Shapland

Outline

I. Introduction

This paper addresses the bootstrapping method for constructing a distribution of reserves. At the highest level, bootstrapping is "sampling with replacement." As you can imagine, there are a variety of ways to implement this. The method shown in this paper samples from the model residuals.

II. Notation

As always, we need to define the quantities used in the model:

- $w =$ the accident year, $1 \leq w \leq n$
- $d =$ the development age, $1 \leq d \leq n - w + 1$
- $k = w + d =$ the diagonal representing the loss information for each accident year w as of accounting year k
- $c(w, d) =$ the cumulative loss from accident year w as of age d
- $c(w, n) = U(w) =$ the total loss from accident year w when claims are at ultimate values
- $q(w, d) =$ the incremental loss for accident year w from $d - 1$ to d
- $f(d) =$ the factor applied to $c(w, d)$ to estimate $q(w, d + 1)$
- $F(d) =$ the factor applied to $c(w, d - 1)$ to estimate $c(w, d)$

III. The Bootstrap Model

The bootstrap model deployed in this paper is the **ODP bootstrap model**. This one is useful in this context because it reproduces the standard chain-ladder (CL) method.

Assuming that each accident year has the same development factor, *development factors* can be defined as follows:

$$\hat{F}(d) = \frac{\sum_{w=1}^{n-d+1} c(w,d)}{\sum_{w=1}^{n-d+1} c(w,d-1)}$$

Assuming that each accident year has a parameter representing its relative level, *ultimate claims* can be estimated as follows:

$$\hat{c}(w,n) = c(w,d) \prod_{i=d+1}^{n} \hat{F}(i)$$

where the level parameters are the current cumulative values for each accident year (i.e., $c(w, n-w+1)$).

The ODP Model

Under the ODP model, we use a GLM to model the *incremental claims* directly. The GLM uses a log link function and an ODP distribution. Under this model, the linear predictor is as follows:

$$\eta_{w,d} = c + \alpha_w + \sum_{r=2}^{d} \beta_r$$

where $\alpha_1 = \beta_1 = 0$

Note that the linear predictor shown in the source does not "sum" the β parameters. However, if you review the Excel exhibits provided by the author, the model is clearly parameterized such that the β parameters must be summed. In addition, solutions for past exam problems always sum the β parameters.

In the formulation above, the α parameters act as adjustments to the constant, c, and the β parameters adjust for the development trends after the first development period.

Alternatively, we can remove the constant, c, which causes the α parameters to act as individual AY level parameters:

$$\eta_{w,d} = \alpha_w + \sum_{r=2}^{d} \beta_r$$

where $\alpha_1 > 0$ and $\beta_1 = 0$. This is the linear predictor used for the remainder of the paper.

The **expected value** and **variance** of incremental claims are as follows:

$$E[q(w,d)] = m_{w,d}$$
$$Var\big(q(w,d)\big) = \phi E[q(w,d)]^z = \phi m_{w,d}^z$$

where $m_{w,d} = e^{\eta_{w,d}}$ and ϕ is the scale parameter (i.e., dispersion parameter).

Notice that the variance is equal to a "power" of the mean times the scale parameter. The power, z, is used to specify the error distribution:

- $z = 0$: Normal distribution
- $z = 1$: ODP distribution
- $z = 2$: Gamma distribution
- $z = 3$: Inverse Gaussian distribution

Thus, under the ODP distribution, $Var\big(q(w,d)\big) = \phi m_{w,d}$.

Estimating the α & β Parameters Using a GLM

Shapland describes the following process for estimating the α & β parameters using a GLM:

- Apply the log-link function to the incremental loss triangle
- Specify the GLM using a system of equations with α & β parameters
- Convert the system of equations into matrix notation, $Y = XA$, where Y is the matrix of log-incremental data, X is the design matrix, and A is the matrix of α & β parameters

- Use a numerical technique (ex. iteratively weighted least squares, Newton-Raphson) to solve for the model parameters in the A matrix that minimize the squared difference between the Y matrix and the solution matrix, $\hat{Y}$. The solution matrix is the column vector whose entries are the applicable $\ln(m_{w,d})$ values

The method above for estimating the α & β parameters is a bit different from the one described in Taylor & McGuire. Furthermore, Shapland's description is not exactly correct (although the Excel exhibits provided with the paper correctly solve for the parameters). For the purposes of the *exam*, we recommend treating Shapland's description as another formulation for solving for the model parameters.

Example: Setting Up the ODP GLM

A GLM with a log link function and an ODP error distribution is fit to a 4 x 4 incremental loss triangle.

The system of equations underlying the GLM are as follows:

- $\ln[q(1,1)] = \alpha_1$
- $\ln[q(2,1)] = \alpha_2$
- $\ln[q(3,1)] = \alpha_3$
- $\ln[q(4,1)] = \alpha_4$
- $\ln[q(1,2)] = \alpha_1 + \beta_2$
- $\ln[q(2,2)] = \alpha_2 + \beta_2$
- $\ln[q(3,2)] = \alpha_3 + \beta_2$
- $\ln[q(1,3)] = \alpha_1 + \beta_2 + \beta_3$
- $\ln[q(2,3)] = \alpha_2 + \beta_2 + \beta_3$
- $\ln[q(1,4)] = \alpha_1 + \beta_2 + \beta_3 + \beta_4$

We convert the system of equations into matrix notation, $Y = XA$, as follows:

$$Y = \begin{bmatrix} \ln[q(1,1)] \\ \ln[q(2,1)] \\ \ln[q(3,1)] \\ \ln[q(4,1)] \\ \ln[q(1,2)] \\ \ln[q(2,2)] \\ \ln[q(3,2)] \\ \ln[q(1,3)] \\ \ln[q(2,3)] \\ \ln[q(1,4)] \end{bmatrix} \qquad X = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix} \qquad A = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}$$

The model parameters are those that minimize the squared difference between the following two matrices:

$$Y = \begin{bmatrix} \ln[q(1,1)] \\ \ln[q(2,1)] \\ \ln[q(3,1)] \\ \ln[q(4,1)] \\ \ln[q(1,2)] \\ \ln[q(2,2)] \\ \ln[q(3,2)] \\ \ln[q(1,3)] \\ \ln[q(2,3)] \\ \ln[q(1,4)] \end{bmatrix} \qquad \hat{Y} = \begin{bmatrix} \ln[m_{1,1}] \\ \ln[m_{2,1}] \\ \ln[m_{3,1}] \\ \ln[m_{4,1}] \\ \ln[m_{1,2}] \\ \ln[m_{2,2}] \\ \ln[m_{3,2}] \\ \ln[m_{1,3}] \\ \ln[m_{2,3}] \\ \ln[m_{1,4}] \end{bmatrix}$$

The ODP Bootstrap Model

One **important property** of the over-dispersed Poisson model is that the fitted incremental claims will *exactly equal* the fitted incremental claims derived using the standard CL volume-weighted factors:

- If we start with the latest diagonal and divide backwards successively by each age-to-age factor, we obtain fitted cumulative claims. Using subtraction, the fitted cumulative claims can be used to determine the fitted incremental claims
- These fitted incremental claims exactly match those obtained using the GLM
- This simplified model is known as the *ODP bootstrap model*

The property above has **three important consequences**:

- A simple link ratio algorithm can be used in place of the more complicated GLM algorithm, while still maintaining an underlying GLM framework
- The use of the age-to-age factors serves as a bridge to the deterministic framework. This allows the model to be more easily explained
- In general, the log link function does not work for negative incremental claims. Using link ratios remedies this problem

Unscaled Pearson Residuals

The bootstrap process involves sampling with replacement from the residuals of the incremental loss data (i.e., actual – fitted). Although the deviance, Pearson, and Anscombe residuals are potential candidates, the **Pearson residuals** are used since they are calculated consistently with the scale parameter, ϕ.

The **unscaled Pearson residuals** and **scale parameter** are calculated as follows:

$$r_{w,d} = \frac{q(w,d) - m_{w,d}}{\sqrt{m^z_{w,d}}}$$

$$\phi = \frac{\sum r^2_{w,d}}{N - p}$$

where N is the number of data cells in the triangle and p is the number of parameters. In general, $p = 2 \cdot (\ of\ AYs) - 1$ since $\beta_1 = 0$.

In the formula above, the numerator is the raw residual. The denominator is the square root of the variance. The variance excludes the scale parameter since it's simply a proportionality constant.

Distribution of the Residuals

Although sampling with replacement assumes the residuals are independent and identically distributed (i.i.d.), it does *not require* the residuals to be normally distributed:

- This is an **advantage** of the ODP bootstrap model since the actual distributional form of the residuals will flow through the simulation process
- This is sometimes referred to as a "semi-parametric" bootstrap model since we are not parameterizing the residuals

Sample Triangles

Sampling with replacement from the residuals can be used to create new sample triangles of incremental claims using the following formula:

$$q^*(w,d) = r^* \sqrt{m^z_{w,d}} + m_{w,d}$$

where r^* is a randomly select residual from the sample.

The new sample triangles can then be cumulated, and age-to-age factors can be calculated and applied to calculate reserve point estimates. This produces a *distribution of point estimates*.

Process Variance in the Future Incremental Losses

The point estimates above do not fully capture the predictive distribution of future losses. Although they incorporate process variance and parameter variance in the simulation of the *historical data* (i.e., each sample triangle is a recreation of history), they fail to incorporate process variance in the simulation of the *future data*. This is because we are producing point estimates for each sample triangle, which is an expected value.

To *incorporate process variance in the simulation of future data*, we can expand the simulation process. Rather than projecting each incremental loss using age-to-age factors, we sample from a Gamma distribution with mean $m_{w,d}$ and variance $\phi m_{w,d}$. In doing so, our distribution of point estimates becomes a *distribution of possible outcomes*, which incorporates process variance and parameter variance in the simulation of the historical and future data.

Degrees of Freedom Adjustment Factor

Historical papers have taken slightly different approaches for implementing the ODP bootstrap model. For example, England & Verrall suggest multiplying the distribution of point estimates by the **degrees of freedom adjustment factor**, f^{DoF}, and the **scale parameter**. The degrees of freedom factor is defined as follows:

$$f^{DoF} = \sqrt{\frac{N}{N-p}}$$

By multiplying the point estimates by the two quantities mentioned above, we allow for over-dispersion of the residuals in the sampling process and add process variance to future incremental losses.

Scaled Pearson Residuals

Another way to add over-dispersion in the residuals is by multiplying the unscaled Pearson residuals by the degrees of freedom adjustment factor. This produces the **scaled Pearson residuals**, defined as follows:

$$r_{w,d}^{S} = r_{w,d} \times f^{DoF}$$

We can think of the scaled Pearson residuals as a "bias-corrected" version of the unscaled Pearson residuals.

Standardized Residuals

The issue with the scaled Pearson residuals is that the f^{DoF} factor does not produce standardized residuals. Assuming our model is properly specified, standardized residuals should have a constant variance. This is important since the ODP bootstrap model assumes that residuals are i.i.d.

If the model is not properly specified, standardized residuals may not display a constant variance. This is known as **heteroscedasticity**. In this case, it might indicate that something other than a Poisson distribution should be used. Alternatively, it might indicate that more predictors are needed.

In order to calculate the standardized residuals, the hat matrix adjustment factor, $f_{w,d}^{H}$, is required. Mathematically:

$$\boxed{f_{w,d}^{H} = \sqrt{\frac{1}{1 - H_{i,i}}}}$$

where $H_{i,i}$ represents the i^{th} point on the diagonal of the hat matrix. Note that the hat matrix is the projection matrix that maps the response vector Y to the solution vector $\hat{Y}$.

As a quick example of the structure of the hat matrix, suppose we are given the following hat matrix for a 3 x 3 loss triangle:

$$\begin{bmatrix} H_{1,1} & H_{1,2} & H_{1,3} & H_{1,4} & H_{1,5} & H_{1,6} \\ H_{2,1} & H_{2,2} & H_{2,3} & H_{2,4} & H_{2,5} & H_{2,6} \\ H_{3,1} & H_{3,2} & H_{3,3} & H_{3,4} & H_{3,5} & H_{3,6} \\ H_{4,1} & H_{4,2} & H_{4,3} & H_{4,4} & H_{4,5} & H_{4,6} \\ H_{5,1} & H_{5,2} & H_{5,3} & H_{5,4} & H_{5,5} & H_{5,6} \\ H_{6,1} & H_{6,2} & H_{6,3} & H_{6,4} & H_{6,5} & H_{6,6} \end{bmatrix}$$

The hat matrix adjustment factors are as follows:

$$\begin{bmatrix} f_{1,1}^{H} = \sqrt{\dfrac{1}{1-H_{1,1}}} & & & & & \\ & f_{2,1}^{H} = \dots & & & & \\ & & f_{3,1}^{H} = \dots & & & \\ & & & f_{1,2}^{H} = \dots & & \\ & & & & f_{2,2}^{H} = \dots & \\ & & & & & f_{1,3}^{H} = \dots \end{bmatrix}$$

The hat matrix adjustment factor is considered as a replacement for and an improvement over the degrees of freedom adjustment factor.

The **standardized residuals** are found by multiplying the hat matrix adjustment factors by the unscaled Pearson residuals:

$$\boxed{r_{w,d}^{H} = \frac{q(w,d) - m_{w,d}}{\sqrt{m_{w,d}^{z}}} \cdot f_{w,d}^{H}}$$

The standardized residuals are the *final residuals that should be sampled from* when running the ODP bootstrap model.

Approximating the Scale Parameter

Although the standardized residuals serve as the foundation of the bootstrapping process, the scale parameter should still be calculated *using the unscaled Pearson residuals*.

In the event that the unscaled Pearson residuals are not readily available, we can approximate the scale parameter using the standardized residuals:

$$\phi^{H} = \frac{\sum\left(r_{w,d}^{H}\right)^{2}}{N}$$

Zero-Value Residuals

When sampling from residuals, we should not include any zero-value residuals. These are normally found in the "corners" of the triangle (i.e., the $(n, 1)$ and $(1, n)$ cells). A common explanation for excluding these residuals is that these cells contain variance…we just don't know what that variance is yet!

Example: ODP Bootstrap Model

An actuary is running an ODP bootstrap model. Given the following information for an insurer as of December 31, 2023:

	Cumulative Paid Losses ($) as of XX Months:		
AY	12 mo.	24 mo.	36 mo.
2021	95	150	180
2022	115	160	
2023	105		

The hat matrix is as follows:

$$\begin{bmatrix} 0.8335 & 0.1665 & 0.0000 & 0.1665 & -0.1665 & 0.0000 \\ 0.1561 & 0.8439 & 0.0000 & -0.1561 & 0.1561 & 0.0000 \\ 0.0000 & 0.0000 & 1.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.3496 & -0.3496 & 0.0000 & 0.6504 & 0.3496 & 0.0000 \\ -0.3278 & 0.3278 & 0.0000 & 0.3278 & 0.6722 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0000 \end{bmatrix}$$

- $\phi = 2.58$

First, let's calculate the standard CL volume-weighted factors:

- 12-24: $\frac{150+160}{95+115} = 1.476$
- 24-36: 1.200

Second, let's calculate the fitted cumulative losses by starting with the latest diagonal and dividing backwards successively by each age-to-age factor:

	Fitted Cumulative Paid Losses ($) as of XX Months:		
AY	12 mo.	24 mo.	36 mo.
2021	$101.61 = \frac{150}{1.476}$	$150 = \frac{180}{1.200}$	180
2022	108.39	160	
2023	105		

Third, let's calculate the fitted incremental losses:

	Fitted Incremental Paid Losses ($) as of XX Months:		
AY	12 mo.	24 mo.	36 mo.
2021	101.61	48.39 = 150 – 101.61	30
2022	108.39	51.61	
2023	105		

Fourth, let's calculate the actual incremental losses:

	Actual Incremental Paid Losses ($) as of XX Months:		
AY	12 mo.	24 mo.	36 mo.
2021	95	55 = 150 – 95	30
2022	115	45	
2023	105		

Fifth, let's calculate the unscaled Pearson residuals:

	Unscaled Pearson Residuals as of XX Months:		
AY	12 mo.	24 mo.	36 mo.
2021	-0.66	$0.95 = \frac{55 - 48.39}{\sqrt{48.39}}$	0.00
2022	0.64	-0.92	
2023	0.00		

Sixth, let's calculate the hat matrix adjustment factors (set to 0 if undefined):

	Hat Matrix Adj. Factors as of XX Months:		
AY	12 mo.	24 mo.	36 mo.
2021	$2.451 = \sqrt{\frac{1}{1-0.8335}}$	1.691	0.000
2022	$2.531 = \sqrt{\frac{1}{1-0.8439}}$	1.750	
2023	0.000		

Seventh, let's calculate the standardized residuals:

	Standardized Residuals as of XX Months:		
AY	12 mo.	24 mo.	36 mo.
2021	$-1.61 = -0.66(2.451)$	1.61	0.00
2022	$1.61 = 0.64(2.531)$	-1.61	
2023	0.00		

Eighth, let's randomly sample from the standardized residuals to create the following sample:

	Sample Standardized Residuals as of XX Months:		
AY	12 mo.	24 mo.	36 mo.
2021	1.61	1.61	-1.61
2022	-1.61	1.61	
2023	-1.61		

Ninth, let's convert the sample standardized residuals into sample incremental losses:

	Sample Incremental Paid Losses ($) as of XX Months:		
AY	12 mo.	24 mo.	36 mo.
2021	$117.82 = 1.61\sqrt{101.61} + 101.61$	59.57	21.19
2022	91.65	63.16	
2023	88.53		

Tenth, let's use the sample incremental losses to create a triangle of sample cumulative losses:

	Sample Cumulative Paid Losses ($) as of XX Months:		
AY	12 mo.	24 mo.	36 mo.
2021	117.82	177.39	198.58
2022	91.65	154.81	
2023	88.53		
Age-to-Age		1.586	1.119

Eleventh, let's project the sample cumulative losses to ultimate using the age-to-age factors:

	Projected Sample Cumulative Paid Losses ($) as of XX Months:			Reserve Point Estimate
AY	12 mo.	24 mo.	36 mo.	
2021	117.82	177.39	198.58	**0.00**
2022	91.65	154.81	**173.31**	**18.50**
2023	88.53	**140.40**	**157.17**	**68.64**
Total				**87.14**

The reserve point estimates above *do not include process variance in the simulation of the future values*. To include this process variance, we replace the future incremental losses with simulated values from a gamma distribution with mean $m_{w,d}$ and variance $\phi m_{w,d}$.

For example, the future incremental losses including process variance might look like this:

	Sample Incremental Paid Losses w/PV ($) as of XX Months:			Reserve Possible Outcome
AY	12 mo.	24 mo.	36 mo.	
2021	117.82	59.57	21.19	**0.00**
2022	91.65	63.16	**6.71**	**6.71**
2023	88.53	**59.13**	**22.80**	**81.93**
Total				**88.64**

The point estimates are now possible outcomes. To simulate from a gamma distribution, we can use the GAMMAINV(RAND(), α, β) function in Excel. The mean and variance of a gamma distribution are $\alpha\beta$ and $\alpha\beta^2$, where α is the shape parameter and β is the scale parameter. Thus,

$\alpha = \frac{m_{w,d}}{\phi}$ and $\beta = \phi$. Note that the Verrall paper uses an alternative parameterization of the gamma distribution.

As an example:

- AY 2022 at 36 months
  - $m_{2,3} = 173.31 - 154.81 = 18.50$
  - $\phi =$ 2.58 (given in problem)
  - GAMMAINV(RAND(), $\frac{18.50}{2.58}$, 2.58) = 6.71
- AY 2023 at 24 months
  - $m_{3,2} = 140.40 - 88.53 = 51.87$
  - GAMMAINV(RAND(), $\frac{51.87}{2.58}$, 2.58) = 59.13

Keep in mind that the simulated gamma values above depend on the random seed (i.e., the random uniform variable produced by the RAND() function). If you use the function above in Excel, it's unlikely you will get the same values since it will be based on a different random seed.

At this point, we have completed one pass through the bootstrapping process. The total point estimate (or possible outcome) produced at the end represents one point (or possible outcome) in the distribution of point estimates (or possible outcomes).

Starting with the random sampling of the standardized residuals, we would redo this entire process a number of times until a full distribution is produced.

Bootstrapping the Incurred Loss Triangle

The ODP bootstrap model can be applied to both paid and incurred data. If applied to paid data, the distribution of possible outcomes represents the total unpaid. If applied to incurred data, the distribution of possible outcomes represents IBNR. Thus, they are not directly comparable.

Shapland covers two approaches for modeling an unpaid loss distribution using incurred data.

Approach 1

Under Approach 1, we run a paid data model in conjunction with the incurred data model. Then, we use the random payment pattern from each iteration of the paid data model to convert the ultimate values from each corresponding incurred model iteration to develop paid losses by accident year.

A **benefit** of this approach is that it allows us to use the case reserves to help predict the ultimate losses, while still focusing on the payment stream for measuring risk.

An **improvement** to this approach would be the inclusion of correlation between the paid and incurred models (possibly in the residual sampling process). For example, if we want to compare iterations showing long payment streams with iterations showing high incurred results, we must consider correlation.

Approach 2

Under this approach, we apply the ODP bootstrap to the Munich Chain-Ladder (MCL) model. The MCL uses the inherent relationship/correlation between paid and incurred losses to predict ultimate losses. When paid losses are low relative to incurred losses, then future paid loss development tends to be higher than average. When paid losses are high relative to incurred losses, then future paid loss development tends to be lower than average.

Two advantages of this approach versus the first approach are as follows:

- It does not require us to model paid losses twice
- It explicitly measures the correlation between paid and incurred losses

Bootstrapping the BF and Cape Cod Models

An **issue** with using the ODP bootstrap process is that iterations for the latest few accident years tend to be more variable than what we would expect given the simulations for earlier accident

years. This is due to the fact that *more* age-to-age factors are used to extrapolate the sampled values to develop point estimates for each iteration.

To **address** this issue, future incremental values can be extrapolated using the BF or Cape Cod method.

Typically, these methods are deterministic. However, they can be made stochastic by converting the deterministic assumptions to stochastic assumptions. For example:

- Instead of simply assuming a priori loss ratios for the BF method, we can add a vector of standard deviations to go with these means
- We can then assume a distribution and simulate a different a priori loss ratio for every iteration of the model

The GLM Bootstrap Model

So far, we have discussed the ODP bootstrap model, where the standard CL factors are used to produce fitted values. Two **limitations** of the ODP bootstrap model are as follows:

- It does not adjust for CY effects
- Since it includes a parameter for each AY and development period beyond the first period, it might over-fit the data

If we operate completely in the GLM framework (known as the *GLM bootstrap model*), we can address these limitations.

The GLM bootstrap model has the following **two drawbacks**:

- The GLM must be solved for each iteration of the bootstrap model, which may slow down the simulation
- The model is no longer directly explainable to others using age-to-age factors

The GLM bootstrap model has the following **benefits**:

- We can specify fewer parameters, which helps avoid over-fitting

- We can add parameters for CY trends. If we include CY trends, the linear predictor changes to the following:

$$\eta_{w,d} = \alpha_w + \sum_{r=2}^{d} \beta_r + \sum_{k=2}^{w+d-1} \gamma_k$$

  where γ_k is the CY trend, $\alpha_1 > 0$, $\beta_1 = 0$, and $\gamma_1 = 0$

- We can model shapes other than triangles
  - If incremental data is missing for the first few diagonals, the remaining values in those rows would not be useful under the CL method since it requires cumulative values
  - Since the GLM bootstrap uses incremental values, the entire trapezoid can be used to fit the model
- We can match the model parameters to the statistical features found in the data
  - For example, modeling with fewer development trend parameters means that the last parameter is assumed to continue past the end of the triangle. This gives us a tail without actually having to specify a tail factor

One **issue with the CY trend model** is that the system of equations no longer has a unique solution. For example:

- Excluding CY trends, a 3 x 3 triangle has five parameters (three α parameters and two β parameters)
- Including CY trends, a 3 x 3 triangle has seven parameters (three α parameters, two β parameters, and two γ parameters)
- Since six cells of the triangle contain data, we now have six knowns and seven unknowns
- To deal with this issue, we start with a model with one α parameter, one β parameter, and one γ parameter. We then add and remove parameters as needed

How do we calculate expected future incremental losses using the GLM bootstrap model?

- Unlike the ODP bootstrap model, we ***do not apply*** age-to-age factors to each sample triangle to produce point estimates. Instead, we fit the same GLM model underlying the residuals to each sample triangle. Then, we use the resulting parameters to produce reserve point estimates (or possible outcomes)
- One **drawback** of this approach is the additional time required to fit a GLM to each sample triangle

---

Example: GLM Bootstrap Model

An actuary is using the GLM bootstrap model with a log link and an ODP error distribution to produce a reserve distribution. Given the following information for an insurer as of December 31, 2023:

	Cumulative Paid Losses ($) as of XX Months:		
AY	12 mo.	24 mo.	36 mo.
2021	50	80	100
2022	60	90	
2023	80		

The fitted parameters are as follows:

- $\hat{\alpha}_1 = 3.947$
- $\hat{\alpha}_2 = 4.064$
- $\hat{\alpha}_3 = 4.382$
- $\hat{\beta}_2 = -0.606$
- $\hat{\beta}_3 = -0.345$

First, let's calculate the fitted incremental paid losses using $m_{w,d} = e^{\alpha_w + \sum_{r=2}^{d} \beta_r}$

	Fitted Incremental Paid Losses ($) as of XX Months:		
AY	12 mo.	24 mo.	36 mo.
2021	$51.76 = e^{3.947}$	$28.24 = e^{3.947-0.606}$	20
2022	$58.24 = e^{4.064}$	31.76	
2023	80		

Second, let's calculate the actual incremental paid losses:

	Actual Incremental Paid Losses ($) as of XX Months:		
AY	12 mo.	24 mo.	36 mo.
2021	50	30	20
2022	60	30	
2023	80		

Third, let's calculate the unscaled Pearson residuals:

	Unscaled Pearson Residuals as of XX Months:		
AY	12 mo.	24 mo.	36 mo.
2021	$-0.25 = \frac{50 - 51.76}{\sqrt{51.76}}$	0.33	0.00
2022	0.23	-0.31	
2023	0.00		

The remainder of the GLM bootstrap model process is as follows:

- Calculate standardized residuals by multiplying the unscaled Pearson residuals by the hat matrix adjustment factors
- Randomly sample from the standardized residuals to create a sample residual triangle
- Convert the sample residual triangle to a sample incremental paid loss triangle
- Fit the same GLM structure underlying the original model to the sample incremental paid loss triangle

- Calculate future expected incremental values using the fitted parameters from the prior step
- Sample from a gamma distribution with mean $m_{w,d}$ and variance $\phi m_{w,d}$ to incorporate process variance into each future cell of the sample incremental paid loss triangle
- Sum up the expected future incremental paid losses to obtain a possible outcome for each AY
- Starting with the random sampling of the standardized residuals, redo this process thousands of times

Notice that we never need to produce sample cumulative paid losses when using the GLM bootstrap model. This is because the GLM models incremental losses directly.

IV. Practical Issues

Negative Incremental Values – GLM Bootstrap Model

Negative incremental values can be problematic when parameterizing a GLM with a log link function since we cannot take the log of a negative number. Depending on the extent of negative values, we have two options for dealing with them.

Option 1: Modified Log-Link Function

If some of the incremental values in a column are negative, but the **sum of the incremental values in the column is positive**, we can use the following modified log-link function to solve the GLM:

$$\begin{cases} \ln[q(w,d)], & q(w,d) > 0 \\ 0, & q(w,d) = 0 \\ -\ln[\text{abs}(q(w,d))], & q(w,d) < 0 \end{cases}$$

The reason the GLM doesn't break down here is because each β parameter is based on the entire column. As long as the column sum is positive, the β parameter will usually be found.

Example: Modified Log-Link Function

Given the following incremental incurred losses as of December 31, 2023:

	Incremental Incurred Losses ($) as of XX Months:		
AY	12 mo.	24 mo.	36 mo.
2021	100	-20	15
2022	110	50	
2023	120		

Using the modified log-link function, the values used to solve the GLM are as follows:

	Log-Incremental Incurred Losses ($) as of XX Months:		
AY	12 mo.	24 mo.	36 mo.
2021	4.605	$-2.996 = -\ln[\text{abs}(-20)]$	2.708
2022	4.700	3.912	
2023	4.787		

Option 2: Subtracting the Largest Negative Value

If the **sum of the incremental values in a column is negative**, the iterative algorithm used to solve for the GLM parameters will diverge. Instead, we must use a different approach:

- Find the largest negative value in the triangle
- Set the constant ψ equal to the largest negative value
- Subtract ψ from every incremental value in the triangle to arrive at the following adjusted incremental values:

$$\boxed{q^+(w,d) = q(w,d) - \psi}$$

- Solve for the GLM parameters based on the logs of the adjusted incremental values
- Calculate the fitted incremental values using $m_{w,d}^{+} = e^{\text{linear predictor}}$

- Adjust the resulting fitted incremental values as follows:

$$\boxed{m_{w,d} = m_{w,d}^{+} + \psi}$$

The reason we add ψ back in add the end is to ensure the final fitted values are on the same basis as the original data.

Example: Subtracting the Largest Negative Value

Given the following incremental incurred losses as of December 31, 2023:

	Incremental Incurred Losses ($) as of XX Months:		
AY	12 mo.	24 mo.	36 mo.
2021	100	-20	15
2022	110	10	
2023	120		

In this case, the sum of the incremental losses in the second column is negative. Thus, we cannot use the simple modification we used in the previous example. Instead, we must use the second modification.

First, let's calculate the adjusted incremental losses by subtracting the largest value in the triangle:

	Adjusted Incremental Incurred Losses ($) as of XX Months:		
AY	12 mo.	24 mo.	36 mo.
2021	120	$0 = -20 - (-20)$	35
2022	130	30	
2023	140		

Second, let's take the log of the adjusted incremental losses (set to 0 if equal to 0):

	Log of Adjusted Incremental Incurred Losses ($) as of XX Months:		
AY	12 mo.	24 mo.	36 mo.
2021	4.787	0.000	3.555
2022	4.868	3.401	
2023	4.942		

Once the GLM parameters have been determined, we can calculate the fitted values, $m_{w,d}^{+}$. Then, we "adjust them back to normal" by adding the negative value we subtracted earlier (i.e., $m_{w,d} = m_{w,d}^{+} + \psi$). This reduces all of the fitted values.

In the *second option* above, some of the final fitted values will be negative since we add the largest negative value back in at the end. This is a problem when calculating residuals and sampled incremental values because we must take the square root of the fitted value.

To deal with this, we *modify* the **unscaled Pearson residuals** and **sampled incremental losses** as follows:

$$r_{w,d} = \frac{q(w,d) - m_{w,d}}{\sqrt{\text{abs}\left(m_{w,d}^{z}\right)}}$$

$$q^{*}(w,d) = r^{*}\sqrt{\text{abs}(m_{w,d}^{z})} + m_{w,d}$$

Negative Incremental Values – ODP Bootstrap Model

Unlike the GLM bootstrap model, the ODP bootstrap is not impacted by negative values when calculating fitted values. The development factors being used to "divide backwards" will simply be less than 1.00.

Since development factors less than 1.00 will produce negative incremental fitted values, we must use the modified unscaled Pearson residuals and sampled incremental losses shown above to run the full bootstrap process.

Negative Values During Simulation

Although we have dealt with the issue of negative incremental values when parameterizing a model, negative values still affect the process variance in the simulation process.

When each future incremental value is sampled from a gamma distribution to add process variance, the gamma parameters must be positive. To simulate negative values from a gamma distribution, we have **two options**:

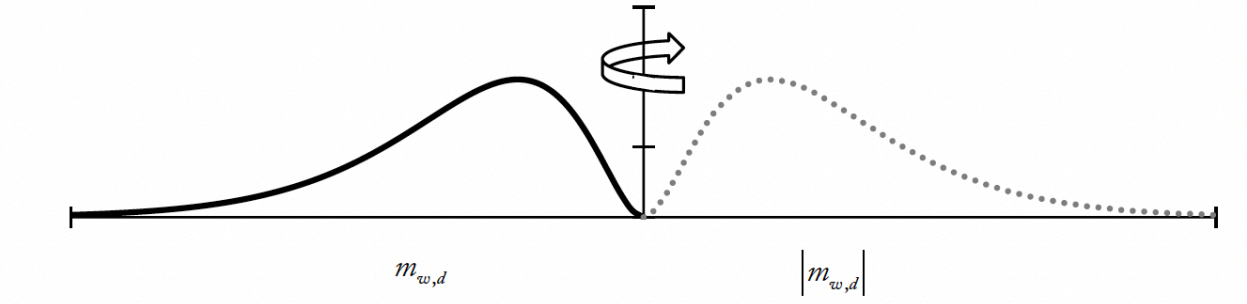
$$m_{w,d} = -\text{Gamma}\left[\text{abs}\left(m_{w,d}\right), \phi\text{abs}\left(m_{w,d}\right)\right]$$
$$m_{w,d} = \text{Gamma}\left[\text{abs}\left(m_{w,d}\right), \phi\text{abs}\left(m_{w,d}\right)\right] + 2m_{w,d}$$

Although the first option is more intuitive, the second option is preferred because the resulting distribution will be skewed to the right, as it should be for a gamma distribution.

Graphically, the two options have the following densities:

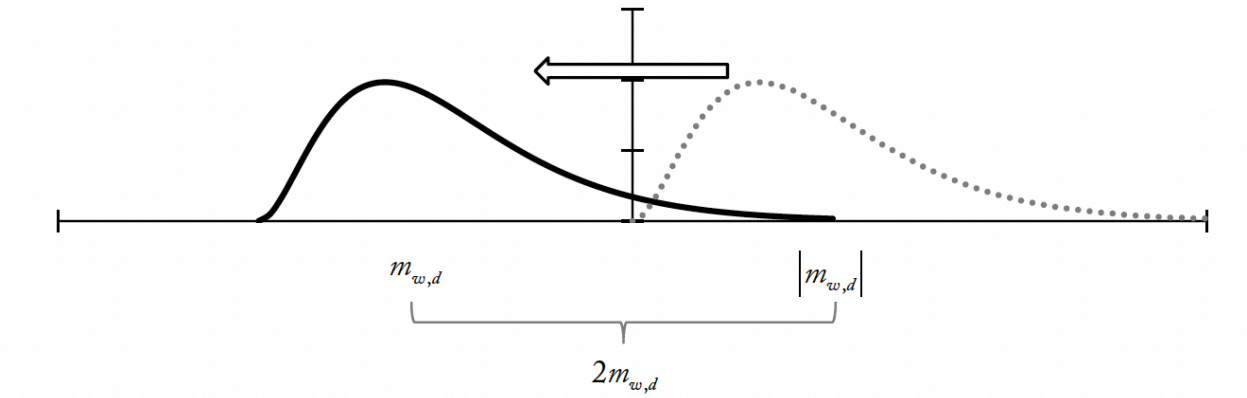
Option 1: Change the sign of the simulated value

$$-\text{Gamma}\left(\text{mean} = \left|m_{w,d}\right|, \text{var} = \phi \cdot \left|m_{w,d}\right|\right)$$



Option 2: Shift the entire distribution to have a mean of $m_{w,d}$

$$\text{Gamma}\left(\text{mean} = \left|m_{w,d}\right|, \text{var} = \phi \cdot \left|m_{w,d}\right|\right) + 2m_{w,d}$$



Under option 1, the density is skewed to the left and all values are negative. Under option 2, the density is skewed to the right and there is a mixture of positive and negative values. However, the mean is the same in both graphs.

Negative Values & Extreme Outcomes

Negative incremental values can cause extreme outcomes. This is most common when newly sampled triangles have negative incremental losses in the first few development columns. When a column of cumulative values sums close to zero and the next column sums to a large number, extremely large age-to-age factors and reserve estimates are produced.

To deal with extreme outcomes, the author provides the following **three options**:

1) Identify the extreme iterations and remove them
    - Be careful to only remove unreasonable extreme iterations so that the probability of extreme outcomes is not understated
2) Recalibrate the model
    - Identify the source of the negative incremental losses and remove it if necessary
    - For example, if the first row has negative incremental values due to sparse data, remove it and reparametrize the model

    - Alternatively, we could create separate models. For example, if negative values are caused by salvage/subrogation, we can model the gross losses and salvage/subrogation separately. Then, we can combine the iterations assuming 100% correlation
- Limit incremental losses to zero
    - This involves replacing negative incremental values with zeroes in the original triangles, zeroes in the sample triangles, OR zeroes in the projected future incremental losses

Non-Zero Sum of Residuals

Although the residuals are theoretically identically distributed with mean zero, they tend to have a non-zero average in practice.

Should the residuals be adjusted so that their average is zero?

- If the average of the residuals is positive, then resampling from the residuals will add variability to the resampled incremental losses. It may also cause the resampled incremental losses to have an average greater than the fitted loss. In this respect, the residuals should be adjusted
- Others argue that the non-zero average of the residuals is a characteristic of the data set. Thus, they should not be adjusted

If a residual average of zero is desired, we can add a single constant to all residuals such that the sum of the shifted residuals is zero.

Using an *N*-Year Weighted Average

At times, we may not use all years in the loss triangle when determining development factors. Instead, we may only use *N* years of data. The method for accomplishing this differs between the GLM bootstrap model and the ODP bootstrap model.

GLM Bootstrap Model

- We use N years of data by excluding the first few diagonals in the triangle, leaving us with $N + 1$ included diagonals
- This changes the shape of the triangle to a trapezoid
- The excluded diagonals are given zero weight in the model and fewer calendar year parameters are required
- When running the bootstrap simulations, we only need to sample residuals for the trapezoid that was used to parameterize the original model. The reason for this is that the GLM models incremental claims directly and can be parameterized using a trapezoid
- Each parameter set is then used to calculate future incremental values

ODP Bootstrap Model

- We calculate N-year weighted-average factors instead of all-year weighted-average factors
- We exclude the first few diagonals when calculating residuals
- Unlike the GLM bootstrap model, we must still sample residuals for the entire triangle when running bootstrap simulations. The reason for this is that the ODP bootstrap model requires cumulative values in order to calculate link ratios
- Once we have cumulative values for each sample triangle, we use N-year average factors to project the sample triangles to ultimate

Why does using N years of data leave us with $N + 1$ included diagonals? As an **example**, assume we have a 5 x 5 triangle, but only want to use a two-year weighted average. The required data in the triangle changes as follows:

AY	12 mo.	24 mo.
1	x_1	y_1
2	x_2	y_2
3	x_3	y_3
4	x_4	y_4

$\Rightarrow$

AY	12 mo.	24 mo.
1		
2		y_2
3	x_3	y_3
4	x_4	y_4

5	x_5		5	x_5	

In this example, the original 12-24 age-to-age factor is $\frac{y_1+y_2+y_3+y_4}{x_1+x_2+x_3+x_4}$. The new 12-24 age-to-age factor for $N = 2$ years of data is $\frac{y_3+y_4}{x_3+x_4}$. The new factor uses $N + 1 = 3$ diagonals:

- The first diagonal is comprised of x_3
- The second diagonal is comprised of x_4 and y_3
- The third diagonal is comprised of y_4

Missing Values

If values are missing from the loss triangle, the following calculations are affected:

- Loss development factors
- Fitted triangle (if the missing value lies on the last diagonal)
- Residuals
- Degrees of freedom

Once again, the method for dealing with missing values differs between the GLM bootstrap model and the ODP bootstrap model.

ODP Bootstrap Model

When using the ODP bootstrap model, there are multiple approaches to managing missing values:

- Estimate the missing value using surrounding values
- Exclude the missing value when calculating the loss development factors
    - If we do this, no corresponding residual will be calculated for the missing value.
    - Similar to N-year weighted averages, we must still sample for the entire triangle so that we can calculate cumulative values during the simulation process
    - Once the sample triangles are calculated, we should exclude the cells corresponding to the missing values from the projection process (i.e., when

calculating the age-to-age factors based on the sample triangles, the missing values should be excluded)

- If the missing value lies on the last diagonal, we can either estimate the value or we can use the value in the second to last diagonal to construct the fitted triangle

GLM Bootstrap Model

If using the GLM bootstrap model, the missing data simply reduces the number of observations used in the model. Similar to the ODP bootstrap model, we could use one of the methods described above to estimate the missing data if desired.

Outliers in the Original Dataset

If extreme values exist in the original triangle, and they are not considered representative of the future variability of the data, we can remove their impact from the model.

The method for dealing with outliers differs between the GLM bootstrap model and the ODP bootstrap model.

ODP Bootstrap Model

If using the ODP bootstrap model, approaches to managing outliers include the following:

- Exclude the outliers completely
    - In this case, we would proceed in the same manner as a missing value
- Exclude the outliers when calculating the age-to-age factors and the residuals (similar to missing values), but include the outlier cells during the sample triangle projection process (different from missing values)
    - The idea here is to remove the extreme impact of the incremental cell by excluding the outlier during the fitting process while still including some non-extreme variability by including the cell in the sample triangle projections

When excluding outliers from the calculation of age-to-age factors, there are **three options**:

1) Exclude the outlier in the numerator
2) Exclude the outlier in the denominator
3) Exclude the outlier in the numerator and the denominator

GLM Bootstrap Model

When using the GLM bootstrap model, outliers are treated similarly to missing data. If the data is not considered representative of real variability, the outlier should be excluded, and the model should be parameterized without it.

Example: Dealing with Outliers in the ODP Bootstrap Model

An actuary is running the ODP bootstrap model and has decided to deal with outliers by excluding them from the calculation of the age-to-age factors. Given the following incremental paid losses (in $) as of December 31, 2023:

AY	12 mo.	24 mo.
2019	100	80
2020	90	120
2021	105	85
2022	100	75

In this case, AY 2020 at 24 months appears to be an outlier. Let's walk through the three options for excluding outliers from the calculation of age-to-age factors:

1) Excluding the outlier in the numerator
    - 12-24 age-to-age factor = $\frac{180+190+175}{100+105+100} = 1.787$
2) Excluding the outlier in the denominator
    - 12-24 age-to-age factor = $\frac{180+210+190+175}{100+90+105+100} = 1.911$
    - Notice there is no exclusion here since the outlier only exists in the numerator

3) Excluding the outlier in the numerator and the denominator
    - 12-24 age-to-age factor = $\frac{180+190+175}{100+105+100} = 1.787$

What do we do if there are a significant number of outliers?

- This might indicate that the model is a poor fit to the data
- For the *GLM bootstrap model* , new parameters could be chosen, or the distribution of the error could be changed (i.e. the z parameter)
- For the *ODP bootstrap model*, an N-year weighted average could be used to provide a better model fit

Shapland also discusses *residual outliers*. Since the bootstrap process does not make a distributional assumption about the residuals, a large number of residual outliers may just mean that the residuals are highly skewed. If the skewness is real, then the outliers in the original dataset driving the residual outliers should be included in the fitting process to replicate the true nature of the residuals.

Heteroscedasticity

Since residuals are assumed to be i.i.d., we are able to apply a residual from one development period/accident period to the fitted loss in any other development period/accident period to produce sampled values. This is known as **homoscedasticity** (i.e., residuals have the same variance).

As mentioned before, if model assumptions are not quite met, it's possible for standardized residuals to exhibit more variability in certain development (or accident) periods than in other development (or accident) periods. This non-constant variance of the residuals is known as **heteroscedasticity**.

When assessing if heteroscedasticity is present, we must *keep credibility in mind*. There are fewer residuals as the development years become older, so comparing development years at the tail-end of the triangle is difficult.

Shapland covers three options to adjust for heteroscedasticity:

- Stratified sampling
- Calculating variance parameters
- Calculating scale parameters

Option 1: Stratified Sampling

Under this option, we do the following:

- Group development (or accident) periods with homogeneous variances
- Sample with replacement from the residuals in each group separately

An **advantage** of this method is that it is straightforward and easy to implement.

A **disadvantage** of this method is that some groups may only have a few residuals in them, which limits the amount of variability in the possible outcomes.

Option 2: Calculating Variance Parameters

Under this option, we do the following:

- Group development (or accident) periods with homogeneous variances
- Calculate the standard deviation of the standardized residuals in each of the "hetero" groups (each group has a different variance)
- Calculate the hetero-adjustment factor, h_i, for each group i as follows:

$$\boxed{h_i = \frac{stdev\left(\bigcup_1^j r_{w,d}^H\right)}{stdev\left(\bigcup_i r_{w,d}^H\right)}}$$

where $stdev\left(\bigcup_1^j r_{w,d}^H\right)$ is the standard deviation for all standardized residuals combined and $stdev\left(\bigcup_i r_{w,d}^H\right)$ is the standard deviation of the standardized residuals in group i. Notice that there is a total of j groups

- Multiply all residuals in group i by h_i as follows:

$$\boxed{r_{w,d}^{iH} = \frac{q(w,d) - m_{w,d}}{\sqrt{m_{w,d}^z}} \times f_{w,d}^H \times h_i}$$

All groups now have the same standard deviation, and we can sample with replacement from among all $r_{w,d}^{iH}$

- Divide the resampled residuals by the corresponding hetero-adjustment factor as follows:

$$\boxed{q^{i*}(w,d) = \frac{r^*}{h_i} \times \sqrt{m_{w,d}^z} + m_{w,d}}$$

This final adjustment ensures that the resampled triangles are consistent with the original triangle in terms of the residual distribution

Option 3: Calculating Scale Parameters

Under this option, we do the following:

- Group development (or accident) periods with homogeneous variances
- Calculate the total scale parameter using the unscaled Pearson residuals as follows:

$$\boxed{\begin{aligned}\phi &= \frac{\sum r_{w,d}^2}{N-p} \\ &= \frac{\frac{N}{N-p}\sum r_{w,d}^2}{N}\end{aligned}}$$

- Calculate the group i scale parameter using the unscaled Pearson residuals as follows:

$$\phi_i = \frac{\sum_{i=1}^{n_i}\left(\sqrt{\frac{N}{N-p}}\cdot r_{w,d}\right)^2}{n_i} = \frac{\frac{N}{N-p}\sum_{i=1}^{n_i} r_{w,d}^2}{n_i}$$

- Calculate the hetero-adjustment factor, h_i, for each group i as follows:

$$h_i = \frac{\sqrt{\phi}}{\sqrt{\phi_i}} = \frac{\sqrt{\frac{\frac{N}{N-p}\sum r_{w,d}^2}{N}}}{\sqrt{\frac{\frac{N}{N-p}\sum_{i=1}^{n_i} r_{w,d}^2}{n_i}}} = \frac{\sqrt{\frac{\sum r_{w,d}^2}{N}}}{\sqrt{\frac{\sum_{i=1}^{n_i} r_{w,d}^2}{n_i}}}$$

- Multiply all standardized residuals in group i by h_i as shown for the variance parameter option

- Divide the resampled residuals by the corresponding hetero-adjustment factor as shown for the variance parameter option

---

Example: Adjusting for Heteroscedasticity Using Variance Parameters

Given the following standardized residuals and standard deviations as of December 31, 2023:

	Standardized Residuals as of XX Months:				
AY	12 mo.	24 mo.	36 mo.	48 mo.	60 mo.
2019	160	40	-90	-140	0
2020	-45	-30	300	120	
2021	-150	-120	-200		
2022	40	100			
2023	0				

Development Period/Range	Standard Deviation
12	113.71
24	94.65
36	262.74
48	183.85
12-24	99.14
12-36	140.36
24-60	148.68
36-60	185.52
12-60	133.82

First, let's group the development periods with homogeneous variances:

- Based on the individual development period standard deviations, there are two reasonable groupings: 12-24 and 36-60

Second, let's calculate the hetero-adjustment factors for the two groups:

- Recall that $h_i = \frac{stdev\left(\cup_1^j r_{w,d}^H\right)}{stdev\left(\cup_i r_{w,d}^H\right)}$
- $h_1 = \frac{133.82}{99.14} = 1.350$
- $h_2 = \frac{133.82}{185.52} = 0.721$

Third, let's apply the hetero-adjustment factors to the standardized residuals to produce hetero-adjusted standardized residuals:

Hetero-Adjusted Standardized Residuals as of XX Months:

AY	12 mo.	24 mo.	36 mo.	48 mo.	60 mo.
2019	$215.98 = 160(1.350)$	54.00	$-64.92 = -90(0.721)$	-100.99	0
2020	-60.74	-40.50	216.40	86.56	
2021	-202.48	-161.99	-144.27		
2022	54.00	134.99			
2023	0				

Now that we have hetero-adjusted standardized residuals, we can randomly sample from the full set of standardized residuals above.

Suppose our sampling mechanism chooses the residual for 2021 at 36 months as the top left cell in the resampled triangle:

	Resampled Hetero-Adjusted Standardized Residuals as of XX Months:				
AY	12 mo.	24 mo.	36 mo.	48 mo.	60 mo.
2019	-144.27	xxx	xxx	xxx	xxx
2020	xxx	xxx	xxx	xxx	
2021	xxx	xxx	xxx		
2022	xxx	xxx			
2023	xxx				

Further suppose that the fitted loss for 2019 at 12 months is $250,000. Then, the resampled incremental loss is $q^{1*}(1,1) = \frac{-144.27}{1.350} \times \sqrt{250{,}000} + 250{,}000 = \$196{,}563$.

One thing students often ask is whether or not the corner zeroes should be included in the calculation of the standard deviations of ranges. In the Excel file provided with the paper, Shapland defaults to including the corner zeroes in the standard deviation calculations. For that reason, we recommend doing the same.

For example, the 36-60 standard deviation given in the problem is the sample standard deviation of -90, 300, -200, -140, 120, and 0. This can be calculated using the STDEV() function in Excel.

Example: Adjusting for Heteroscedasticity Using Scale Parameters

The same actuary that adjusted for heteroscedasticity using variance parameters wants to compare it to the adjustment using scale parameters. Given the following unscaled Pearson residuals as of December 31, 2023:

	Unscaled Pearson Residuals as of XX Months:				
AY	12 mo.	24 mo.	36 mo.	48 mo.	60 mo.
2019	120	30	-50	-95	0
2020	-15	-20	225	90	
2021	-125	-100	-190		
2022	30	80			
2023	0				

- The standardized residuals and corresponding standard deviations are the same as the prior example

First, let's group the development periods with homogeneous variances:

- Based on the individual development period standard deviations, there are two reasonable groupings: 12-24 and 36-60

Second, let's calculate the total scale parameter using the unscaled Pearson residuals:

- Recall that $\phi = \frac{\Sigma(r_{w,d})^2}{N-p}$
- Determining N
  - $N = 15$ since there are 15 cells in the triangle
- Determining p
  - Since we were not told otherwise, it's reasonable to assume that the model has 5 AY parameters (i.e., $\alpha_1 - \alpha_5$) and 4 development period parameters (i.e., $\beta_2 - \beta_5$)
  - Since hetero-adjustment factors act as parameters, we need to consider them as well. Although there are two hetero-adjustment factors, we only include one as a parameter since we could simply rebase the factors so that one of them is 1.000
  - In general, we need to include "**the number of hetero groups – 1**" as parameters
  - Thus, $p = 5 + 4 + 1 = 10$

- $\phi = \frac{120^2+(-15)^2+\cdots+90^2+0^2}{15-10} = 31{,}040$

Third, let's calculate the scale parameters for each hetero group:

- Recall that $\phi_i = \frac{\frac{N}{N-p}\sum_{i=1}^{n_i} r_{w,d}^2}{n_i}$
- $\phi_1 = \frac{\frac{15}{15-10}(120^2+(-15)^2+\cdots+(-100)^2+80^2)}{9} = 16{,}283$
- $\phi_2 = \frac{\frac{15}{15-10}((-50)^2+(225)^2+\cdots+(90)^2+0^2)}{6} = 53{,}175$

Fourth, let's calculate the hetero-adjustment factors for the two groups:

- Recall that $h_i = \frac{\sqrt{\phi}}{\sqrt{\phi_i}}$
- $h_1 = \frac{\sqrt{31{,}040}}{\sqrt{16{,}283}} = 1.381$
- $h_2 = \frac{\sqrt{31{,}040}}{\sqrt{53{,}175}} = 0.764$
- Alternatively, we could have calculated h_i using the final version of the formula where the $\frac{N}{N-p}$ terms are cancelled out. The nice thing about this version of the formula is that it doesn't require us to calculate the number of parameters

Fifth, let's apply the hetero-adjustment factors to the standardized residuals to produce hetero-adjusted standardized residuals:

Hetero-Adjusted Standardized Residuals as of XX Months:

AY	12 mo.	24 mo.	36 mo.	48 mo.	60 mo.
2019	$220.91 = 160(1.381)$	55.23	$-68.76 = -90(0.764)$	-106.96	0
2020	-62.13	-41.42	229.21	91.68	
2021	-207.10	-165.68	-152.80		
2022	55.23	138.07			
2023	0				

Now that we have hetero-adjusted standardized residuals, we can randomly sample from the full set of standardized residuals above.

Suppose our sampling mechanism chooses the residual for 2021 at 36 months as the top left cell in the resampled triangle:

	Resampled Hetero-Adjusted Standardized Residuals as of XX Months:				
AY	12 mo.	24 mo.	36 mo.	48 mo.	60 mo.
2019	-152.80	xxx	xxx	xxx	xxx
2020	xxx	xxx	xxx	xxx	
2021	xxx	xxx	xxx		
2022	xxx	xxx			
2023	xxx				

Further suppose that the fitted loss for 2019 at 12 months is \$250,000. Then, the resampled incremental loss is $q^{1*}(1,1) = \frac{-152.80}{1.381} \times \sqrt{250{,}000} + 250{,}000 = \$194{,}663$.

Before we end our discussion on heteroscedasticity, we should discuss its impact on the *incorporation of process variance* into future incremental values:

- Adjusting for heteroscedasticity using variance parameters
    - The total scale parameter should be adjusted to a group-specific scale parameter using $\phi_i = \frac{\phi}{h_i^2}$
    - Thus, when simulating future incremental losses, we draw from a gamma distribution with mean $m_{w,d}$ and variance $\phi_i m_{w,d}$
- Adjusting for heteroscedasticity using scale parameters
    - The group-specific scale parameters should be used when adding process variance to future incremental losses
    - Thus, when simulating future incremental losses, we draw from a gamma distribution with mean $m_{w,d}$ and variance $\phi_i m_{w,d}$

Heteroecthesious Data

Heteroecthesious data refers to incomplete or uneven exposures at interim evaluation dates. The two most common types of heteroecthesious data triangles are as follows:

- Partial first development period triangles
- Partial last calendar period triangles

Shapland discusses both of these from the viewpoint of the **ODP bootstrap model**.

Partial First Development Period Data

This occurs when the **first development column has a different exposure period** than the rest of the columns. For example, with annual data evaluated as of June 30, partial first development period data would have development periods ending at 6 months, 18 months, 30 months, etc.

How does this impact the bootstrap process?

- Parameterizing the model
    - **Not a problem** since the unscaled Pearson residuals use the square root of the fitted value to make them all "exposure independent"
- Projecting future incremental values
    - For each sample incremental triangle, **we must reduce** the projected future values for the latest AY to correspond to the earned exposure
    - For **example**, if the latest AY is only 6 months old, then the future values for the latest AY must be cut in half. No adjustments are needed for the remaining AYs since they are fully earned (i.e., at least 12 months old)
    - After adjusting the future values for the latest AY, we incorporate process variance using a gamma distribution as discussed before

Partial Last Calendar Period Data

This occurs when the latest diagonal is less than an annual period. For example, partial last calendar period data would have development periods ending at 12 months, 24 months, 36

months, etc. for all of the data in the triangle except the latest diagonal, which might have development periods ending at 6 months, 18 months, 30 months, etc.

How does this impact the bootstrap process?

- Parameterizing the model
    - We must **annualize the exposures in the latest diagonal** to make them consistent with the rest of the triangle
    - The fitted values (and consequently, the residuals) are based on this annualized triangle
- Projecting future incremental values
    - For each annualized sample incremental triangle, age-to-age factors are calculated and **interpolated**
    - The latest diagonal in the sample triangle is **adjusted back** to the partial period present in the original data
    - The cumulative values are then multiplied by the interpolated age-to-age factors to project future cumulative & incremental values
    - Once again, **we must reduce** the projected future values for the latest AY to correspond to the earned exposure
    - After adjusting the future values for the latest AY, we incorporate process variance using a gamma distribution as discussed before

Exposure Adjustment

A common issue in real data is exposures that have changed dramatically over the years. Examples include LOBs with rapid growth or run-off business.

Shapland discusses exposure adjustments under both the ODP bootstrap and GLM bootstrap models.

ODP Bootstrap Model

To adjust for this, we divide the claim data by earned exposures for each accident year. Then, the entire bootstrap process is run on the exposure-adjusted data.

After the "incorporate process variance step" is completed, we multiply the results by the earned exposures to restate them in terms of total values.

GLM Bootstrap Model

Similar to the ODP bootstrap model, the GLM bootstrap model is fit to the exposure-adjusted losses. However, there are two differences in the model fit as compared to the ODP bootstrap model:

- The fit is **exposure-weighted**, meaning that exposure-adjusted losses with higher exposure are assumed to have a lower variance
- The exposure adjustment could allow fewer AY parameters to be used

Once the fit is complete, the entire bootstrap process is run on the exposure-adjusted data. The GLM fits to the sample incremental triangles would need to be exposure-weighted fits to align with the fit to the original triangle.

After the "incorporate process variance step" is completed, we multiply the results by the earned exposures to restate them in terms of total values.

Tail Factors

Tail factors are commonly used to extrapolate development beyond the end of the triangle. We can incorporate them into the bootstrapping process.

ODP Bootstrap Model

Instead of using a deterministic tail factor, we can assign a distribution (ex. normal or lognormal) to the tail factor parameter. This makes the tail factor stochastic within the ODP bootstrap model.

Since the tail factor represents an "accumulation" of incremental factors, a single factor may not produce accurate incremental values. Thus, it might be useful to split the tail factor into its incremental counterparts using a decay model.

For **example**, suppose that the selected 60-ultimate tail factor is 1.05. This could be extrapolated over three years using a decay model. If the decay percentage is 50%, the 60-ultimate, 72-ultimate, and 84-ultimate tail factors would be 1.05, 1.025 (1 + 0.05(0.50)), and 1.013 (1+0.025(0.50)). This is what Shapland does in the Excel files provided with the text.

A **rule of thumb** for the tail factor standard deviation is 50% or less of the tail factor minus 1. So, for our example above, the standard deviation should be less than or equal to 0.50(1.05 – 1) = 0.025.

GLM Bootstrap Model

Earlier in the outline, we mentioned that modeling with fewer development trend parameters means that the last parameter is assumed to continue past the end of the triangle. This effectively gives us a tail without actually having to specify a tail factor.

To implement this, we simply assume that the final development period will continue to apply incrementally until its effect on the future incremental claims is negligible.

If we fit a GLM bootstrap model with development year *and* calendar year parameters, we would assume both continue past the end of the sample triangle until the effects on the future incremental claims are negligible.

Fitting a Distribution to the Residuals

Due to the limited number of observations in a loss triangle, it's difficult to determine whether the most extreme observation is a 1-in-100 event, 1-in-1000 event, etc.

If we believe that extreme observations are *not captured well* in the loss triangle, we can parameterize a distribution for the residuals (such as normal) and resample using the distribution. This is known as the **parametric bootstrap**.

V. Diagnostics

In order to assess the quality of a stochastic model, various diagnostic tests should be run. The diagnostic tools are designed for three purposes:

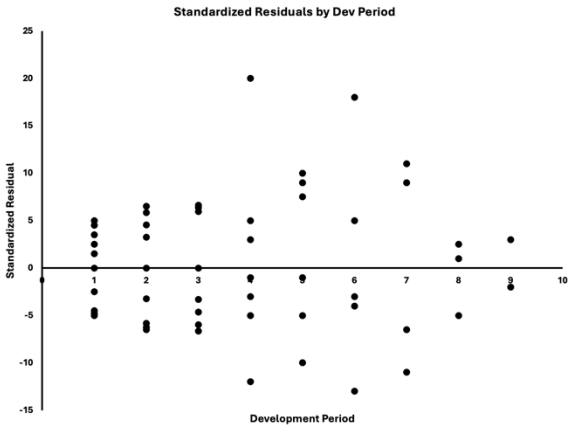
1) Test various assumptions in the model
2) Gauge the quality of the model fit
3) Guide the adjustment of model parameters

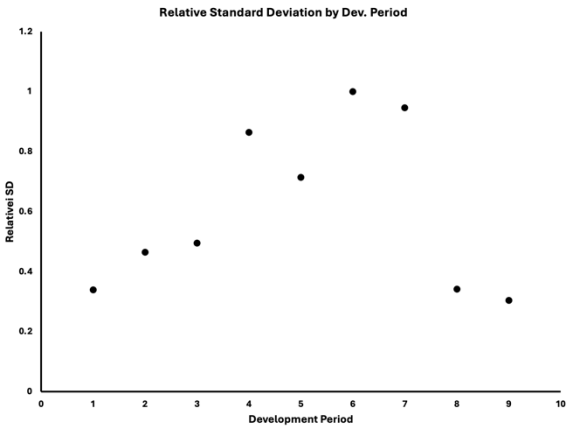
Residual Graphs

Residual graphs can be used to test the assumption that residuals are independent and identically distributed. We can graph the standardized residuals by development period, accident period, calendar period, or fitted incremental loss.

In each case, we should be able to draw a **relatively horizontal line** through the residuals. In addition, we want the residuals to show a **constant spread** across the graph. If the spread is non-constant, then heteroscedasticity exists, and hetero adjustments need to be made before bootstrapping the residuals.

To help *visualize* how residuals should be grouped for hetero adjustments, we can graph *relative standard deviations* and look for natural groupings. For example, consider the following plots:





In the plots above, there seem to be three groups of residuals (periods 1-3, periods 4-7, and periods 8-9). We can calculate hetero-adjustment factors for these groups and adjust them to a common standard deviation. After adjusting the residuals to a common standard deviation, we can run the graphs again to ensure that the residuals are homoscedastic.

Normality Test

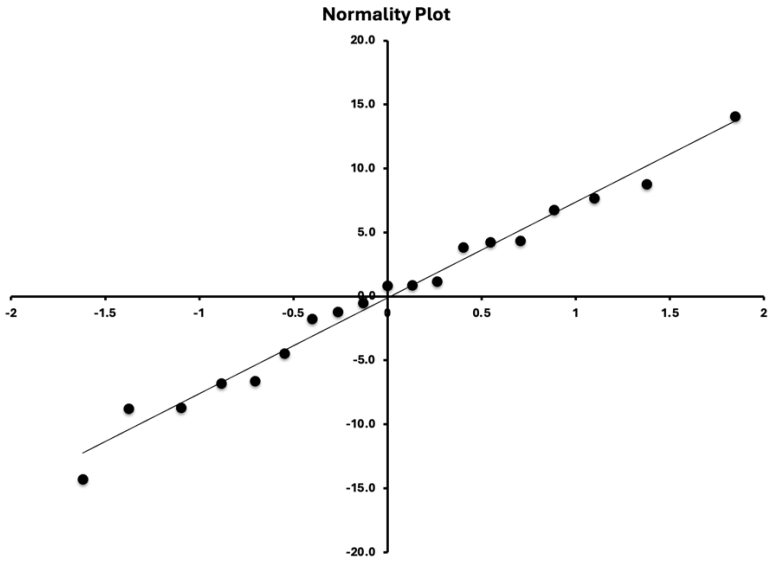
Although we do not require the residuals to be normally distributed, it's still helpful to compare residuals against a normal distribution. This allows us to compare parameter sets and assess the skewness of the residuals.

We can assess normality using plots and test values.

Normality Plot

A normality plot shows the relationship between theoretical quantiles of a standard normal distribution (i.e., the x-axis) and the empirical quantiles of the observed data (i.e., the y-axis). In our case, the observed data are the standardized residuals.

If the data is normally distributed, then the points in the plot will lie along a diagonal line. For **example**, consider the following normality plot for a set of standardized residuals:



In the graph above, the data points are tightly distributed around the diagonal line. This suggests that the standardized residuals are normally distributed.

Test Values

A number of test values can be used to assess normality:

- **p-value** from Shapiro's Test for Normality
    - If the residuals are normally distributed, the p-value should be large (i.e., greater than 5%)
- **R^2**
    - The diagonal line of the normality plot can be thought of as the regression line through the points
    - If the residuals are normally distributed, the R^2 should be close to 1
- **AIC/BIC**
    - p-values and R^2 fail to penalize for the number of parameters used in the model
    - To address this limitation, we can look at AIC and BIC. Both tests use the difference between each residual and its normal counterpart from the normality plot to calculate the Residual Sum Squared (RSS)
    - AIC and BIC are defined as follows:

$$\boxed{\begin{aligned} AIC &= 2p + n \times \left[\ln\left(\frac{2 \times \pi \times RSS}{n}\right) + 1\right] \\ BIC &= n \times \ln\left(\frac{RSS}{n}\right) + p \times \ln(n) \end{aligned}}$$

    - If residuals are normally distributed, the AIC and BIC should be small

For the normality plots and test values, we can run them before and after making heteroscedasticity adjustments. This enables us to see if applying the hetero groupings improved the fit of the model.

Keep in mind that failing the normality assumption does not mean the bootstrap model is poor. It just means that the residuals are not normally distributed.

Outliers

Shapland uses **box-whisker plots** of the standardized residuals to identify outliers in the data:

- The border of the box is formed using the inter-quartile range (the 25th to 75th percentiles)
- The median is graphed within the box
- The whiskers extend to the largest values within three times the inter-quartile range
- Values beyond the whiskers are considered outliers and are identified with a point

Once outliers are identified, judgment must be used to decide if they should be removed from the model fit:

- If the outliers represent extreme events *not expected to happen again*, they can be given *zero weight in the GLM* (or alternatively, we could remove them from the age-to-age factor calculation as described earlier)
- If the outliers represent extreme events that could happen again, they should be kept in the model

When residuals are not normally distributed, outliers tend to be more common. In this case, we don't necessarily want to remove them because they are representative of the shape of the data.

Parameter Adjustment

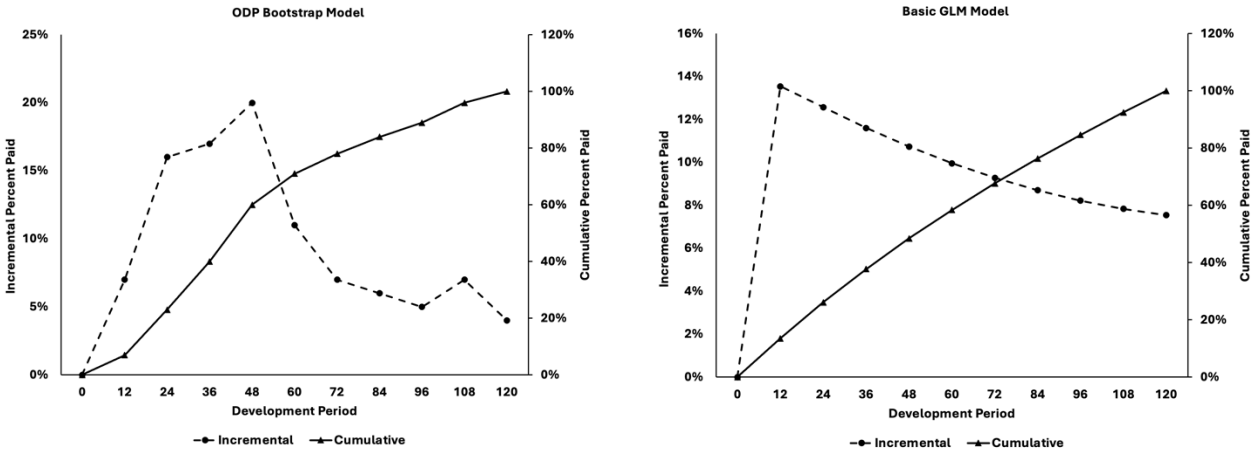
Including parameters for every accident period and development period may lead to an over-parameterized model (i.e., over-fitting the noise in the data). This is effectively what the ODP bootstrap model is doing.

The benefit of the GLM bootstrap model is that we can reduce the number of parameters until we find the optimal mix. The *principle of parsimony* states that a model with fewer parameters is preferred as long as the goodness of fit is not markedly different.

To find the optimal mix of parameters in the GLM bootstrap model, we do the following:

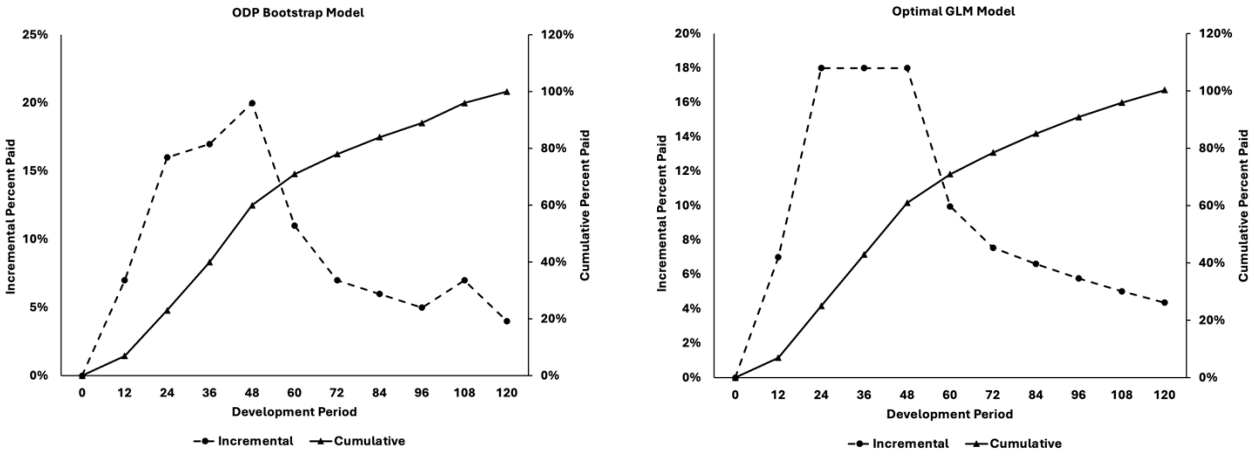
- Start with a "basic" GLM model which includes one parameter for accident, development, and calendar periods
- Check the residual plots for the basic model
    - If the residuals by accident period, development period, and calendar period are not randomly scattered around zero, then we should consider adding parameters
    - In addition, if certain parameters are not statistically significant, we should remove them
    - The implied development pattern should look like a smoothed version of the chain-ladder development pattern
- We keep adding/removing parameters until we see a proper residual plot

The following graphs show the implied development patterns for the ODP bootstrap model and a "basic" GLM model:



In the ODP bootstrap model, we have a different factor for each development period. This produces a jagged pattern in the incremental paid percentages. In the basic GLM model, we have a single development period parameter (i.e., a development period trend). This over-smooths the development pattern resulting in a poor fit.

The following graphs show the implied development patterns for the ODP bootstrap model and an "optimal" GLM model:



In the optimal GLM model, we have a set of parameters that balances parsimony and goodness of fit. In this case, the implied development pattern is a smoothed version of the ODP bootstrap model development pattern that still achieves the overall shape.

As we mentioned earlier, the optimal model should show no pattern in the residuals, indicating that all of the significant development period trends have been captured.

Estimated Unpaid Model Results

For each bootstrap model run, descriptive statistics (ex. mean, percentiles, standard deviation) relating to the bootstrap simulations can provide an additional diagnostic tool.

For the **estimated unpaid model results by AY**, Shapland recommends zooming in on the standard error and coefficient of variation (CoV).

Standard Error

We should see the following trends in the standard error:

- Standard error should *increase* from older to more recent years
    - This is because the standard error should follow the magnitude of the results

- Total standard error should be larger than the standard error for any individual year

CoV

We should see the following trends in the CoV

- CoV should *generally decrease* from older to more recent years
    - This is because older accident years have fewer payments remaining, which causes all of the variability to be reflected in the coefficient
    - For more recent years, random variations in the remaining payments tend to offset each other, which reduces overall variability
- Total CoV should be less than any individual year

The CoV may rise in the most recent years due to a couple of reasons:

- With an increasing number of parameters in the model, parameter uncertainty increases when moving from the oldest years to the most recent years. This parameter uncertainty may overpower the process uncertainty, causing an increase in variability
- The model may simply be overestimating the variability in the most recent years. In this case, the BF or Cape Cod models may need to be used in place of the CL model

Lastly, *since AY are assumed to be independent*, we should note that the standard error or CoV for all years combined will be *less* than the sum of the standard error or CoV for individual years. These points should *not be used as diagnostic tests* because they should be true of all models, including ones that are unreasonable.

Mean & Standard Deviation of Incremental Values

The mean and standard deviation of each incremental value (aggregated across all simulations) provide another diagnostic tool.

If the future mean or standard deviation values appear inconsistent, then this may help identify the source of any CoV issues that have surfaced (i.e., is it the mean or standard deviation causing the problem?)

Example: Reviewing the Mean & Standard Deviation of Incremental Values

Suppose we have the following incremental means from the simulated data after running the ODP bootstrap model as of December 31, 2023:

	Simulated Incremental Means ($) as of XX Months:				
AY	12 mo.	24 mo.	36 mo.	48 mo.	60 mo.
2019	2,005	1,502	800	402	207
2020	2,113	1,555	753	352	**178**
2021	2,225	1,409	727	**372**	**220**
2022	2,497	1,713	**844**	**368**	**196**
2023	2,365	**1,638**	**776**	**394**	**212**

In the table above, the means down every development period appear consistent. There are no future means that seem out of place.

Further suppose we have the following incremental standard deviations from the simulated data after running the ODP bootstrap model:

	Simulated Incremental SDs ($) as of XX Months:				
AY	12 mo.	24 mo.	36 mo.	48 mo.	60 mo.
2019	1,012	327	161	142	125
2020	1,216	313	202	148	**146**
2021	985	298	183	**164**	**181**
2022	1,165	304	**245**	**205**	**170**
2023	1,106	**500**	**320**	**231**	**191**

In the table above, the standard deviations down every development period show some inconsistencies. In particular, the standard deviations of the future incremental values appear much higher than the standard deviations for the historical period.

The resulting CoVs are as follows:

	Simulated CoVs as of XX Months:				
AY	12 mo.	24 mo.	36 mo.	48 mo.	60 mo.
2019	50.5%	21.8%	20.1%	35.3%	60.4%
2020	57.5%	20.1%	26.8%	42.0%	**82.0%**
2021	44.3%	21.1%	25.2%	**44.1%**	**82.3%**
2022	46.7%	17.7%	**29.0%**	**55.7%**	**86.7%**
2023	46.8%	**30.5%**	**41.2%**	**58.6%**	**90.1%**

In the table above, the higher standard deviations are driving higher CoVs for the future incremental values. In addition, the inconsistencies appear to be more severe in the latest two accident years. This is not completely unexpected since the CoV may rise in the latest few accident years due to the reasons mentioned earlier.

VI. Using Multiple Models

Multiple stochastic models should be run when performing a bootstrap analysis. Each model should go through diagnostic testing and all model output should be reviewed. Once the models have been reviewed, weights should be assigned to the models to combine results.

There are two methods for combining the results for multiple models:

- Run models with the same random variables
- Run models with independent random variables

Run Models with the Same Random Variables

Each model is run with the exact same random variables (i.e., random residuals in terms of position).

For **example**, suppose two models are run and each produces 50 residuals. When performing the random sampling portion of the bootstrap process, the first model chooses residual number 40

from the list of possible residuals and uses it in the first cell of the sample triangle. The second model would also choose residual number 40 from its list of possible residuals and use it in the first cell of its sample triangle. This process continues until both sample triangles are built.

Once all of the models have been run, the incremental values for each model are weighted together for each iteration by accident year.

This method for combining models causes correlation in model results since each model is run using the same set of random residuals.

Run Models with Independent Random Variables

Each model is run with its own random variables (i.e., different samples of random residuals in terms of position). In this case, the position of the random residual in the first model has no impact on the position of the random residual in other models. Once all of the models have been run, the weights are used to select a model for each iteration by accident year.

For **example**, suppose we are estimating unpaid losses using CL and BF bootstrap models. Further suppose the weights are 25% and 75%, respectively. For each iteration by accident year, we would draw a uniform random variable on (0, 1). If the drawn uniform random variable is "< 0.25", then the CL unpaid losses would be used for that iteration/AY combination. Otherwise, the BF unpaid losses would be used.

The result is a weighted mixture of models, where the CL and BF model results represent approximately 25% and 75% of the iterations by AY, respectively.

For detailed examples of each of the "model combining" methods above, see the Cookbook.

Other Notes on Weighting the Results of Different Models

Shapland discusses some other miscellaneous items regarding the weighting of results from different modes:

- The process of weighting the results of different models produces an *actuarial best estimate* of a distribution
- The weights can be determined using *Bayesian* methods that account for the quality of each model's forecasts
- The weights may produce results by accident year that do not reconcile with case reserves
    - For **example**, the weighted results for an accident year may produce negative IBNR. If negative IBNR is not reasonable for that particular LOB, then the weighted distributions can be "adjusted"
    - If the actuary believes that the shape and width of the distribution is appropriate, then they can add a fixed amount to the accident year to produce positive IBNR
    - Alternatively, if the actuary believes that the shape of the distribution is appropriate but would like to adjust the width, they can multiply the results by a factor to produce positive IBNR

Fitting Distributions to the Data

Once model results have been combined and tabulated, distributions can be fit to the data to smooth results.

For **example**, we can fit normal, lognormal, and gamma distributions to the total unpaid claim distributions.

The smoothed results can be used to:

- Assess the quality of the fit
- Parameterize a DFA (i.e., dynamic financial analysis) model
- Estimate extreme values
- Estimate TVaR (i.e., Tail Value at Risk)

The benefit of using smoothed results is that some of the random noise is prevented from distorting the calculations of specific metrics.

Estimated Cash Flow Results

Similar to the estimated unpaid model results, we can review a model's output by future calendar year. This is equivalent to reviewing "estimated cash flow results."

The **main difference** between future CYs (cash flows) and historical AYs (unpaid losses) is that the standard errors and CoVs move in opposite directions:

- For historical AYs, standard errors increase and CoVs decrease as we move from older to more recent years
- For future CYs, standard errors decrease and CoVs increase as we move further out into the future
    - Ex. If we are reviewing future cash flows for unpaid claims as of 12/31/23, the first future CY is 1/1/24 – 12/31/24

This makes intuitive sense for CYs. As we move further out into the future, the CY cash flows will decrease, which also leads to a decrease in the absolute standard error. However, relative to its mean, the variability increases substantially.

Estimated Ultimate Loss Ratio Results

The estimated ultimate loss ratios by AY are calculated using ***all*** simulated values, not just the values beyond the latest diagonal of the historical triangle:

- Since the simulated values represent additional possibilities of what could have happened in the past, we are able to estimate the ***complete*** variability in the loss ratio from day one until all claims are paid and settled
    - If we are **only interested in the future volatility of the loss ratios**, we can add the estimated unpaid claim estimates to the actual cumulative paid values and divide by the premiums
- By using all simulated values, we should have a fairly constant CoV across the accident years (since the standard errors should be proportional to the mean)
    - An increasing CoV would indicate large parameter uncertainty as described earlier

Distribution Graphs

We can create a total unpaid distribution histogram by dividing the range of all values generated from the simulation into equal size buckets and then counting the number of simulations that fall within each bucket (Shapland uses 100 buckets in the paper).

Dividing the number of simulations in each bucket by the total number of simulations results in a frequency/probability for each bucket.

Since simulation results tend to appear jagged, a *Kernel density* can be fit to the data to provide a smoothed distribution.

We can also *compare the individual models* used to create the best estimate by graphing each model's Kernel density function on the same graph.

Correlation

Once best estimates have been calculated for each business segment, results can be aggregated to the business unit level. Since business segments tend have specific correlations, results cannot simply be added together.

Shapland described two useful correlation process for the bootstrap model:

1) Location mapping
2) Re-sorting

Location Mapping

For each iteration:

- Sample residuals for segment 1
- Track the location in the residual triangle where each sampled residual was taken
- For all other segments, sample the residuals from their residuals triangles using the same locations

Advantages of location mapping are as follows:

- Easily implemented
- Does not require an estimated correlation matrix

Disadvantages of location mapping are as follows:

- Requires all business segments to have the same size data triangles with no missing data
- Since the correlation of the original residuals is used, we cannot test other correlation assumptions for stress testing purposes

Re-sorting

Re-sorting relies on algorithms such as Iman-Conover (rank correlation algorithm) or copulas to induce a desired correlation.

For **example**, we could induce correlation among business segments by re-sorting the residuals until the rank correlation between each business segment matches the desired correlation specified by a correlation matrix.

Advantages of re-sorting are as follows:

- Data triangles can be different shapes/sizes by segment
- Can use different correlation assumptions
- Different correlation assumptions may have other beneficial impacts on the aggregate distribution
    - Ex. We can use a copula with a heavy tail distribution to strengthen the correlation between segments in the tails, which is important for risk-based capital modeling