresponding distribution conditional only on Model B

$$[MSE|f(y) = f_B(y)]$$

if the unknown model error inherent in this weighted averaging of models, $w(\xi_A) + (1-w)(\xi_B)$, is preferred to relying solely on the unknown model error inherent in Model A,ξ_A, or the unknown model error inherent in Model B,ξ_B.

It should be noted that the word "preferred" is used rather than a mathematical relationship such as "less than" in the context of this discussion because this is a philosophical approach. Ideally, we wish to develop a solution that eliminates model error but in the absence of being able to do so, a reasonable alternative is to attempt to recognize our uncertainty in whatever model error remains.

4 Model Error

Before progressing further, it may be helpful to differentiate model error from other types of error. Previously, model risk was defined as "the risk that the methods are not appropriate to the circumstances or the models are not representative of the specified phenomenon."

Many actuarial projection methodologies (i.e. models) can be shown to have no model error when applied in a controlled environment under specific limitations; however, these conditions rarely exist, if at all, in practice. For example, the approach used to extrapolate link ratios into the "tail" of a traditional chain ladder model can introduce model error. An important point to make about model error is that its resulting bias on the actuary's prediction, if any, should be unknown.

4.1 User Error

User error is different from model error. User error occurs when actions, or inactions, of the actuary lead to the **expectation** that the resulting prediction will be biased high or low. Generally accepted actuarial practice is based on the presumption that an actuary's work product is void of significant or material user error, and hence this type of error should not be incorporated as a component of uncertainty in the actuary's estimate.

4.2 Historical Error

Implicit within most actuarial projection methodologies is the assumption that observations of patterns and trends in the past are indicative of patterns and trends in the future, but future conditions can change and result in materially different processes and outcomes that are often too speculative to estimate. This type of error is a subset of model error and while some changes to future conditions may be reasonably estimable and therefore can be incorporated as an element of uncertainty within the MSE, actuaries oftentimes consider this type of error to be out of scope of their analysis. If so, then the approaches discussed herein will also exclude uncertainty associated with this type of error.

Regardless of the type of error that may exist in a prediction, a goal should be to minimize error within each model to the extent appropriate. Unfortunately, model error often still exists and should therefore be incorporated into the actuary's estimate of uncertainty.

5 Incorporating Model Error

At this point we are ready to introduce a methodology for incorporating model error into an estimate of uncertainty. Various suitable methods exist for estimating the MSE conditional on a single model in isolation so it will be assumed that this analysis has already been performed for each model relied upon by the actuary to derive the central point estimate. This methodology is a simulation-based approach as opposed to a mathematical approach aimed at computing the formulas discussed previously and is perhaps best described through a simplistic example.

5.1 Weighted Sampling

Consider a single actuarial central estimate, $\hat{y}$, to be based on a 50%-50% weighting of estimates produced from two projection methodologies, Model A and Model B, such that:

$$\hat{y} = \sum_{m=A,B} w_m \hat{y}_m$$

Where,

$$\hat{y}_A = the\ prediction\ from\ Model\ A;$$

$$\hat{y}_B = the\ prediction\ from\ Model\ B;$$

$$w_A = 0.5;\ and$$

$$w_B = 0.5$$

Assume that two distributions of the MSE conditional on Model A and separately for Model B are already estimated and that each distribution is comprised of a series of 10 simulations where each simulation, denoted x_i, is shown in Figure 4.

Figure 4. Single prediction model simulations

Model A Simulations

Sim	Value
1	3.4
2	2.5
3	1.8
4	3.8
5	4.4
6	3.0
7	2.0
8	6.0
9	3.7
10	6.4

Model B Simulations

Sim	Value
1	3.6
2	4.6
3	5.2
4	4.4
5	3.4
6	3.6
7	4.4
8	3.9
9	3.4
10	3.0

E.g. simulation x_5 from Model A equals 4.4

A distribution reflecting the inclusion of model error can be estimated by taking a weighted sample without replacement of simulations from Model A and Model B in accordance with their weights. To accomplish this with the example given above, we first create a matrix where we use the weights as the basis for sampling between Model A and Model B for each of the 10 simulations. Because this matrix defines which model to sample for each simulation, we will refer to it as a "Model Matrix," which is shown in Figure 5.

Figure 5. Single prediction Model Matrix

	Wt
Model A	50%
Model B	50%

Model Matrix

Sim	Method
1	B
2	A
3	A
4	B
5	A
6	A
7	B
8	B
9	A
10	A

Once a Model Matrix is created, we select the value corresponding to the simulation number and model to create a series of sampled simulations, which are shown in Figure 6.

Figure 6. Single prediction sampled simulations

Model A Simulations

Sim	Value
1	3.4
2	2.5
3	1.8
4	3.8
5	4.4
6	3.0
7	2.0
8	6.0
9	3.7
10	6.4

Model B Simulations

Sim	Value
1	3.6
2	4.6
3	5.2
4	4.4
5	3.4
6	3.6
7	4.4
8	3.9
9	3.4
10	3.0

	Wt
Model A	50%
Model B	50%

Model Matrix

Sim	Method
1	B
2	A
3	A
4	B
5	A
6	A
7	B
8	B
9	A
10	A

Sampled Simulations

Sim	Value
1	3.6
2	2.5
3	1.8
4	4.4
5	4.4
6	3.0
7	4.4
8	3.9
9	3.7
10	6.4

If we increase the number of simulations in this example to a larger sample size the MSE of the resulting distribution can be estimated by computing the variance of the simulations and the mean of the resulting distribution will be equal to the actuarial central estimate.

Figure 7 shows the results of the distribution before and after incorporating model error when the number of simulations in this example is increased to 10,000.

Figure 7. Single prediction weighted sampling

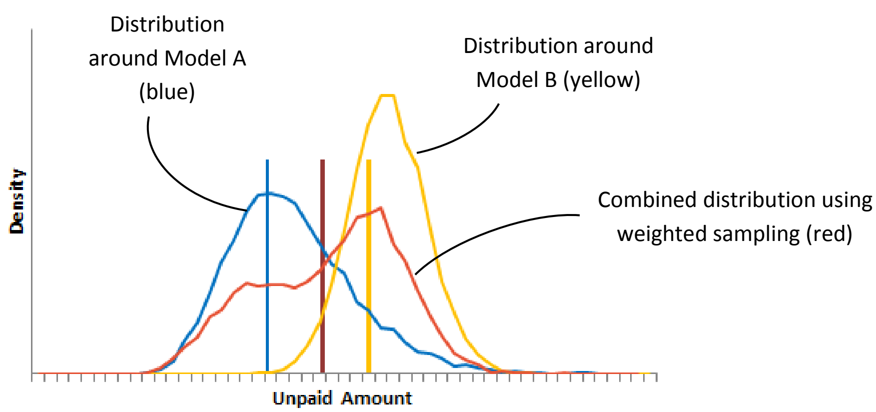
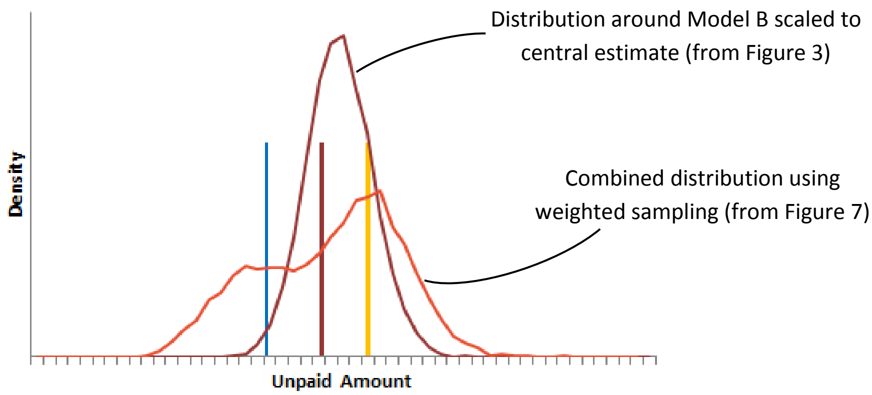


Figure 8 compares weighted sampling in this example to multiplicative scaling Model B's simulations to the central estimate.

Figure 8. Single prediction weighted sampling versus multiplicative scaling



5.2 Considerations

Before we progress the methodology further, it is worth discussing a few points about this approach thus far.

5.2.1 Simulations

It should be noted that in this example, Model B is generated 4 times and Model A is generated 6 times in the Model Matrix. Ideally each model would have been generated an equal number of times since the weighting between the models were equal but the low sample count has led to sample error. For statistically significant sample sizes, we would expect each model in this example to be generated close to 50% of the time.

Sample error must also be considered when evaluating the resulting distribution. Although there is no single number of simulations that is suitable for every circumstance, the user should incorporate a sufficient number to adequately represent the range of potential outcomes, especially if the user is interested in evaluating outcomes generated for extreme tail probabilities.

5.2.2 Individual Model Distributions

Weighted sampling assumes that a distribution of the MSE reflecting the combined effects of process variance and parameter variance is already developed for each model in isolation. Various approaches to estimating the distribution and deriving simulations exist in the literature and example approaches include but are not limited to:

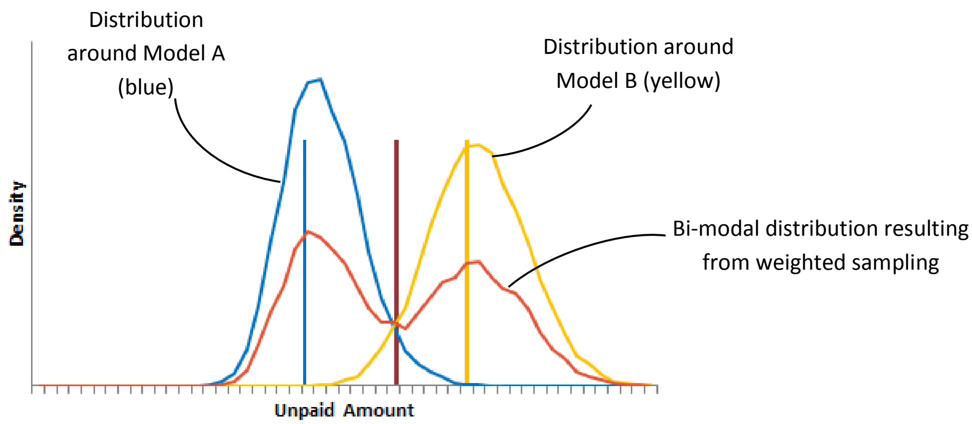
- Simulated approaches – Bootstrapping, Markov-Chain Monte-Carlo simulation or straightforward simulation of outcomes from an assumed distribution using benchmark statistical properties, for example, can be used;
- Analytical approaches – The methodology presented by Thomas Mack is an example of approaches that estimate the statistical properties underlying a model. From this, the user can simulate outcomes once a distributional form is selected; and

- Replicating and scaling – Simulations generated for a particular model can be scaled, either additively or multiplicatively, to the mean of a different model such that an implied distribution of the different model is approximated.

5.2.3 Lumpiness

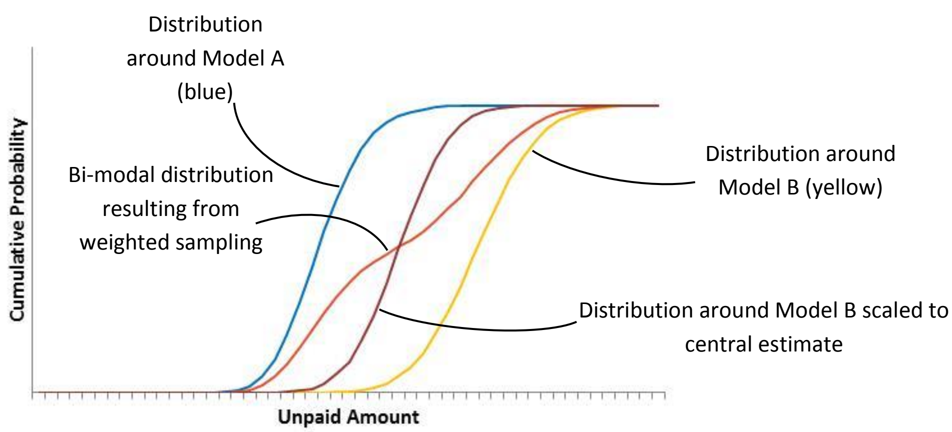
In practice, the user may find the resulting probability density function from weighted sampling to be lumpy, in that there may be multiple modes to the distribution. Figure 9 shows a comparison of weighted sampling from two underlying distributions.

Figure 9. Multi-mode distribution



As a result, it may be challenging to interpret relative probabilities associated with particular outcomes but it is less of an issue when evaluating probabilities associated with a range of outcomes as shown by the corresponding cumulative probability density function for the same example in Figure 9, shown as Figure 10 (also shown in Figure 10 is the distribution around Model B scaled to the selected central estimate).

Figure 10. Multi-mode cumulative probability function



If the shape of the probability density function resulting from weighted sampling is determined to be problematic, the following adjustments could be made:

- Compute the indicated coefficient of variation from the resulting lumpy distribution and re-simulate a newly defined distribution with the same mean and coefficient of variation. Figures 11 and 12 show an example where the lumpy distribution was re-simulated using a Gamma distribution with the same mean and coefficient of variation. It should be noted that a potentially undesired consequence of this adjustment is that probabilities associated with various outcomes within the distribution will be different.
- Probabilities within the range of outcomes where the nodes occur can be re-distributed according to some user-selected smoothed distribution, such as a uniform distribution. An advantage of this adjustment approach is that tail probabilities are unaffected. Figures 13 and 14 show an example of this approach with the probability density graph and the cumulative probability graph, respectively. Note that the actuary should use caution with this approach and be aware that in achieving a more intuitive 'shape' to the distribution, the mean and the coefficient of variation should be maintained.

Figure 11. Re-simulated distribution – probability density

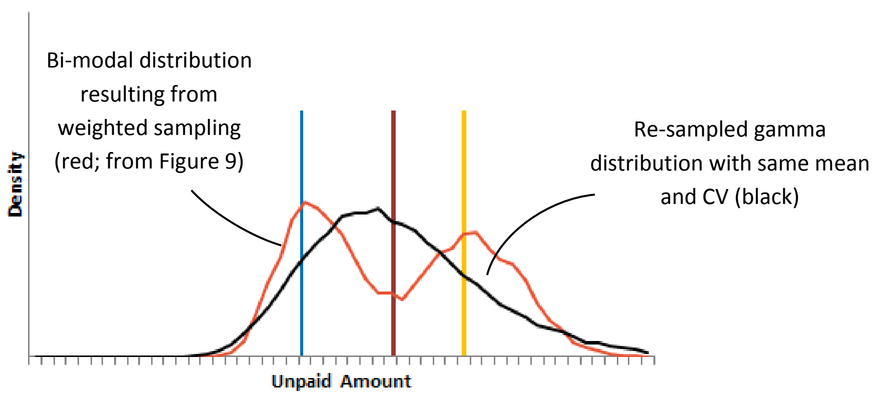


Figure 12. Re-simulated distribution – cumulative probability

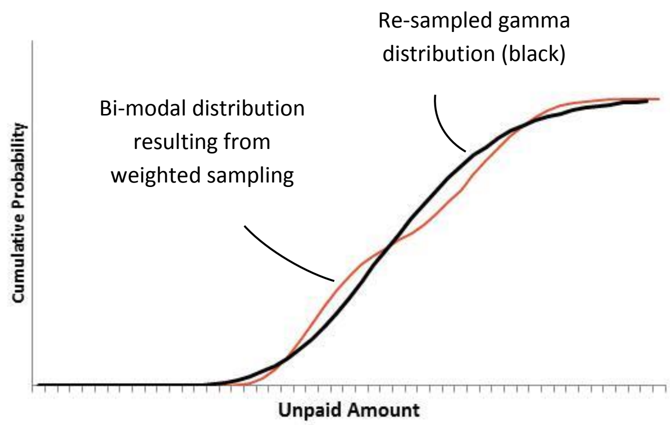


Figure 13. Re-distributed distribution – probability density

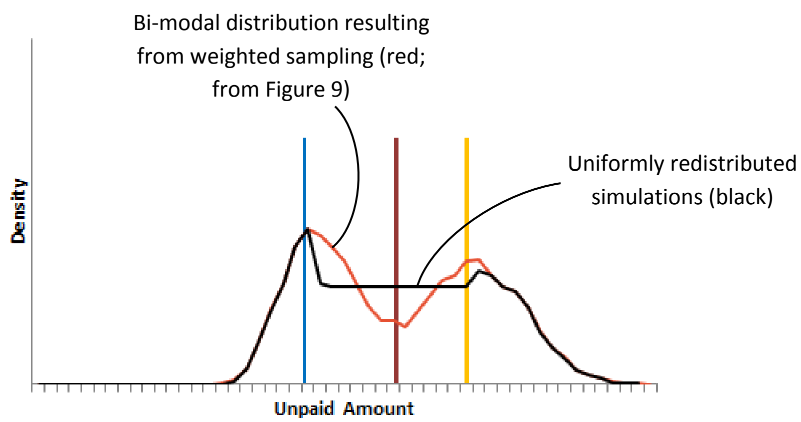
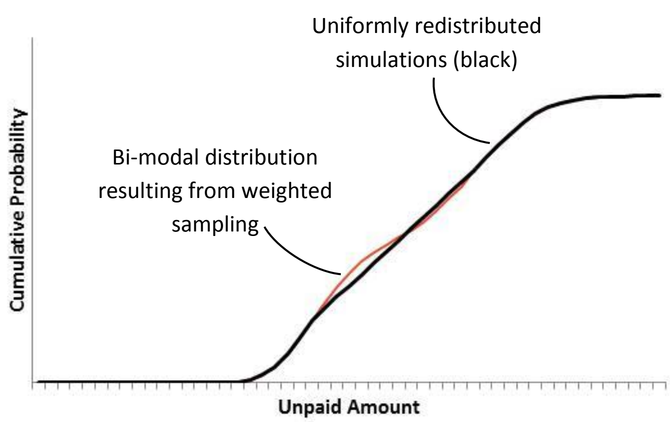


Figure 14. Re-distributed distribution – cumulative probability



5.2.4 Assigning Weights to Models

Assigning weight to a model when using the weighted sampling approach implies that the actuary believes the model is a reliable predictor because otherwise the user may be introducing additional variability that is attributable to user error. Bad practices can exist without harm to deriving a central point estimate, such as having two models that are known to be biased but offset each other so that the average produces a reasonable point estimate (e.g. "two wrongs can make a right" philosophy), but this practice should not be used when estimating uncertainty. In such cases where the models have any known bias, the user may want to consider scaling as a solution instead of weighted sampling.

5.2.5 Effect on MSE

The effect that weighted sampling has on the MSE depends on two factors:

1. The dispersion in the means of the underlying models before weighted sampling; and
2. The MSE of the model distributions before weighted sampling.

As the mean of each model converges to the same point, the resulting MSE using weighted sampling will essentially be an average of the MSE from the various models before weighted sampling. As the mean of each model diverges, the resulting MSE will increase and can be larger than the MSE before weighted sampling of each underlying model.

6 Aggregating Variability

The weighted sampling approach described thus far is an approach to incorporating model error for a ***single*** prediction. Projection methodologies used by actuaries often generate ***multiple*** predictions where each prediction corresponds to a certain subset of claims generally grouped according to a predefined time interval (e.g. accident year, report year, policy quarter, etc.), which we will refer to generically as an origin period. Weighted sampling is suitable for estimating the distribution of any single origin period prediction, however, a separate and more complex approach must be considered when aggregating the variability across multiple origin period predictions.

Consider a situation where each model used by the actuary generates a prediction, $\hat{y}_m$, for multiple different origin periods, t, such that:

$$\hat{y}_{m,t} = [\hat{y}_{m,t=1}, \hat{y}_{m,t=2}, \hat{y}_{m,t=3}, \dots]$$

and the actuary's selected central estimate for each origin period, t, is

$$\hat{y}_t = \sum_{m=A,B,\dots} w_{m,t}\hat{y}_{m,t}$$

where $w_{m,t}$ corresponds to the weight assigned to model m and origin period t and

$$\sum_{m=A,B,\dots} w_{m,t} = 1$$

Then we wish to derive an approach for aggregating the Mean Squared Error of predictions across all origin periods,

$$MSE = E\left[\left(\sum_{t=1}^{N}\left(y_t - \sum_{m=A,B,\ldots} w_{m,t}\hat{y}_{m,t}\right)\right)^2\right] = ?$$

6.1 Weighted Sampling Revisited

Expanding on the previous example in Section 5.1, consider actuarial central estimates for three separate origin periods, $\hat{y}_{t=1}, \hat{y}_{t=2}$ *and* $\hat{y}_{t=3}$, to be based on a 50%-50% weighting of predictions produced from two projection methodologies, Model A and Model B, such that:

$$For\ origin\ periods\ t = 1, 2\ and\ 3$$

$$\hat{y}_t = \sum_{m=A,B,\ldots} w_{m,t}\hat{y}_{m,t}$$

Where,

$$\hat{y}_{A,t} = the\ prediction\ from\ Model\ A\ for\ origin\ period\ t;$$

$$\hat{y}_{B,t} = the\ prediction\ from\ Model\ B\ for\ origin\ period\ t;$$

$$w_{m,t} = \begin{bmatrix} w_{A,1} & w_{A,2} & w_{A,3} \\ w_{B,1} & w_{B,2} & w_{B,3} \end{bmatrix} = \begin{bmatrix} 0.5 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0.5 \end{bmatrix}$$

Assume that distributions of the MSE for each origin period conditional on Model A and separately for Model B are already estimated and that each origin period distribution is comprised of a series of 10 simulations where each simulation, denoted x_i, is shown in Figure 15.

Figure 15. Multiple prediction model simulations

Model A Simulations

Sim	t = 1	t = 2	t = 3
1	3.4	5.8	28.8
2	2.5	12.5	28.0
3	1.8	6.5	24.0
4	3.8	8.8	20.0
5	4.4	8.7	14.5
6	3.0	10.7	14.0
7	2.0	9.4	16.9
8	6.0	7.6	24.9
9	3.7	9.7	25.0
10	6.4	8.6	29.0

Model B Simulations

Sim	t = 1	t = 2	t = 3
1	3.6	12.0	19.9
2	4.6	13.3	26.9
3	5.2	16.1	27.2
4	4.4	11.3	22.7
5	3.4	17.2	26.9
6	3.6	11.3	15.7
7	4.4	10.7	22.9
8	3.9	13.3	22.6
9	3.4	13.5	20.4
10	3.0	13.2	15.0

Once again, a distribution incorporating model error can be estimated for each origin period by taking a weighted sample without replacement of simulations from the distributions of Model A and Model B for each origin period independently in accordance with their weights. As before, this is accomplished by creating a Model Matrix, shown in Figure 16, where the weights are used as the basis for sampling between Model A and Model B for each set of origin period simulations.

Figure 16. Multiple prediction Model Matrix

Weighting Selections

	t = 1	t = 2	t = 3
Model A	50%	50%	50%
Model B	50%	50%	50%

Model Matrix

Sim	t = 1	t = 2	t = 3
1	B	B	B
2	A	B	A
3	A	B	A
4	B	B	A
5	A	A	B
6	A	A	A
7	B	B	A
8	B	A	B
9	A	B	A
10	A	A	B

Then based on the Model Matrix, we select the value corresponding to the simulation number, model and origin period to create a series of sampled simulations, which can be used as a distribution incorporating model error for each origin period's actuarial central estimate as shown in Figure 17.

Figure 17. Multiple prediction sampled simulations

Model Matrix

Sim	t = 1	t = 2	t = 3
1	B	B	B
2	A	B	A
3	A	B	A
4	B	B	A
5	A	A	B
6	A	A	A
7	B	B	A
8	B	A	B
9	A	B	A
10	A	A	B

Sampled Simulations

Sim	t = 1	t = 2	t = 3
1	3.6	12.0	19.9
2	2.5	13.3	28.0
3	1.8	16.1	24.0
4	4.4	11.3	20.0
5	4.4	8.7	26.9
6	3.0	10.7	14.0
7	4.4	10.7	16.9
8	3.9	7.6	22.6
9	3.7	13.5	25.0
10	6.4	8.6	15.0

The weighted sampling approach works for multiple separate estimates much in the same way it works for a single estimate; however, dependencies need to be considered before aggregating uncertainty across multiple origin periods. In this example, a total distribution of the three origin periods remains unanswered as depicted in Figure 18.

Figure 18. Multiple prediction weighted sampling

Sampled Simulations

Sim	t = 1	t = 2	t = 3	Total
1	3.6	12.0	19.9	?
2	2.5	13.3	28.0	?
3	1.8	16.1	24.0	?
4	4.4	11.3	20.0	?
5	4.4	8.7	26.9	?
6	3.0	10.7	14.0	?
7	4.4	10.7	16.9	?
8	3.9	7.6	22.6	?
9	3.7	13.5	25.0	?
10	6.4	8.6	15.0	?

6.2 Dependencies

If it can be assumed that within each model the predictions for each origin period are independent then an aggregate distribution representing the total of the three origin periods above can be created quite easily by summing across the values generated above for each simulation (assuming the weighted sampling used to derive the Model Matrix was generated randomly).

Unfortunately, the assumption of independence among different origin periods within a particular model is generally not true. Instead, origin period dependencies are generally inherent within the structure of a model and the process of weighted sampling among various different models for each origin period independently (as described in this example thus far) will break these origin period dependencies. Before discussing an approach to establishing a dependency, if any, among origin periods, it is useful to consider how origin period dependencies may exist within the components that make up uncertainty.

6.2.1 Origin Period Dependency – Process Error

Given that the actual outcome, y, is assumed to be a random variable, we would not expect there to be any dependency in the order in which actual outcomes occur. Therefore, it is usually assumed that the outcome of any given origin period is independent of the outcomes in any other origin period.

6.2.2 Origin Period Dependency - Parameter Error

Parameter variance measures the uncertainty in the actuary's estimate of the model's parameters used to generate a prediction. For many actuarial models, the same parameters and assumptions are used to generate predictions for all origin periods, and as such, any change to a parameter estimate or assumption will permeate through some or all of the origin periods and result in a dependency. Approaches, such as Bootstrapping, produce results which enable the user to measure this dependency.

6.2.3 Origin Period Dependency - Model Error

The model we use to predict $\hat{y}$ is likely an imperfect representation of the true model that defines the actual outcome, y, and as such may result in an unknown tendency to overestimate or underestimate the intended measure. The degree to which a model's error, if any, is dependent across different origin periods is debatable and may depend on the circumstances.

In certain circumstances, it may be argued that a model's error will be consistent across all origin periods. Consider a hypothetical example where the only difference between two chain-ladder models is the approach used to select the tail factor, which results in different values being chosen. Because the tail factor affects the predictions for all origin periods, any error may affect all origin periods.

In other circumstances, it may be argued that error, if any, in any given model may not be consistent across origin periods. For example, chain ladder models tend to be sensitive to the magnitude of cumulative amounts to which the link-ratios are applied and it may be that the cumulative amounts across origin periods exhibit an amount of reasonable volatility with respect to their size relative to historical experience simply because the volume of business being analyzed is not statistically voluminous. If the volatility observed is somewhat random across the origin periods, then the corresponding error in the model, if any, may also be random across origin periods as a result of this attribute.

Because it can be argued that model error dependency may or may not exist across origin periods, we discuss two different approaches to aggregating the weighted sampling distributions across origin periods so that a range of model error dependency assumptions can be used.

6.3 Rank Tying

One approach to aggregating the weighted sampling results across origin periods is to borrow a dependency structure from one of the underlying sampled models. Since process variance does not usually create a dependency across origin periods, any dependency observed is wholly attributable to parameter variance in standard models.

Continuing with the example discussed in Section 6.1, we can create another type of matrix, called a Rank Matrix, that identifies the "rank order" of each simulation within a given model and origin period where the largest value of all simulated values is assigned a rank value of 1. Then, the second largest value of all simulated values within that same model and origin period is assigned a rank value of 2. This process is repeated until all simulations are assigned a rank order value. The Rank Matrix for Model A and Model B are shown in Figure 19.

Figure 19 – Rank Matrix for Model A and Model B

Model A Simulations

Sim	t = 1	t = 2	t = 3
1	3.4	5.8	28.8
2	2.5	12.5	28.0
3	1.8	6.5	24.0
4	3.8	8.8	20.0
5	4.4	8.7	14.5
6	3.0	10.7	14.0
7	2.0	9.4	16.9
8	6.0	7.6	24.9
9	3.7	9.7	25.0
10	6.4	8.6	29.0

Model A Rank Matrix

Sim	t = 1	t = 2	t = 3
1	6	10	2
2	8	1	3
3	10	9	6
4	4	5	7
5	3	6	9
6	7	2	10
7	9	4	8
8	2	8	5
9	5	3	4
10	1	7	1

Model B Simulations

Sim	t = 1	t = 2	t = 3
1	3.6	12.0	19.9
2	4.6	13.3	26.9
3	5.2	16.1	27.2
4	4.4	11.3	22.7
5	3.4	17.2	26.9
6	3.6	11.3	15.7
7	4.4	10.7	22.9
8	3.9	13.3	22.6
9	3.4	13.5	20.4
10	3.0	13.2	15.0

Model B Rank Matrix

Sim	t = 1	t = 2	t = 3
1	7	7	8
2	2	4	3
3	1	2	1
4	4	9	5
5	8	1	2
6	6	8	9
7	3	10	4
8	5	5	6
9	9	3	7
10	10	6	10

Currently, the weighted sample results for each origin period in Figure 18 produces a different Rank Matrix from the Rank Matrix of Model A and Model B because the underlying Model Matrix was generated randomly in accordance with the weights and therefore broke the origin period links intrinsic to the underlying models. Figure 20 shows the implied Rank Matrix from Figure 18 which is crossed out to denote that the origin period dependencies may not be appropriate.

Figure 20. Rank Matrix from weighted sampling

Sampled Simulations

Sim	t = 1	t = 2	t = 3
1	3.6	12.0	19.9
2	2.5	13.3	28.0
3	1.8	16.1	24.0
4	4.4	11.3	20.0
5	4.4	8.7	26.9
6	3.0	10.7	14.0
7	4.4	10.7	16.9
8	3.9	7.6	22.6
9	3.7	13.5	25.0
10	6.4	8.6	15.0

Sampled Implied Rank Matrix

Sim	t = 1	t = 2	t = 3
1	7	4	7
2	9	3	1
3	10	[illegible]	4
4	[illegible]	[illegible]	6
5	2	8	2
6	8	[illegible]	10
7	[illegible]	6	8
8	[illegible]	10	5
9	6	2	3
10	1	9	9

If we select Model B as the model to use as the basis for dependency in aggregating simulations across all origin periods, then all we have to do is reorder our sampled simulation values in Figure 20 within each origin period separately so that the Rank Matrix of Model B is replicated. Then we can aggregate across each simulation as shown in Figure 21 (differences in the total occur because of rounding).

Figure 21. Reordered simulations using Model B Rank Matrix

Sampled Simulations

Sim	t = 1	t = 2	t = 3
1	3.6	12.0	19.9
2	2.5	13.3	28.0
3	1.8	16.1	24.0
4	4.4	11.3	20.0
5	4.4	8.7	26.9
6	3.0	10.7	14.0
7	4.4	10.7	16.9
8	3.9	7.6	22.6
9	3.7	13.5	25.0
10	6.4	8.6	15.0

Model B Simulation Ranks

Sim	t = 1	t = 2	t = 3
1	7	7	8
2	2	4	3
3	1	2	1
4	4	9	5
5	8	1	2
6	6	8	9
7	3	10	4
8	5	5	6
9	9	3	7
10	10	6	10

Sampled Simulations: Re-ordered by Model B

Sim	t = 1	t = 2	t = 3	Total
1	3.6	10.7	16.9	31.2
2	4.4	12.0	25.0	41.4
3	6.4	13.5	28.0	47.8
4	4.4	8.6	22.6	35.6
5	3.0	16.1	26.9	46.1
6	3.7	8.7	15.0	27.5
7	4.4	7.6	24.0	35.9
8	3.9	11.3	20.0	35.2
9	2.5	13.3	19.9	35.7
10	1.8	10.7	14.0	26.5

Note that the resulting reordered simulations are not color-coded because the link to the Model Matrix no longer exists.

The Rank Tying approach is a means to combine the simulations across origin periods while maintaining the same parameter variance dependency structure associated with one of the underlying projection models. In essence, this approach assumes that the introduction of model uncertainty does not produce any additional dependency across origin periods.

6.4 Model Tying

The Model Tying approach attempts to incorporate dependencies associated with model error into the aggregate estimate. In order to accomplish this, we will need to revisit the case study in Section 6.1 and revert to the step where the Model Matrix was created in Figure 16. The Model Matrix in Figure 16 and underlying model simulations in Figure 15 are summarized in Figure 22.

Figure 22. Multiple prediction Model Matrix

Model A Simulations

Sim	t = 1	t = 2	t = 3
1	3.4	5.8	28.8
2	2.5	12.5	28.0
3	1.8	6.5	24.0
4	3.8	8.8	20.0
5	4.4	8.7	14.5
6	3.0	10.7	14.0
7	2.0	9.4	16.9
8	6.0	7.6	24.9
9	3.7	9.7	25.0
10	6.4	8.6	29.0

Model B Simulations

Sim	t = 1	t = 2	t = 3
1	3.6	12.0	19.9
2	4.6	13.3	26.9
3	5.2	16.1	27.2
4	4.4	11.3	22.7
5	3.4	17.2	26.9
6	3.6	11.3	15.7
7	4.4	10.7	22.9
8	3.9	13.3	22.6
9	3.4	13.5	20.4
10	3.0	13.2	15.0

Weighting Selections

	t = 1	t = 2	t = 3
Model A	50%	50%	50%
Model B	50%	50%	50%

Model Matrix

Sim	t = 1	t = 2	t = 3
1	B	B	B
2	A	B	A
3	A	B	A
4	B	B	A
5	A	A	B
6	A	A	A
7	B	B	A
8	B	A	B
9	A	B	A
10	A	A	B

Under the Model Tying approach, we will rearrange the Model Matrix with the goal of maximizing the degree to which the same model is selected across as many origin periods as possible within a given simulation. In this specific example, we want to maximize the degree to which 'A's in one origin period are grouped with 'A's in other origin periods, and the degree to which 'B's are grouped with 'B's. The resulting reordered Model Matrix might look like the example in Figure 23.

Figure 23. Reordered Model Matrix

Model Matrix

Sim	t = 1	t = 2	t = 3
1	B	B	B
2	A	B	A
3	A	B	A
4	B	B	A
5	A	A	B
6	A	A	A
7	B	B	A
8	B	A	B
9	A	B	A
10	A	A	B

Model Matrix: Reordered

Sim	t = 1	t = 2	t = 3
1	A	B	A
2	A	A	A
3	B	B	B
4	A	A	A
5	A	A	A
6	B	B	B
7	A	A	A
8	B	B	B
9	A	B	A
10	B	B	B

Note that sampling error in this example means that we do not achieve an exact 50/50 split reflecting the weights chosen in each year between Model A and Model B so 'perfect strings' are not possible for all simulations.

With the reordered Model Matrix, we are now ready to select the value corresponding to the simulation number, model and origin period to derive our values for each origin period as shown in Figure 24. Also, the total can be derived by aggregating across each simulation (differences in the total occur because of rounding). It should be noted that the resulting distributions for each origin period from this approach should produce similar results to the distributions derived from weighted sampling because the reordered Model Matrix maintains the exact same weighting between the models.

Figure 24. Model Tying simulations

Model Matrix: Reordered

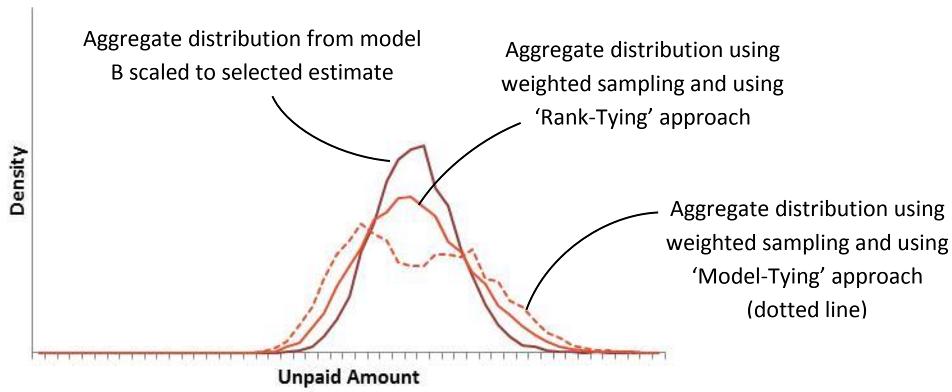
Sim	t = 1	t = 2	t = 3
1	A	B	A
2	A	A	A
3	B	B	B
4	A	A	A
5	A	A	A
6	B	B	B
7	A	A	A
8	B	B	B
9	A	B	A
10	B	B	B

Sampled Simulations: Using reordered Model Matrix

Sim	t = 1	t = 2	t = 3	Total
1	3.4	12.0	28.8	44.2
2	2.5	12.5	28.0	43.0
3	5.2	16.1	27.2	48.6
4	3.8	8.8	20.0	32.7
5	4.4	8.7	14.5	27.6
6	3.6	11.3	15.7	30.6
7	2.0	9.4	16.9	28.2
8	3.9	13.3	22.6	39.7
9	3.7	13.5	25.0	42.2
10	3.0	13.2	15.0	31.3

Figure 25 shows the resulting aggregate distribution for all three origin periods combined resulting from Model Tying, Rank Tying to Model B's dependency structure and scaling the distribution (multiplicatively) around Model B to the selected central estimate when the number of simulations in this example is increased to 10,000. All three approaches have the same mean value, which is equal to the actuarial selected central estimate for all three origin periods combined.

Figure 25. Aggregating multiple predictions: Model Tying versus Rank Tying to Model B



The difference between Model Tying and Rank Tying occurs only in the aggregate results. Rank Tying uses the parameter variance dependency attributable to only one of the models whereas Model Tying incorporates parameter variance dependencies from all models in accordance with their weights. Rank Tying excludes origin period dependencies associated with model error whereas Model Tying incorporates origin period dependency associated with model error.

6.5 Aggregation Considerations

A few points about using the Rank Tying or Model Tying approaches are noteworthy.

6.5.1 Broken Strings

With respect to the Model Tying approach, a broken string refers to a Model Matrix simulation where the same model is not identified for all origin periods. Examples of broken strings and perfect strings are shown in Figure 26.

Figure 26. Broken strings versus perfect strings

Model Matrix: Reordered

Sim	t = 1	t = 2	t = 3	
1	A	B	A	← 'Broken' string
2	A	A	A	
3	B	B	B	
4	A	A	A	
5	A	A	A	
6	B	B	B	
7	A	A	A	
8	B	B	B	
9	A	B	A	← 'Broken' string
10	B	B	B	

Broken strings can occur because of sample error as demonstrated in the previous example or because of the particular weighting attributed to the various models by origin period. A broken string is noteworthy for two reasons. First, a broken string raises the question of how to address parameter variance dependency since values are being pulled from different models within that particular simulation. One solution is to pre-sort the simulations within each model in ascending order by some measure, such as the total unpaid claim estimate across all origin periods, before applying the Model Matrix. The result will be an approximate Rank Tying of parameter variance dependency between models.

Second, a broken string implies that a dependency associated with model error does not run throughout all origin periods in that particular simulation. This should be considered a desirable effect if the broken string was caused by the particular weighting chosen for each model and origin period.

6.5.2 Increasing Complexity

The example used for Rank Tying and Model Tying was simplistic in that it used only two models, three origin periods and equal weights across all origin periods. The Rank Tying and Model Tying approaches are scalable to multiple models, an increased number of origin periods and varying weights across origin periods, however, some considerations are worth noting.

As mentioned previously, Rank Tying superimposes the parameter variance dependency structure from a single model. As the number of models is increased the relevance of any single parameter variance dependency structure is diminished accordingly. If Rank Tying is used, preference for the selected parameter variance dependency structure should be given to one of the models that contribute to the largest proportion of the total unpaid claim estimate.

Increasing the number of models and origin periods and varying the weights with Model Tying may result in broken strings and a situation where there are multiple solutions for the Model Matrix. Weightings among models should be sensible such that broken strings produce a desirable effect on the resulting distribution. An example of a desirable effect is if the actuary believes that a particular model is appropriate and hence given weight in the actuarial central estimate for only a subset of origin periods. As a result, a perfect string will not exist across all origin periods if the weight for some origin periods is zero.

With regards to multiple solutions for the Model Matrix, consider the following example in Figure 27 where we have three models used to estimate three origin periods:

Figure 27. Multiple prediction model simulations

Model A Simulations

Sim	t = 1	t = 2	t = 3
1	3.4	5.8	28.8
2	2.5	12.5	28.0
3	1.8	6.5	24.0
4	3.8	8.8	20.0
5	4.4	8.7	14.5
6	3.0	10.7	14.0
7	2.0	9.4	16.9
8	6.0	7.6	24.9
9	3.7	9.7	25.0
10	6.4	8.6	29.0

Model B Simulations

Sim	t = 1	t = 2	t = 3
1	3.6	12.0	19.9
2	4.6	13.3	26.9
3	5.2	16.1	27.2
4	4.4	11.3	22.7
5	3.4	17.2	26.9
6	3.6	11.3	15.7
7	4.4	10.7	22.9
8	3.9	13.3	22.6
9	3.4	13.5	20.4
10	3.0	13.2	15.0

Model C Simulations

Sim	t = 1	t = 2	t = 3
1	3.6	12.5	19.4
2	4.6	14.1	26.2
3	5.3	17.3	26.5
4	4.4	11.7	22.1
5	3.4	18.5	26.2
6	3.6	11.7	15.4
7	4.5	11.1	22.3
8	3.9	14.0	22.0
9	3.4	14.2	19.9
10	3.0	13.9	14.7

We can, again, create a Model Matrix, shown in figure 28, based on the selected weights from each of the Models A, B and C across 10 simulations:

Figure 28. Multiple predictions Model Matrix

Weighting Selections

	t = 1	t = 2	t = 3
Model A	33%	33%	33%
Model B	33%	33%	33%
Model C	33%	33%	33%

Model Matrix

Sim	t = 1	t = 2	t = 3
1	B	C	B
2	C	B	A
3	A	A	A
4	C	C	C
5	A	A	A
6	B	C	B
7	A	A	B
8	B	B	C
9	A	C	B
10	C	B	C

Under the Model Tying approach, we rearrange the Model Matrix with the goal of maximizing the degree to which the same model is selected across as many origin periods as possible within a given simulation. Two unique solutions exist and are shown in Figure 29:

Figure 29. Multiple solutions

Model Matrix

Sim	t = 1	t = 2	t = 3
1	B	C	B
2	C	B	A
3	A	A	A
4	C	C	C
5	A	A	A
6	B	C	B
7	A	A	B
8	B	B	C
9	A	C	B
10	C	B	C

Model Matrix: Re-ordered (solution 1)

Sim	t = 1	t = 2	t = 3
1	C	C	C
2	C	C	C
3	A	A	A
4	B	C	B
5	C	C	C
6	A	B	B
7	B	B	B
8	A	A	A
9	A	A	A
10	B	B	B

Model Matrix: Re-ordered (solution 2)

Sim	t = 1	t = 2	t = 3
1	A	C	C
2	B	B	B
3	A	A	A
4	C	C	C
5	A	A	A
6	C	C	B
7	C	C	C
8	A	A	A
9	B	B	B
10	B	B	B

Removing common strings in Figure 30 helps identify the differences:

Figure 30. Isolated differences

Model Matrix

Sim	2009	2010	2011
1	B	C	B
2	C	B	A
3	A	A	A
4	C	C	C
5	A	A	A
6	B	C	B
7	A	A	B
8	B	B	C
9	A	C	B
10	C	B	C

Model Matrix: Re-ordered (solution 1)

Sim	t = 1	t = 2	t = 3
1			
2			
3			
4	B	C	B
5	C	C	C
6	A	B	B
7			
8			
9			
10			

Model Matrix: Re-ordered (solution 2)

Sim	t = 1	t = 2	t = 3
1	A	C	C
2	B	B	B
3			
4			
5			
6	C	C	B
7			
8			
9			
10			

Although both solutions maximize origin period dependency as measured on the Model Matrix, the origin period dependency measured on the sampled simulations (i.e. values) between both solutions may differ and the preferred solution may depend on the circumstances.

6.5.3 Effects on MSE

It is difficult to make blanket statements about the impact between Rank Tying and Model Tying approaches on the overall variance of aggregate origin period predictions because it will depend on each unique situation. With regards to model error, the dependency assumed in Model Tying will generally increase the aggregate variance as compared to Rank Tying in situations where the predictions of the underlying models diverge in the same direction relative to the actuarial central estimate across origin periods. However, model error dependency assumed in Model Tying can reduce the aggregate variance in situations where the predictions of the underlying models fluctuate between being greater and less than the actuarial central estimate across origin periods.

With regards to parameter variance, the dependency assumed in Rank Tying is unaffected by the complexity in the number of models, origin periods and weights, and the dependency structure selected may be different from the dependency structures observed in other models. On the other hand, parameter variance dependency structures across models will be averaged under Model Tying and their effect may be diminished as the complexity of the approach increases.

7 Summary

It has been shown that the uncertainty in a prediction, as defined by the mean squared error, is comprised of the sum of three components: process variance, parameter variance and squared bias. Suitable approaches exist in the literature to measure these components and its corresponding distribution when a single model is considered in isolation. When multiple models are considered reasonable indicators of unpaid claims, it may be appropriate to incorporate model uncertainty into the actuary's distribution of uncertainty. Various approaches for incorporating model uncertainty were introduced. The first approach, called weighted sampling, is an approach that can be used to incorporate model uncertainty into a single prediction. Rank Tying and Model Tying are approaches that can be used to incorporate model uncertainty into an aggregation of multiple predictions that exhibit dependencies in either parameter or model uncertainty. These approaches are somewhat more complex to apply but are nevertheless important to consider when measuring the aggregate uncertainty of multiple predictions.

References

Actuarial Standards Board, *Actuarial Standard of Practice No. 43 Property/Casualty Unpaid Claim Estimates*, June 2007, Updated for Deviation Language Effective May, 1, 2011, Doc. No. 159, http://www.actuarialstandardsboard.org/pdf/asops/asop043_159.pdf

England, P. D., and Verrall, R. J., *Analytic and bootstrap estimates of prediction errors in claims reserving*, Insurance: Mathematics and Economics, 1999, 25, pp. 281-293

England, P. D., *Addendum to Analytic and bootstrap estimates of prediction errors in claims reserving*, 2001, Actuarial Research Paper No. 138a, Department of Actuarial Science and Statistics, City University, London, EC1V 0HB

England, P. D., and Verrall, R. J., *Stochastic Claims Reserving in General Insurance*, British Actuarial Journal, 8, III, 2002, pp. 443-544, http://www.actuaries.org.uk/system/files/documents/pdf/sm0201.pdf

England, P.D., and Verrall, R. J., *Predictive Distributions of Outstanding Liabilities in General Insurance*, Annals of Actuarial Science, 2006, Vol. 1, No. 2, pp. 221-270, http://cassknowledge.co.uk/sites/default/files/article-attachments/371~~richardverrall_-_predictive_distributions_of_general_insurance_outstanding_liabilities.pdf

Mack, T, *Measuring the Variability of Chain Ladder Reserve Estimates*, Casualty Actuarial Society Forum, Spring 1994, pp. 102-182, http://www.casact.org/pubs/forum/94spforum/94spf101.pdf

Appendix A

Excerpts of the following case study are used throughout this paper. In this appendix we will discuss the complete case study and will highlight relevant sections corresponding to the Figures displayed in the body of the paper.

Overview of data and selections

- This case study is based on data spanning a nine year history of origin periods, where an origin period represents an accident year.
- Development factor models (i.e. chain ladder models) were applied to each of the paid ('model A') and incurred ('model B') data in order to project to ultimate.
- A 'central estimate' was selected based on a simple average of the two development factor models for each accident year.
- Distributions reflecting process and parameter variance for each model were achieved using stochastic methods. The type of stochastic methods used is irrelevant for this illustration, but in this instance a 'practical stochastic' method was applied to Model A and a Bootstrapping approach to Model B. 'Practical stochastic' in this instance is used to describe a process whereby the analyst generates samples from a selected distribution with a user-defined mean and coefficient of variation.

For the purpose of this case study we are going to concentrate on results for just the three most recent accident years, however, any totals shown will represent the cumulative results of the full nine years of accident period history (rounding may occur with totals).

Central Estimate

The table in Figure A.1 summarizes the point estimates produced by each model for 'prior' years (1997 – 2008), 2009, 2010 and 2011 accident years, alongside the weighting used to determine the selected central estimate and the resulting amount of that estimate.

Figure A.1. Selected central estimates

	Model A	Weight	Model B	Weight	Selected Central Estimate
Prior	$2,784	50%	$8,783	50%	$5,783
2009	2,774	50%	3,838	50%	3,306
2010	8,275	50%	12,871	50%	10,573
2011	19,114	50%	23,534	50%	21,324
Total	$32,947		$49,026		$40,987

Implicit in the equal weightings used in this case study is the assumption that each model is an equally reliable predictor of the final outcome. The challenge is to estimate the corresponding uncertainty around this prediction that adequately reflects this inherent assumption.

Distributions conditional on each model

We begin the process of estimating uncertainty by developing distributions around each of the underlying models that reflect both process and parameter variance. The table in Figure A.2 summarizes the results of the stochastic uncertainty analyses performed around each of the underlying models in terms of the prediction error ("Pred. Error", $000s) of the resulting distributions as well as the coefficient of variation ("CV", prediction error as a percentage of the mean), for the most recent three accident years and in total.

Figure A.2. Summary of uncertainty conditional on each model

Model A

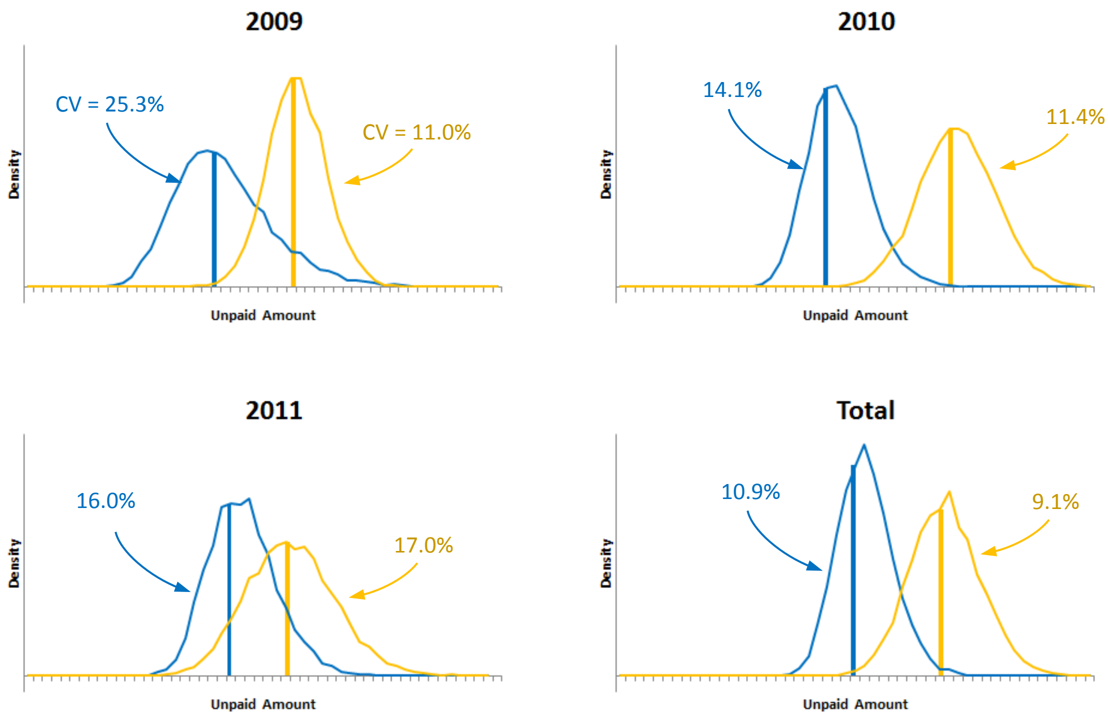
	Mean	Pred. Error	CV
2009	$2,774	$702	25.3%
2010	8,275	1,167	14.1%
2011	19,114	3,058	16.0%
Total	$32,947	$3,595	10.9%

Model B

	Mean	Pred. Error	CV
2009	$3,838	$423	11.0%
2010	12,871	1,465	11.4%
2011	23,534	3,995	17.0%
Total	$49,026	$4,441	9.1%

These distributions are also shown graphically in Figure A.3 along with the means (represented by the vertical bar) and corresponding CV's from each model (blue line is Model A, yellow line is Model B).

Figure A.3. Distributions around Model A and Model B



It should be noted that the distributions for each origin period and in total are not generated independently but rather collectively as a single process defined by the stochastic methods. As a result, origin period dependencies exist and can be measured. As a precursor for what is to come, each origin period can be treated as a 'single period prediction' as discussed in the paper through weighted sampling, however, the intrinsic origin period dependencies created by these stochastic methods will be broken. Rank Tying and Model Tying are options to restoring some sort of origin period dependency in order to recreate a 'total' aggregate distribution.

Distribution around selected central estimate using scaling

Once we have generated our distributions reflecting process and parameter uncertainty for each of the underlying models, we are faced with the challenge of producing a distribution around our selected central estimate.

One commonly-used approach is to select an underlying model and scale the associated simulated output from that model in an appropriate manner (see Section 2, Scaling).

In this example, we might select underlying Model B as our preferred model and choose multiplicative scaling to generate a distribution of simulated outcomes with a mean equal to our selected central estimate.

Figure A.4 summarizes the statistical properties of our distribution around our selected central estimate derived by multiplicatively scaling the simulations from Model B. Again, we show the prediction error ("Pred. Error", $000s) of the resulting distribution as well as the coefficient of variation ("CV", prediction error as a percentage of the mean), for the most recent three accident years and in total.

Figure A.4. Summary of uncertainty for selected central estimate using scaling

Uncertainty Summary: Selected - Scaling ($000s)

	Mean	Pred. Error	CV
2009	$3,306	$364	11.0%
2010	10,573	1,203	11.4%
2011	21,324	3,620	17.0%
Total	$40,987	$3,958	9.7%

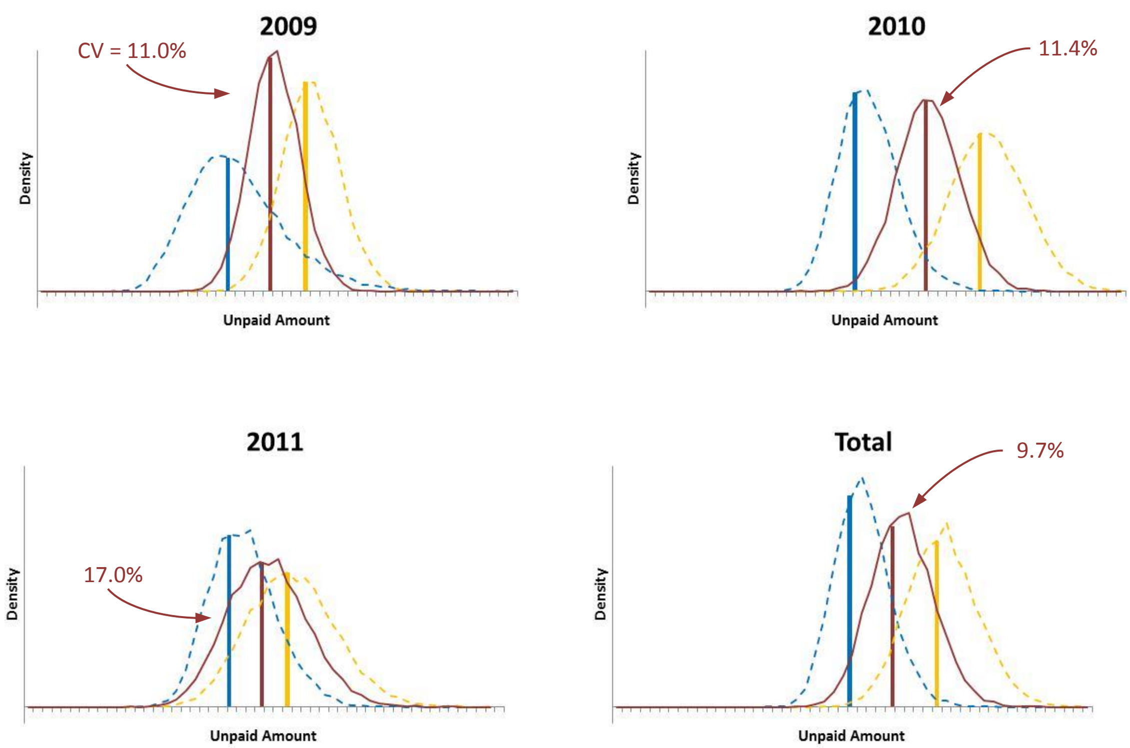
Note that, because we selected to use multiplicative scaling, the mean of the distribution is equal to our selected central point estimate and the coefficients of variation for each accident year are equivalent to the corresponding measure from the distribution developed around Model B.

Had we selected to scale additively, the mean of our distributions would still align with our selected central estimate but the coefficient of variation for each accident year would change when compared to Model B. Under additive scaling, the prediction error for each accident year remains equivalent instead of the CV.

Note also that the 'Total' coefficient of variation from multiplicative scaling is not equivalent to the 'Total' coefficient of variation from Model B. This is due to differences in the magnitude of scaling for each year.

Figure A.5 shows these scaled distributions for each accident year and in total. The selected mean and the scaled distributions are shown as solid green lines, and the distributions and means from our underlying models are shown as blue (Model A) and yellow (Model b) broken lines.

Figure A.5. Distributions using scaling

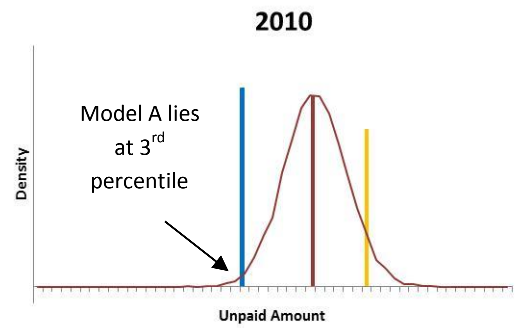


It should be noted that the graph shown in Figure A.5 for 2009 is similar to the graph shown in Figure 3 in the main text.

With regards to scaling, we are simply *borrowing* a distribution from one of our underlying models, which the actuary is forced to select. This may not adequately reflect the assumption that both models are considered to be equally valid as implied by the equal weighting used in the selection of the central estimate.

Furthermore, we may end up in a situation where our selected scaled distribution around our central estimate implies that the prediction from one of our underlying models is a relatively unlikely outcome. If we consider the 2010 accident year, our scaling approach suggests that the point estimate projected by Model A, as shown as the blue bar in Figure A.6, lies at the 3rd percentile of our range of probable outcomes.

Figure A.6. Distribution using scaling for 2010



Distribution around selected estimate using weighted sampling

We can instead employ weighted sampling for each accident year in a manner that reflects the weights selected for the determination of our selected central estimate that perhaps better represents the full distribution of possible outcomes suggested by the underlying models (see Section 5.1, Weighted Sampling).

For each accident year, we sample randomly and without replacement from each of the underlying distributions – in this case, we select 50% of the sample from the distribution around Model A and 50% from the distribution around Model B.

The table in Figure A.7 summarizes the statistical properties of our distribution around our selected central estimate derived by weighted sampling from each of the underlying models. Again, we show the prediction error ("Pred. Error", $000s) of the resulting distribution as well as the coefficient of variation ("CV", prediction error as a percentage of the mean), for the last three accident years.

Figure A.7. Summary of uncertainty using weighted sampling

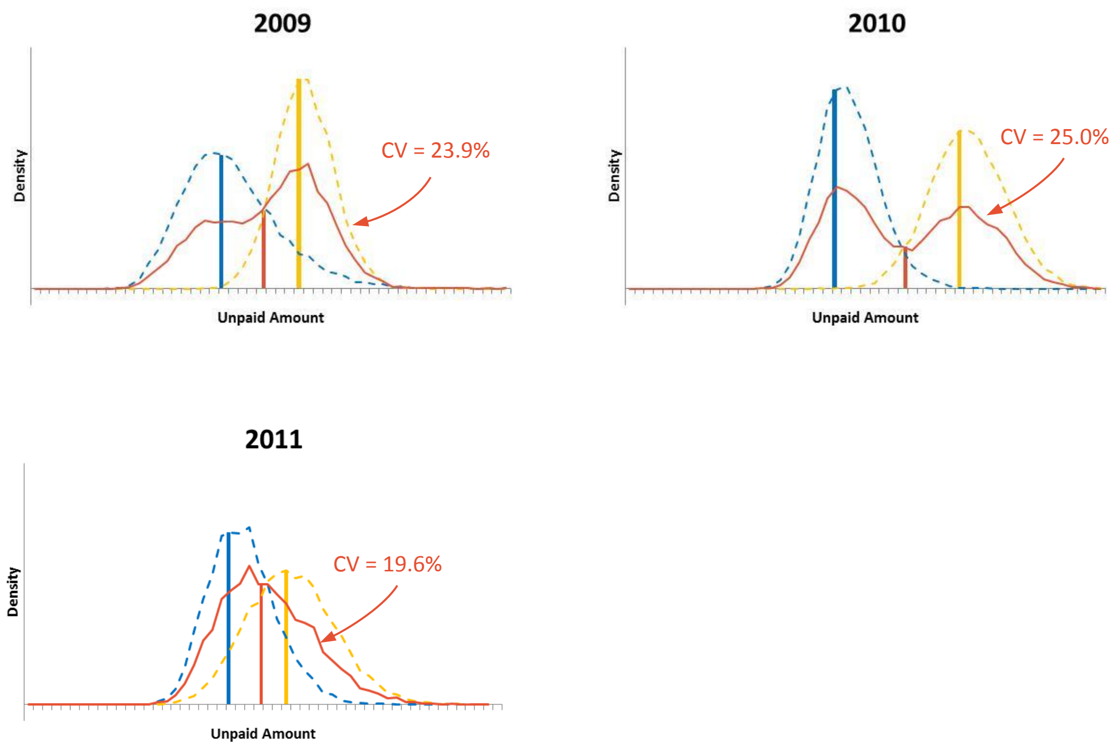
Uncertainty Summary: Selected - Wtd Sampling ($000s)

	Mean	Pred. Error	CV
2009	$3,306	$790	23.9%
2010	10,573	2,646	25.0%
2011	21,324	4,174	19.6%
Total	$40,987	?	?

As noted previously, the origin period dependencies intrinsic in the stochastic methods have been broken as a result of weighted sampling so the total aggregate distribution is no longer discernible.

The graphs in Figure A.8 show these distributions for each of the last three accident years. The selected mean and the weighted sampling distributions are shown as solid red lines, and the distributions and the means from our underlying models are again shown as broken lines.

Figure A.8. Distributions using weighted sampling



It should be noted that the graphs shown in Figure A.8 for 2009 and 2010 are similar to the graphs shown in the main text as Figures 7 and 9, respectively.

Weighted sampling will produce distributions for each accident year in isolation (as discussed for single period predictions in Section 5.1). In order to create a distribution around the selected total central estimate of unpaid claims across multiple accident years we must decide how to reintroduce an origin period dependency.

As suggested by this paper, we have the options of using either:

- Rank Tying, which reorders the year-by-year simulations such that a pre-defined accident-year correlation is targeted (as discussed in Section 6.3); or
- Model Tying, which uses a Model Matrix designed in such a manner to maximize the degree to which the same model is selected across as many different accident years as possible within a given simulation (as discussed in Section 6.4)

If using Rank Tying, the analyst should produce the Rank Matrix that is to be used to reorder the simulation. In this example, we have selected to use the Rank Matrix from the simulated distribution around Model B.

The tables in Figure A.9 summarize the point estimates and statistical properties of our distribution around each of:

- Model A;
- Model B;
- Selected central estimate using multiplicative scaled simulations from Model B;
- Selected central estimate using weighted sampling and Rank Tying accident years according to the correlation matrix suggested by Model B; and
- Selected central estimate using weighted sampling and optimized Model Tying.

Figure A.9. Summary comparing uncertainty from various models

Point Estimate Selection Summary ($000s)

	Model A	Model B	Selected: Scaled	Selected: Wtd Sample (Rank Tying)	Selected: Wtd Sample (Model Tying)
2009	$2,774	$3,838	$3,306	$3,306	$3,306
2010	8,275	12,871	10,573	10,573	10,573
2011	19,114	23,534	21,324	21,324	21,324
Total	$32,947	$49,026	$40,987	$40,987	$40,987

Uncertainty Summary: Comparison (Prediction Error)

	Model A	Model B	Selected: Scaled	Selected: Wtd Sample (Rank Tying)	Selected: Wtd Sample (Model Tying)
2009	$702	$423	$364	$790	$785
2010	1,167	1,465	1,203	2,646	2,664
2011	3,058	3,995	3,620	4,174	4,187
Total	$3,595	$4,441	$3,958	$5,854	$8,973

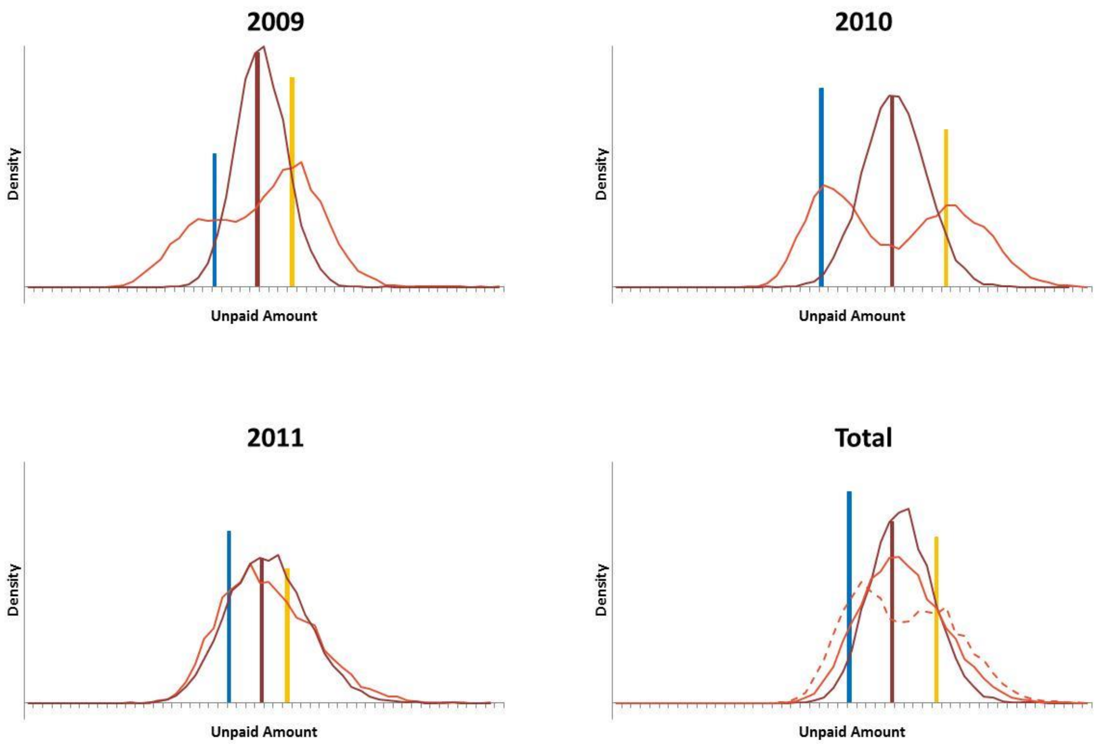
Uncertainty Summary: Comparison (CVs)

	Model A	Model B	Selected: Scaled	Selected: Wtd Sample (Rank Tying)	Selected: Wtd Sample (Model Tying)
2009	25.3%	11.0%	11.0%	23.9%	23.9%
2010	14.1%	11.4%	11.4%	25.0%	25.0%
2011	16.0%	17.0%	17.0%	19.6%	19.6%
Total	10.9%	9.1%	9.7%	14.3%	21.9%

As before, graphs assist in the interpretation and comparison of these results and the associated distributions. Such graphs corresponding to Figure A.9 can be viewed in Figure A.10. Please note:

- The blue and yellow columns represents the point estimate prediction from Models A and B
- The red column represents the selected central estimate
- The burgundy line represents the distribution around the selected central estimate using multiplicative scaling
- The red line represents the distribution around the selected central estimate using weighted sampling
- In the 'Total' graph, the distribution is shown around the total aggregate point estimate using:
    - Rank Tying (solid red line)
    - Model Tying (broken red line)
    - Scaling (solid burgundy line)

Figure A.10. Comparison of distributions using weighted sampling and scaling



Appendix B

In statistics, the Mean Squared Error (MSE) measures the difference between an estimate and what the true value is. Consider a random variable, y and a predicted variable, $\hat{y}$. The mean squared error (MSE) is:

$$E[(y-\hat{y})^2]$$

Expanding this term through additive properties yields:

$$E[(y-\hat{y})^2] = E[(y-\hat{y} + (E[y]-E[y]) + (E[\hat{y}]-E[\hat{y}]))^2]$$

Reordering yields

$$= E\left[\left((y-E[y])-(\hat{y}-E[\hat{y}])+E[y]-E[\hat{y}]\right)^2\right]$$

A series of expanding terms and subsequent simplification yields,

$$\begin{aligned}= E\big[&(y-E[y])^2-(y-E[y])(\hat{y}-E[\hat{y}])+E[y](y-E[y])-E[\hat{y}](y-E[y])+(\hat{y}-E[\hat{y}])^2\\ &-(y-E[y])(\hat{y}-E[\hat{y}])-E[y](\hat{y}-E[\hat{y}])+E[\hat{y}](\hat{y}-E[\hat{y}])+E[y]^2\\ &+E[y](y-E[y])-E[y](\hat{y}-E[\hat{y}])-E[y]E[\hat{y}]+E[\hat{y}]^2-E[\hat{y}](y-E[y])\\ &+E[\hat{y}](\hat{y}-E[\hat{y}])-E[y]E[\hat{y}]\big]\end{aligned}$$

$$\begin{aligned}= E[&(y-E[y])^2-2(y-E[y])(\hat{y}-E[\hat{y}])+2E[y](y-E[y])-2E[\hat{y}](y-E[y])+(\hat{y}-E[\hat{y}])^2\\ &-2E[y](\hat{y}-E[\hat{y}])+2E[\hat{y}](\hat{y}-E[\hat{y}])+E[y]^2-2E[y]E[\hat{y}]+E[\hat{y}]^2]\end{aligned}$$

$$\begin{aligned}= E[&(y-E[y])^2-2y\hat{y}+2yE[\hat{y}]+2\hat{y}E[y]-2E[y]E[\hat{y}]+2yE[y]-2E[y]^2-2yE[\hat{y}]\\ &+2E[\hat{y}]E[y]+(\hat{y}-E[\hat{y}])^2-2\hat{y}E[y]+2E[y]E[\hat{y}]+2\hat{y}E[\hat{y}]-2E[\hat{y}]^2+E[y]^2\\ &-2E[y]E[\hat{y}]+E[\hat{y}]^2]\end{aligned}$$

$$= E[(y-E[y])^2-2y\hat{y}+2yE[y]-E[y]^2+(\hat{y}-E[\hat{y}])^2+2\hat{y}E[\hat{y}]-E[\hat{y}]^2]$$

$$= E[(y-E[y])^2]-E[2y\hat{y}]+E\big[2yE[y]\big]-E[E[y]^2]+E[(\hat{y}-E[\hat{y}])^2]+E\big[2\hat{y}E[\hat{y}]\big]-E[E[\hat{y}]^2]$$

$$= E[(y-E[y])^2]-2E[y\hat{y}]+2E\big[yE[y]\big]-E[E[y]^2]+E[(\hat{y}-E[\hat{y}])^2]+2E\big[\hat{y}E[\hat{y}]\big]-E[E[\hat{y}]^2]$$

$$= E[(y-E[y])^2]-2E[y\hat{y}]+2E[y]E[y]-E[y]^2+E[(\hat{y}-E[\hat{y}])^2]+2E[\hat{y}]E[\hat{y}]-E[\hat{y}]^2$$

$$= E[(y-E[y])^2]-2E[y\hat{y}]+2E[y]^2-E[y]^2+E[(\hat{y}-E[\hat{y}])^2]+2E[\hat{y}]^2-E[\hat{y}]^2$$

$$= E[(y-E[y])^2]-2E[y\hat{y}]+E[y]^2+E[(\hat{y}-E[\hat{y}])^2]+E[\hat{y}]^2$$

If we assume y and $\hat{y}$ are independent, then $E[y\hat{y}] = E[y]E[\hat{y}]$ and

$$= E[(y-E[y])^2]-2E[y]E[\hat{y}]+E[y]^2+E[(\hat{y}-E[\hat{y}])^2]+E[\hat{y}]^2$$

Reordering yields,

$$= E[(y - E[y])^2] + E[(\hat{y} - E[\hat{y}])^2] + E[y]^2 - 2E[y]E[\hat{y}] + E[\hat{y}]^2$$

which simplifies to,

$$= E[(y - E[y])^2] + E[(\hat{y} - E[\hat{y}])^2] + (E[y] - E[\hat{y}])^2$$