Arius®

Stochastic User Guide

Milliman

IT TAKES VISION

Milliman, Inc.
3424 Peachtree Road, NE Suite 1900
Atlanta GA 30326 USA

Tel +1 800 404 2276
Fax +1 404 237 6984

ActuarialSoftware.com

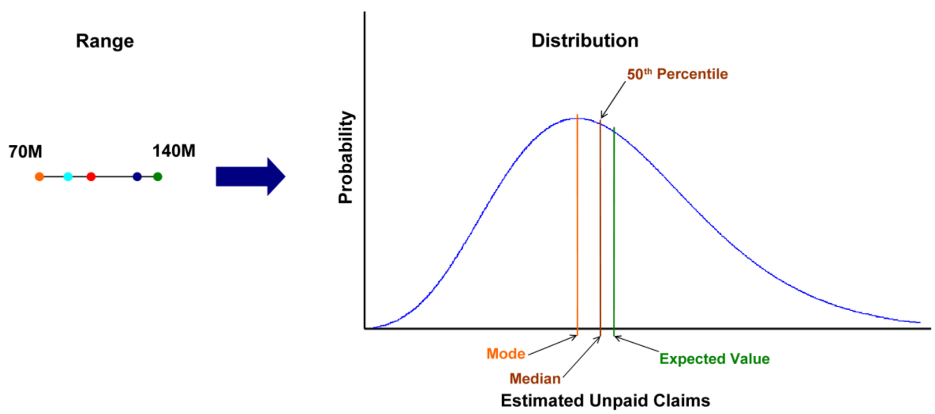
Table of Contents

1. General Background

In recent years, risk managers and other decision-makers responsible for loss reserving have been working in an information environment that is in a state of flux, but the actuarial components of this environment are only beginning to change. The actuarial world has been dominated by algorithmic or deterministic methods (i.e., results determined via defined steps from clearly stated assumptions) that each project a single point estimate of future liabilities from a set of existing data. These methods can be used with various combinations of assumptions to produce a "range of reasonable estimates" from which the actuary would judgmentally select the "best" estimate.

There is something beguiling about a range of estimates; especially one purported to be reasonable. But as every actuary knows, determining "reasonability" is ultimately a subjective process. It is influenced by a host of non-scientific but significant factors that, in their totality, define the specific business culture of any particular loss reserving process. Unfortunately, no range of estimates can account for every possible outcome. Indeed, these point estimates are each designed to reflect a central estimate of the possible outcomes and constitute the search for the pattern which leads to the "best central estimate."

In order to create a clearer statistical picture of the loss reserving process, actuaries have been turning to mathematical or stochastic models to capture the organic nature of real world loss information. As organizations begin to add stochastic models to their traditional approaches, corporate decision-makers will find that they have significantly more information regarding the unpaid claim liability estimates. For example, a traditional deterministic point estimate provides no information as to the risk that the ultimate result might eventually exceed the estimate. On the other hand, a stochastic-based estimate can provide a wealth of statistical information, or risk profile, about the unpaid claim liabilities (e.g., the 75th percentile, which is a reserve level at which there is a 25% chance that future payments might ultimately exceed the reserve).



Graph 1-1:
Ranges vs. Distributions

To be sure, the range of reasonable estimates produced by deterministic methods is still important. But the overriding task in the information-rich stochastic environment is to develop a model that captures the statistical features of the underlying data, and to search for "the models" which lead to a best estimate of "all" possible outcomes.

For example, one kind of stochastic model, the bootstrap, demonstrates the strengths, adaptability and utility of this approach. Bootstrapping originated with statisticians and is not new, nor unique to insurance. It is a tried and true model that looks to the dynamic nature of the data as the basis for simulations that generate a realistic distribution of possible outcomes. With respect to unpaid claim liability estimates, bootstrapping provides an estimated risk profile for a specific claims portfolio. It also provides important information for many other uses beyond just setting reserves (e.g., capital and return requirements, impact of diversification, evaluation of business strategies, etc.).

Like all models and methods, bootstrapping depends on the quality of the assumptions upon which it is based. Indeed, the actuarial judgments required in the deterministic world are no less important in the stochastic world Instead, the nature of these judgments are likely to change and, in some ways, may become a more important part of the process. Unlike traditional deterministic methods, however, which typically do not employ statistical tests to validate their overall reasonableness, the bootstrap model is open to a variety of diagnostic tools to help judge its underlying assumptions and to adjust its parameters to more realistically model the data at hand.

In summary, the information-rich environment created by stochastic modeling also helps the corporate decision-maker account for the needs of important audiences whose concerns drive the loss reserving process—regulators and rating agencies concerned with solvency, shareholders and investors concerned with investment return, and Board members concerned with making the best decision possible about the use of capital, and all parties concerned with better understanding the uncertain nature of unpaid claim liabilities. Different though their motives and goals may be, they can all be placed into a quantitative framework in order to better gauge risk using the power of stochastic analysis.[1]

METHODS VS. MODELS

A 2005 research paper by the Casualty Actuarial Society[2] defines two general categories of techniques for estimating unpaid claims[3] as follows:

Method*: A systematic procedure for estimating future payments for loss and allocated loss adjustment expense. Methods are algorithms or series of steps followed to determine an estimate.*

Model: *A mathematical or empirical representation of how losses and allocated loss adjustment expenses emerge and develop. The model accounts for known and inferred properties and is used to project future emergence and development.*

Using these definitions, the primary feature that distinguishes a model from a method is that a model can be used to estimate a “distribution of possible outcomes” whereas a method will only produce a single point estimate. In practice, a variety of assumptions and methods are typically used to generate a “range” of point estimates.

A variety of loss estimation models4 have been developed in recent years, though bootstrap and maximum likelihood models like the ones implemented in this system fall into the more sophisticated end of the spectrum. For example, while a bootstrap tool is based on a widely used and straightforward method (the link ratio method), it uses simulation techniques and statistical

1 Shapland, Mark R., “Risk Management Frontiers: The Quest for More Reserve Information”, Risk Management Magazine, June 2007, pp. 38-42.

2 CAS Working Party on Quantifying Variability in Reserve Estimates, “The Analysis and Estimation of Loss & ALAE Variability: A Summary Report”, 2005 CAS Fall Forum, pp. 29-146. Hereafter referred to as the “Reserve Variability Report.”

3 The term “loss” in this document is intended to include both loss and allocated loss adjustment expenses, unless noted otherwise. The Working Party research paper uses the more generic term “future payments.”

4 As a point of clarification, some methods can be “turned into” models and methods can be thought of as simplified models. For example, the Thomas Mack model can be viewed as an “extension” of the standard chain ladder method in the sense that it calculates statistics which can be used to “add” a distribution to the chain ladder estimate.

information derived from the data to generate a robust estimate of a distribution of possible outcomes.

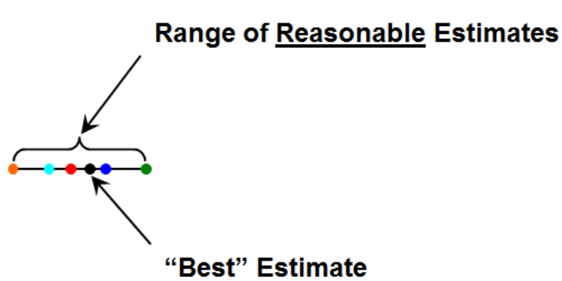
For both methods and models, the term “reserve” can also add to the confusion as it is commonly used to refer to both the estimated unpaid claims and the provision booked in the financial statements. In this manual we will restrict the use of the term “reserve” to the value used for financial reporting purposes. We will use the term “unpaid claim estimates" to refer to the actuarial evaluation of the value of these liabilities.

RANGES VS. DISTRIBUTIONS

Frequently, a casualty actuary will use either a range or distribution as an expression of the degree of uncertainty in the unpaid claim estimate. The approach used to develop a range or distribution of unpaid claim estimates may vary, and may even be dictated by the user’s intended purpose, or by the perspective of the actuary. A major distinction between a range and a distribution is that a “range” is generally considered to be either a set of point estimates or a subset of the possible outcomes whereas a “distribution” generally describes “all” possible outcomes.[5]

Another distinction is that if point estimates are used to determine the “range” then the statistical meaning of the points cannot readily be determined—e.g., we do not know if they represent a mean, median, or mode estimate[6]—whereas a “distribution” does have statistical meaning, e.g., the mean, median, mode, percentiles, and confidence intervals can be determined.

In order to bridge the gap between a traditional deterministic analysis and a stochastic analysis, we can use the term **central estimate** to mean a point estimate that is intended to convey a measure of central tendency rather than a deliberately “high” or “low” estimate – i.e., a central estimate is part of a subset of potential point estimates. In a deterministic analysis, the actuary would typically use a weighted average of a **range of reasonable estimates**[7] to determine their “best estimate.” This is illustrated in Graph 1-2.



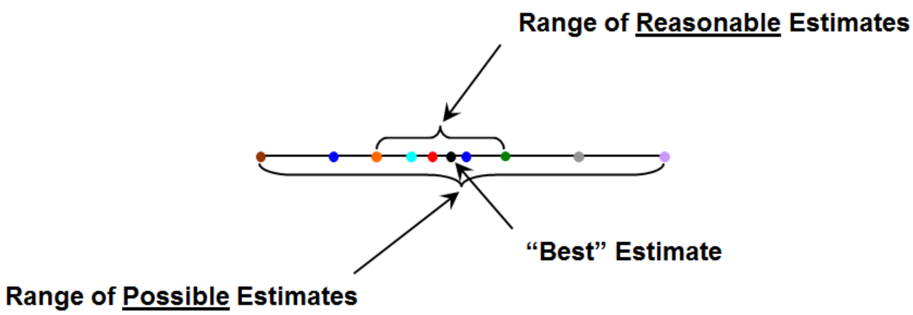
Graph 1-2:
Range of Reasonable Estimates & Best Estimate

5 While a purely statistical distribution will include all possible outcomes as defined by that distribution, the estimation of unpaid claims involves significant uncertainties that cannot be completely estimated. Thus, “all” should be thought of as a reasonable estimate of a distribution to the extent that it can be estimated using historical data (for example, “all” might not account for a potential catastrophe that has never occurred in the historical data being modeled).

6 Indeed, in Appendix 3, page 22, of Actuarial Standard of Practice No. 43, Property/Casualty Unpaid Claim Estimates, the subcommittee noted that “most traditional actuarial methods are meant to produce some measure of central tendency. But what measure? There are several different measures of central tendency, including (for example) mean, median, mode, and truncated mean. The subcommittee believed that ‘mean’ best represented the central tendency measure implicitly underlying most traditional actuarial methods, even if such traditional methods are not statistical in nature.”

7 A “range of central estimates” is analogous to a “range of reasonable estimates” as defined in Actuarial Standard of Practice No. 36.

In order to express more of the uncertainty of the "best estimate", one alternative is to expand the range of reasonable estimates to include other point estimates. For example, the actuary could include point estimates in which the average age-to-age ratios are deliberately higher and/or lower than expected in order to show deliberately higher and/or lower point estimates. In contrast to a range of reasonable estimates, a **range of possible estimates** is intended to show a wider range than what is used to determine the best estimate. However, note that the width of both of these ranges are subjective based on the judgment of the actuary and neither can be used to convey statistical information. These ranges are illustrated in Graph 1-3.

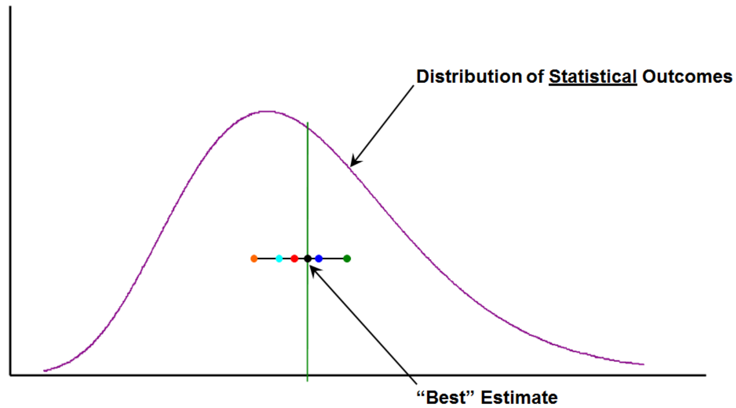


Graph 1-3:
Range of Reasonable Estimates vs. Range of Possible Estimates

Another distinction between a range and a distribution is that the incremental values projected in a point estimate will essentially have the random movements "averaged" or "smoothed" out, whereas the incremental values in a **possible outcome** will include random movements. This is an important distinction since a point estimate (in part or in total) may not be a possible outcome even though it still has a valid meaning for accounting purposes. For example, for the roll of a fair die the possible outcomes are 1, 2, 3, 4, 5 and 6 but the central estimate is 3.5, which is not one of the possible outcomes. The central estimate of 3.5 is an appropriate estimate for most accounting purposes,[8] even though it is not a possible outcome, whereas the possible outcomes convey statistical information about the risk of the central estimate being redundant or deficient.

Another common process for describing the uncertainty of the central estimate is to "add" statistical information to the central estimate. For example, the standard Thomas Mack approach can be used to calculate the standard deviation of the unpaid claims by accident year and for all years combined. Using the standard deviation and an assumed distribution, the actuary can then calculate statistical values corresponding to a distribution around the central estimate. This is illustrated in Graph 1-4.

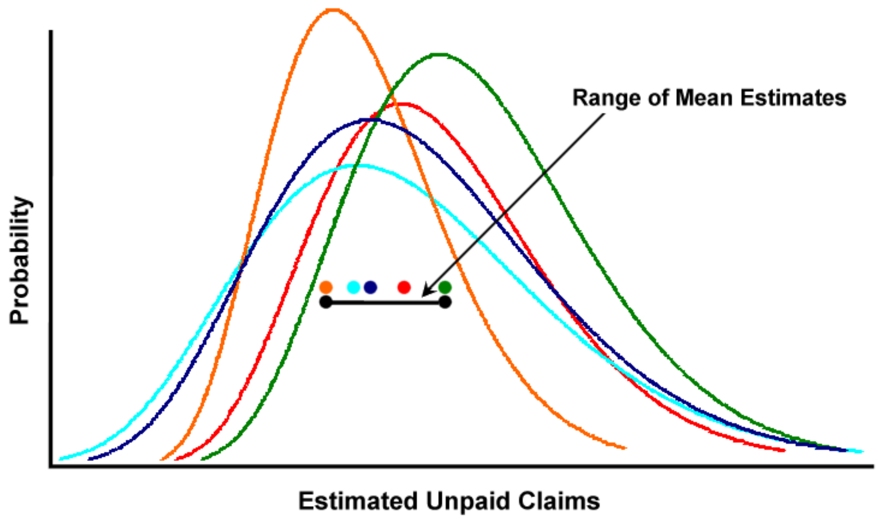
8 Since the possible outcomes are symmetric, the central estimate is also the mean and median values in this example. Some accounting principles describe using the mode (the most likely outcome), but six different modes is not realistic when compared to insurance liabilities. Discounting and other Fair Value accounting rules would modify this example, but they are ignored here for simplicity.



Graph 1-4:
Distribution of Statistical Outcomes

A **distribution of statistical outcomes** can be defined as a distribution based on a central estimate, standard deviation and a statistical distribution. This is an "improvement" over a range of central estimates in the sense that statistical information (e.g., percentiles, confidence intervals, etc.) can be determined, but the central estimate is assumed to be the mean of the distribution and implied points of the distribution may or may not be possible outcomes. In contrast, a **distribution of possible outcomes** can be defined as a distribution in which each of the points in the distribution is based on incremental values that are all random and statistically possible. This transition to a model that produces a distribution of possible outcomes does not imply that we should only be concerned with finding the one "best" model. Indeed, since all models use simplifying assumptions no one model will always produce the "best" distribution any more than any one method will always produce the "best" central estimate.

In many ways, the use of various models is analogous to the use of a variety of methods to develop a range of central estimates. As part of the analysis of various methods, actuaries explicitly and implicitly use diagnostic tools to evaluate the results of each method in order to determine if the point estimate from each specific method (and its assumptions) is reasonable. For example, ratios of incurred to paid claims can show any trends in the adequacy of case reserves over time which would inform the actuarial judgment with respect to the paid and incurred chain ladder methods. As another example, reviewing the accident year data and age-to-age factors informs judgment about the quality of the estimate for each year and whether the overall estimate from a particular method is reasonable.



Graph 1-5:
Reasonable Distributions of Possible Outcomes

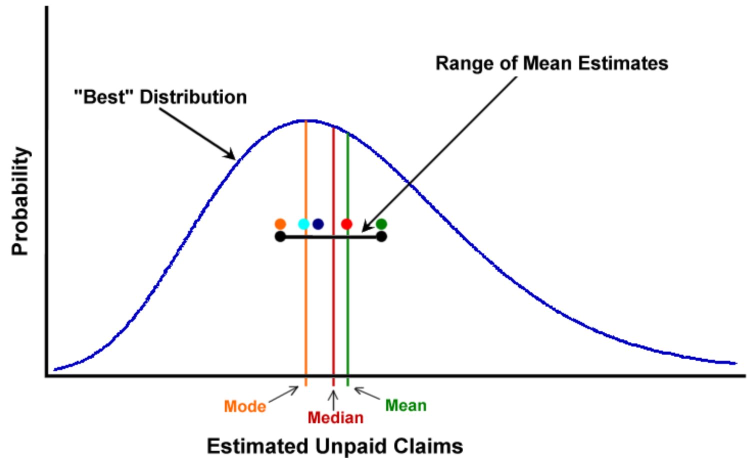
Similarly, in addition to the "deterministic" diagnostic tools actuaries are generally familiar with, a variety of statistically based diagnostic tools are available to determine if a model is reasonable.[9] For example, the statistically based diagnostic tests, described more completely in Section 5 and Appendix B, can be used to adjust model assumptions to match the statistical features of the data and to review the statistical quality of the model results. This process will result in a reasonable distribution from each reasonable model.[10] Graph 1-5 illustrates a variety of reasonable distributions from different reasonable models. Just like for methods, no one model is definitive so different models can be used to illustrate the uncertainty of a particular estimate (e.g., the means in this illustration).[11]

Similar to the process used to weight different reasonable point estimates, it is possible to credibility weight the reasonable distributions in order to derive a "best estimate of the distribution," as illustrated in Graph 1-6. From the "best estimate of the distribution" it is then possible to determine the "best" estimate of the mode, median, and mean. Comparing Graphs 1-5 and 1-6, you can see the link between statistically based ranges and a resulting weighted "best distribution." Adding Graph 1-2 to the comparison, you can also contrast a deterministic range with a statistically based range.

9 See Sections 3-3.1, pp. 43-51, of the Reserve Variability Report for a discussion of twenty criteria, or diagnostic tests, for evaluating the reasonability of a model. See Section 5 and Appendix B of this manual for the diagnostics used in the Milliman model.

10 This also assumes that models deemed not to be reasonable have been removed.

11 As a point of clarification, we could also compare a "range of mode estimates" or a "range of median estimates" to the deterministic "range of central estimates."



Graph 1-6: Best Estimate of a Distribution

THE EFFECTS OF CORRELATION AMONG LINES

Just as distributions can provide more information to your analysis of a particular line of business, they can also provide additional help when you summarize your analysis of several different lines.

You know from experience that different types of insurance tend to react to market stimuli in different ways. If two lines of business, say, Auto BI and Auto PD, both typically tend to behave similarly to market conditions, they are said to be “positively correlated.” (The opposite could also be true, resulting in “negatively correlated” lines.) Stated another way, if the two lines are positively correlated, and Auto BI produced higher than expected losses, we would anticipate Auto PD to also produce higher than expected results.

How does this relate to distributions? You could summarize results to arrive at an aggregate mean by summing the means of the individual lines. But if you want an **aggregate distribution**, you need to factor in the effects of this correlation into your calculation. For a simple example, if two lines are less than 100% correlated, you would probably need less total reserves at the 75% level than the sum of the two lines’ individual 75% levels, due to their propensity to not go “bad” at the same time. Arius provides extensive capabilities to measure and account for this potential correlation effect among multiple lines of business. Appendix C provides additional details about this subject.

[THIS PAGE INTENTIONALLY LEFT BLANK]

2. What is a Bootstrap Model?

A LITTLE HISTORY

The term “bootstrapping” originates in German literature, from legends about Baron von Münchhausen, who was known for the often unbelievable tales he told of his fantastic adventures. Supposedly he could lift himself out of a swamp or quicksand by pulling himself up by his own hair. In later versions of similar tales, he pulled himself out of the sea by pulling up on his bootstraps, thus forming the basis for the term bootstrapping.

The term has taken on broad application in many fields of study, including physics, biology and medical research, computer science, and statistics. Though ultimately having grown to mean different things in each of the above applications, they generally start from the premise of using simple information, and building it bit by bit into more complex systems, often using only one’s own data to estimate or project more complex information about that data (like the Baron, who needed nothing more than what he already had to pull himself out of the sea).

AN UNPAID CLAIM ESTIMATION MODEL

Bradley Efron, chairman of the Department of Statistics at Stanford University, is credited with expanding the concept of the bootstrap estimate into the realm of statistics. In his work, "bootstrap" means that one available sample gives rise to many others by re-sampling the existing data (not unlike pulling yourself up by your own bootstrap). He suggests duplicating the original sample as many times as computing resources will allow, and then treating this expanded sample as a virtual population. Then samples are drawn with replacement from this population to verify the estimators.

Several writers in actuarial literature have applied this concept to the process of loss reserving. The most commonly-cited examples are from England and Verrall (1999 and 2002), Pinheiro, et al. (2001 and 2003), and Kirschner, et al. (2002). In its simplest form, they suggested using a basic chain ladder technique to square a triangle of paid losses, repeating that randomly and stochastically a large number of times, and then evaluating the distribution of the outcomes. The model generates a distribution of possible outcomes, rather than the chain ladder’s typical point estimate, thus providing more information about the potential results. Assuming the users understand the data, and how well the data fits the model, they can draw more effective inferences, given the resulting mean, standard deviation and various percentiles available.

STRENGTHS OF THE BOOTSTRAP APPROACH

A primary advantage to using a bootstrap model (or any of Arius’ stochastic models) is to estimate the distribution of possible outcomes, which in turn provides information about the “riskiness” of the portfolio of claims. For example, without an estimated distribution it is virtually impossible to directly estimate the amount of capital required[12] or how likely it is that the ultimate value of the claims will exceed a certain amount.

Another advantage of a bootstrap model is that it involves aggregate loss triangles of both paid and incurred losses and volume-weighted age-to-age factors (i.e., the chain ladder method), which should

12 Without an estimated distribution, required capital could be “estimated” using industry benchmark ratios or other rules of thumb, but these do not directly account for the specific risk profile under review.

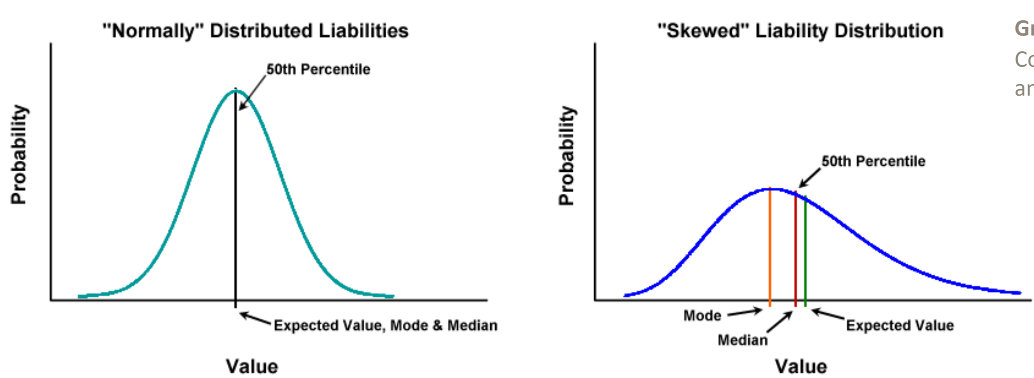
be familiar to most people.[13] The use of aggregate triangle data means that the required data should be readily available in most cases and perhaps even readily tie to published financial statements.

While the bootstrap model has components that are commonly known and understood, it also has components that are sophisticated enough to produce a robust estimate of the distribution of possible outcomes. Even for users with a modest mathematical background, the sophisticated components are straightforward and easy to learn.

Because the framework of the bootstrap model is based on the chain ladder method, other methods can also be used with a bootstrap model, thus giving us multiple models we can credibility-weight into a more robust distribution than we could derive with only one model. In the Milliman model, the user can model using both paid and incurred data models based on the chain ladder, Bornhuetter-Ferguson, and Cape Cod methods, for a total of six different models. The basic mechanics of these models are illustrated in Section 3 and Appendix A.

Another advantage of a bootstrap model is that it can be specifically “tailored” to the statistical features found in the data under analysis. This is particularly important as the results of any simulation model are only as good as the model used in the simulation process. If the model does not “fit” the data then the results of the simulation may not be a very good estimate of the distribution of possible outcomes.

A final advantage of a bootstrap model is that it can reflect the fact that insurance loss distributions are generally “skewed to the right.” Rather than the generally recognized normal distribution (which is often used as a simplifying assumption in other models), wherein the outcome is equally likely to be higher or lower than expected, in a right skewed distribution the higher end possibilities are further away from the average compared to the lower end – i.e., when an outcome is better than expected, there is a limit on how good it could be, but when an outcome is worse than expected the degree to which the outcome can be worse can be much greater. These two different types of distribution are illustrated in the Graph 2-1.



Graph 2-1:
Comparison of Symmetric and Skewed Distributions

13 The volume-weighted age-to-age factors are derived from a Generalized Linear Model (GLM), but understanding the theoretical background is not a prerequisite for using the factors.

This “distributional effect” is greatly influenced by the presence of (or lack of) large claims and large reinsurance recoveries, among other things, which can be generally reflected in a bootstrap model.

BOOTSTRAP MODEL+ SHORTCOMINGS

Like all models and methods, the quality of a bootstrap model depends on the quality of the assumptions. A number of diagnostics are available to help evaluate how well a model fixes a particular data set (or vice versa). We will elaborate on some of the important model diagnostics in Section 5 and Appendix C.

Another aspect of bootstrap models that could be considered a disadvantage is that they are more complex than standard methods and thus more time consuming to create. However, once a flexible model has been developed, as Milliman has done, they can be used as efficiently as most standard methods.

A corollary to this is that at first bootstrap models can appear more difficult to explain and understand. In part, this could be due to more widespread use of standard methods compared to the newer bootstrap models, but we will endeavor to lay out the process used by the Milliman models in order to assist with understanding.

Another potential weakness of a bootstrap model is the limited number of data points used to parameterize the model (e.g., 53 in a typical 10 x 10 triangle). This makes it hard to determine whether the most extreme observation is a one-in-100 or a one-in-1,000 event (or simply, in this example, a one-in-53 event). Of course, the nature of the extreme observations in the data will also affect the level of the extreme simulations in the results. In order to overcome this potential weakness, other sampling options are included in the Milliman model so that the sampling process is not limited to the available data.

SOLVENCY USING THE BOOTSTRAP APPROACH

An emerging area of use for the bootstrap model is for solvency regulation. In Europe the new Solvency II regime is requiring risk bearing entities to calculate reserve risk and Solvency Capital Requirements (SCR) on a “one-year time horizon” basis. The concept behind this one-year time horizon is to project new results based on different scenarios after one year of time after the current financial reporting period. Since the bootstrap model projects results incrementally into the future, it is ideally suited to effectively recalculate future estimates of unpaid claims associated with each outcome of simulated results after one year.

In addition to the SCR, the new technical provisions in Europe require a risk margin calculated based on the runoff of the SCR over time. Accordingly, we have included new methods of using the bootstrap models to not only simulate on a one year time horizon, but an N-year horizon so that the users can effectively “run-off” the implied capital (or SCR) at the one year time horizon.

[THIS PAGE INTENTIONALLY LEFT BLANK]

3. Overview: How the Simulation Models Work

A BASIC ODP BOOTSTRAP PAID LOSS CHAIN LADDER SIMULATION

For purposes of providing a simple review of the algorithm's steps, we will follow the most basic form, a simulation of possible future outcomes based on the paid loss triangle and the basic chain ladder approach.

With random sampling from a triangle of residuals (i.e., the model error terms), the model simulates a large number of "sample" triangles, uses the chain ladder model to estimate the future payment triangles (lower right), including the random nature of those payments, and then calculates a distribution from the many possible outcomes of future payments.

In very simple terms, the model performs the following steps (of course, the reality is a bit more complex):

1. Use a triangle of cumulative paid losses as input. Calculate the average age-to-age development factors. (Initially we're using the chain ladder model with volume weighted averages.)
2. Calculate a new triangle of "fitted values" – i.e., use the average age-to-age factors to "undevelop" each value in the latest diagonal to form a new triangle of values predicted by the model assumptions.
3. Working from the incremental versions of these triangles, calculate a triangle of residuals using the fitted triangle and the original data. These are called "unscaled Pearson residuals" in the model.
4. Standardize the residuals so that they are independent and identically distributed (i.i.d.) and calculate the scale parameter (used for the process variance in Step 7).
5. Create a new incremental sample triangle by selecting randomly with replacement from among the triangle of standardized Pearson residuals.[14]
6. Develop and square that sample triangle, adding tail factors, and estimating ultimate losses.
7. Add process variance to the future incremental values from Step 6 (which will change the "estimated ultimate" to a "possible outcome").
8. Calculate the total future payments (estimated unpaid amounts) for each year and in total for this iteration of the model.
9. Repeat the random selection, new triangle creation, and resulting unpaid calculations in Steps 5 through 8, *X* times.

The result from the *X* simulations is an estimate of a distribution of possible outcomes. From this we can calculate the mean, standard deviation, percentiles, etc.

Appendix A has a numerical example of the Paid Chain Ladder simulation model process, as well as the other models, described in Section 3.

14 Other options for simulating the sample triangle include simulating the residuals form a normal distribution or using the fitted incremental values as the mean of a normal, lognormal or gamma distribution (similar to adding process variance in Step 8).

A BASIC ODP BOOTSTRAP INCURRED LOSS CHAIN LADDER SIMULATION

We can also use incurred loss data and follow the same steps as outlined above for paid data. When paid data is used, the resulting distribution is for total unpaid amounts since we are estimating the difference between the ultimate payments and the payments to date. However, when incurred data is used (with the same algorithm) the resulting distribution is for the total incurred but not reported (IBNR) amount because we are estimating the difference between the ultimate values and the incurred to date.

In order to make apples-to-apples comparisons of each distribution of unpaid claims, as opposed to IBNR, the Milliman incurred model includes an additional step to convert the squared triangle of incremental incurred amounts to a squared triangle of incremental paid amounts.

So why not just add the case reserves to the distribution of IBNR to get a distribution of total unpaid claims that will match the paid method? This would indeed result in a distribution of total unpaid claims, but it would not be a consistent comparison since there would be no variation in the case reserve portion of the distribution. By using the steps outlined below for the incurred data, the simulated sample triangles will include variations in the case reserves which results in a complete apples-to-apples comparison of the unpaid claims.

In very simple terms, the model performs the following steps:

1. Use a triangle of cumulative incurred losses as input. Calculate the average age-to-age development factors. (Initially we're using the chain ladder model with volume weighted averages.)
2. Calculate a new triangle of "fitted values" – i.e., use the average age-to-age factors to "undevelop" each value in the latest diagonal to form a new triangle of values predicted by the model assumptions.
3. Working from the incremental versions of these triangles, calculate a triangle of residuals using the fitted triangle and the original data. These are the "unscaled Pearson residuals."
4. Standardize the residuals so that they are (i.i.d.) and calculate the scale parameter (used for the process variance in Step 7).
5. Create a new incremental sample triangle by selecting randomly with replacement from among the triangle of standardized Pearson residuals.[15]
6. Develop and square that sample triangle, adding tail factors, and estimating ultimate losses.
7. Add process variance to the future incremental values from Step 6 (which will change the estimated ultimate) and calculate the ultimate value for each year estimated by this iteration of the model.
8. Simulate in parallel using paid losses (Steps 1 to 7). Adjust the simulated incremental paid losses so that the total ultimate value for each accident year matches the total ultimate value for the incurred losses in Step 7 (which adjusts the incurred ultimate to a random paid pattern).
9. Calculate the total future payments (estimated unpaid amounts) for each year and in total for this iteration of the model, using the paid losses after they are adjusted to match the ultimate incurred amounts.

Note:
The paid and incurred calculations are identical through Step 7.

Note:
Since the ultimate values are the same, the loss ratio distribution will still be based on the incurred data.

15 The same options for simulating a sample triangle with paid loss data also apply to incurred loss data.

10. Repeat the random selection, new triangle creation, and resulting adjusted unpaid calculations in Steps 5 through 9 *X* times.

The result from the *X* simulations is an estimate of the distribution of possible outcomes. From this we can calculate the mean, standard deviation, percentiles, etc.

A BASIC ODP BOOTSTRAP BORNHUETTER-FERGUSON SIMULATION

We have extended both the paid and incurred chain ladder models by incorporating the Bornhuetter-Ferguson (BF) method into the model steps. The BF bootstrap model requires an additional set of parameters for the ultimate losses (i.e., the mean or *a priori*) per year. Optional parameters for the BF model include:

- Ultimate Premiums (to express the ultimate losses as a loss ratio); or
- Ultimate Exposures (to express the ultimate losses as a pure premium); and
- Coefficient of variation of the *a priori* ultimate losses (to add uncertainty to the ultimate loss parameters).

Replacing Step 6 of the basic chain ladder model, the ultimate loss will be calculated using the BF method. Using the paid BF with premiums as an example:

Ultimate Loss = Paid to date + (1 – %Paid) x Ultimate Premium x Loss Ratio

The loss ratio is simulated from the distribution you selected if you want to include uncertainty in the ultimate. The paid-to-date has been simulated in Step 5. The BF method gives the total remaining unpaid amount for that year.

The total unpaid needs to be divided into incremental losses. This can be done in two different ways.

- Deterministic Option: Use the age-to-age factors to determine the proportion that should fall in each incremental period; or
- Statistical Option: Using a Bayesian weighting of the column sums and row sums to determine the proportion that should fall in each incremental period.[16]

After this step, we can proceed to Step 7 of the normal chain ladder model, where we add process variance to the future incremental values.

The incurred BF model uses the same extra step to convert the total IBNR to total unpaid like the incurred chain ladder model, except that it will use a parallel paid BF model instead of a paid chain ladder model in Step 8 and the simulated loss ratios used in the incurred BF model will also be used in the parallel paid BF model.[17] Keeping with the simplicity theme, we only illustrate the paid BF model in Appendix A.

16 The Statistical option is based on "Verrall, Richard J. 2004. A Bayesian Generalized Linear Model for the Bornhuetter-Ferguson Method of Claims Reserving. North American Actuarial Journal, Vol. 8-3. p. 67-89." This option can be a useful alternative when the development pattern is not stable.

17 The paid BF model uses the same loss ratio assumptions (e.g., a priori and CoV) as the incurred BF model, but loss ratios for each iteration are simulated independently. Only for the paid BF used as part of the incurred model are the same simulated loss ratios used.

A BASIC ODP BOOTSTRAP CAPE COD SIMULATION

Similarly, we have extended both the paid and incurred chain ladder models by incorporating the Cape Cod method into the model steps. The Cape Cod bootstrap model requires these additional parameters:

- Premium index factors (to adjust ultimate premiums to current rate level);
- Loss trends (to adjust for loss cost inflation);
- Weights (to mark which periods to include in the loss ratio weighted average); and
- Decay rate (to systematically reduce the weight given to each year the further each year is from the year being calculated).

Step 6 of the basic chain ladder model will use the Cape Cod method to calculate the ultimate loss. The advantage of the Cape Cod over the Bornhuetter-Ferguson is that the uncertainty of the ultimate values can be based on the stochastic iterations instead of an input variable.

The calculation of the incremental values for the next step can be done using the same two options (i.e., the “Deterministic” and “Statistical” options) as the BF model.

After this step, we can proceed to Step 7 of the normal chain ladder model, where we add process variance to the future incremental values.

The incurred Cape Cod model uses the same extra step to convert the total IBNR to total unpaid as the incurred chain ladder model, except that it will use a parallel paid Cape Cod model instead of a paid chain ladder model in Step 8. Keeping with the simplicity theme, we will illustrate only the paid Cape Cod model in Appendix A.

USING THE ODP PROCESS ALGORITHM[18]

A Paid Loss Chain Ladder Simulation

For purposes of providing a simple review of the OPD Process algorithm’s steps for calculating risk on an N-Year time horizon, we will follow the most basic form, a simulation of possible future outcomes based on the paid loss triangle and the chain ladder approach.

In very simple terms, the model performs the following steps, many of which are consistent with the ODP Bootstrap basic chain ladder simulation noted earlier:

1. Use a triangle of cumulative paid losses as input. Calculate the average age-to-age development factors (using volume weighted averages).
2. Calculate a new triangle of “fitted values” – i.e., use the average age-to-age factors to “undevelop” each value in the latest diagonal to form a new triangle of values predicted by the model assumptions.
3. Working from the incremental versions of these triangles, calculate a triangle of residuals using the fitted triangle and the original data. These are called “unscaled Pearson residuals” in the model.

18 The ODP Process Algorithm could alternatively be called the Wüthrich & Merz algorithm.

4. Standardize the residuals so that they are independent and identically distributed (i.i.d.) and calculate the scale parameter (used for the process variance in Step 7).
5. Create a new incremental sample triangle by selecting randomly with replacement from among the triangle of standardized Pearson residuals.[19]
6. Develop and square that sample triangle, adding tail factors, and estimating ultimate losses.
7. Add process variance to the future incremental values from Step 6 (which will change the estimated ultimate to a possible outcome).
8. Use the original data triangle and the first N future diagonals to recalculate the age-to-age factors up to the Nth diagonal and use these factors project the remaining expected unpaid.
9. Calculate the total N year future payments and remaining point estimate for each year and in total for this iteration of the model.
10. Repeat the random selection, new triangle creation, and resulting unpaid calculations in Steps 5 through 9, *X* times.

The result from the *X* simulations is an estimate of a distribution of possible outcomes for the first N diagonals and point estimates beyond the N diagonals. From this we can calculate the mean, standard deviation, percentiles, etc. To calculate the Claim Development Result, the average from the basic ODP Bootstrap model is subtracted from each iteration of this model. It is important to note that for this algorithm the first N diagonals will be identical to the results for the basic paid chain ladder algorithm.

An Incurred Loss Chain Ladder Simulation

In very simple terms, the model performs the following steps, many of which are consistent with the basic incurred chain ladder noted earlier:

1. Use a triangle of cumulative incurred losses as input. Calculate the average age-to-age development factors (using volume weighted averages).
2. Calculate a new triangle of "fitted values" – i.e., use the average age-to-age factors to "undevelop" each value in the latest diagonal to form a new triangle of values predicted by the model assumptions.
3. Working from the incremental versions of these triangles, calculate a triangle of residuals using the fitted triangle and the original data. These are the "unscaled Pearson residuals."
4. Standardize the residuals so that they are (i.i.d.) and calculate the scale parameter (used for the process variance in Step 7).
5. Create a new incremental sample triangle by selecting randomly with replacement from among the triangle of standardized Pearson residuals.
6. Develop and square that sample triangle, adding tail factors, and estimating ultimate losses.
7. Add process variance to the future incremental values from Step 6 (which will change the estimated ultimate) and calculate the ultimate value for each year estimated by this iteration of the model.

Note:
The paid and incurred calculations are identical through Step 7.

19 Other options for simulating the sample triangle also still apply.

8. Simulate in parallel using paid losses (Steps 1 to 7). Adjust the simulated incremental paid losses so that the total ultimate value for each accident year matches the total ultimate value for the incurred losses in Step 7 (which adjusts the incurred ultimate to a random paid pattern).
9. Use the original incurred data triangle and the first N future incurred diagonals to recalculate the age-to-age factors up to the Nth diagonal and use these factors project the remaining expected IBNR.
10. Use the original paid data triangle and the first N future paid diagonals to recalculate the age-to-age factors up to the Nth diagonal and use these factors to project the remaining expected unpaid for the parallel paid loss portion of the incurred model.
11. Use the original paid triangle and the adjusted incremental paid from Step 8 for the first N diagonals only. Calculate the difference between the “ultimate” incurred values by year in Step 9 and the cumulative paid values up to the Nth future diagonal in Step 11, and then allocate this remaining expected incurred unpaid to the remaining future incremental values using the remaining expected paid pattern from Step 10 (which adjusts the remaining incurred ultimate to an expected payment pattern).
12. Repeat the random selection, new triangle creation, and resulting adjusted unpaid calculations in Steps 5 through 11 ***X*** times.

The result from the ***X*** simulations is an estimate of the distribution of possible outcomes for the first N diagonals and point estimates beyond the N diagonals. From this we can calculate the mean, standard deviation, percentiles, etc. To calculate the Claim Development Result, the average from the basic ODP Bootstrap model is subtracted from each iteration of this model. It is important to note that for this algorithm the first N diagonals will be identical to the results for the basic incurred chain ladder algorithm.

A Bornhuetter-Ferguson Simulation

The adjustments for the paid and incurred Bornhuetter-Ferguson (BF) methods for the ODP Process algorithm are similar to the adjustments to the basic chain ladder methods. The BF bootstrap model requires an additional set of parameters for the ultimate losses (i.e., the mean or *a priori*) per year, as noted above.

Replacing Step 6 of the basic chain ladder model, the ultimate loss will be calculated using the BF method. The paid-to-date has been simulated in Step 5. The BF method gives the total remaining unpaid amount by year, as noted above. To maintain internal consistency, the sampled *a priori* loss ratios by year used for the basic BF model are also used for the ODP Process algorithm without resampling them.

The total unpaid needs to be divided into incremental losses. For the ODP Process algorithm, only the Deterministic Option noted above can be used. After this step, we can proceed to Step 7 of the basic chain ladder model, where we add process variance to the future incremental values. Step 8 of the paid ODP Process algorithm also uses the same sampled a priori ratios to estimate the remaining unpaid, again without resampling.

The incurred BF model uses the same extra step to convert the total IBNR to total unpaid like the incurred chain ladder model, except that it will use a parallel paid BF model instead of a paid chain ladder model in Step 8 and the simulated loss ratios used in the incurred BF model will also be used in the parallel paid BF model.[20] It is important to note that for these algorithms the first N diagonals will be identical to the results for the basic paid and incurred BF algorithms, respectively. Keeping with the simplicity theme, we only illustrate the paid BF model in Appendix A.

A Cape Cod Simulation

Similarly, we have extended both the paid and incurred chain ladder models by incorporating the Cape Cod method into the model steps. The Cape Cod bootstrap model requires these additional parameters, as noted above.

Step 6 of the basic chain ladder model will use the Cape Cod method to calculate the ultimate loss. The calculation of the incremental values for the next step can only be done using the Deterministic option for the ODP Process algorithm.

After this step, we can proceed to Step 7 of the normal chain ladder model, where we add process variance to the future incremental values.

The incurred Cape Cod model uses the same extra step to convert the total IBNR to total unpaid as the incurred chain ladder model, except that it will use a parallel paid Cape Cod model instead of a paid chain ladder model in Step 8. It is important to note that for these algorithms the first N diagonals will be identical to the results for the basic paid and incurred CC algorithms, respectively. Keeping with the simplicity theme, we will illustrate only the paid Cape Cod model in Appendix A.

USING THE ODP RESIDUAL ALGORITHM[21]

A Paid Loss Chain Ladder Simulation

For purposes of providing a simple review of the OPD Residual algorithm's steps for calculating risk on an N-Year time horizon, we will follow the most basic form, a simulation of possible future outcomes based on the paid loss triangle and the basic chain ladder approach.

In very simple terms, the model performs the following steps, many of which are consistent with the ODP Bootstrap basic chain ladder simulation noted earlier:

1. Use a triangle of cumulative paid losses as input. Calculate the average age-to-age development factors (using volume weighted averages).
2. Calculate a new triangle of "fitted values" – i.e., use the average age-to-age factors to "undevelop" each value in the latest diagonal to form a new triangle of values predicted by the model assumptions.
3. Working from the incremental versions of these triangles, calculate a triangle of residuals using the fitted triangle and the original data. These are called "unscaled Pearson residuals" in the model.
4. Standardize the residuals so that they are independent and identically distributed (i.i.d.).

20 The paid BF model uses the same loss ratio assumptions (e.g., a priori and CoV) as the incurred BF model, but loss ratios for each iteration are simulated independently. Only for the paid BF used as part of the incurred model are the same simulated loss ratios used.

21 The "Residual" (or "Recursive") approach to simulating based on a one-year time horizon is described by Alexandre Boumezoued et al. in "One-year Reserve Risk Including a Tail Factor: Closed Formula and Bootstrap Approaches," 2012. See Help | Technical References for the compete paper.

5. Create a new incremental sample trapezoid by selecting randomly with replacement from among the triangle of standardized Pearson residuals and using the age-to-age factors from Step 1 to determine the future fitted values.
6. Develop and square that sample trapezoid, adding tail factors, and estimating ultimate losses.
7. Sum the future incremental values from Step 6 (which is a possible outcome for the N-Year time horizon and expected values for the remaining future periods).
8. Repeat the random selection, new trapezoid creation, and resulting unpaid calculations in Steps 5 through 7, *X* times.

The result from the *X* simulations is an estimate of a distribution of possible outcomes for the first N diagonals and point estimates beyond the N diagonals. From this we can calculate the mean, standard deviation, percentiles, etc. To calculate the Claim Development Result, the average from the basic ODP Bootstrap model is subtracted from each iteration of this model. It is important to note that for this algorithm the first N diagonals will be different from the results for the simple paid chain ladder algorithm. It is important to note that for this algorithm the sample triangles will be identical to the results for the basic paid chain ladder algorithm, but all of the future values will differ.

An Incurred Loss Chain Ladder Simulation

In very simple terms, the model performs the following steps, many of which are consistent with the basic incurred chain ladder noted earlier:

1. Use a triangle of cumulative incurred losses as input. Calculate the average age-to-age development factors (using volume weighted averages).
2. Calculate a new triangle of “fitted values” – i.e., use the average age-to-age factors to “undevelop” each value in the latest diagonal to form a new triangle of values predicted by the model assumptions.
3. Working from the incremental versions of these triangles, calculate a triangle of residuals using the fitted triangle and the original data. These are the “unscaled Pearson residuals.”
4. Standardize the residuals so that they are independent and identically distributed (i.i.d.).
5. Create a new incremental sample trapezoid by selecting randomly with replacement from among the triangle of standardized Pearson residuals and using the age-to-age factors from Step 1 to determine the future fitted values.
6. Develop and square that sample trapezoid, adding tail factors, and estimating ultimate losses.
7. Using the ***basic OPD Bootstrap chain ladder model*** results, adjust the simulated incremental paid losses so that the total ultimate value for each accident year matches the total ultimate value for the incurred losses (which adjusts the incurred ultimate to a random paid pattern).
8. Simulate in parallel using paid losses (Steps 1 to 6).
9. Calculate the difference between the “ultimate” incurred values by year in Step 6 and the cumulative paid values for the triangle in Step 7, and then allocate this remaining expected incurred unpaid to the remaining future incremental values using the remaining expected paid pattern from Step 8 (which adjusts the remaining incurred ultimate to an expected payment pattern).
10. Repeat the random selection, new trapezoid creation, and resulting adjusted unpaid calculations in Steps 5 through 9, *X* times.

Note: The paid and incurred calculations are identical through Step 6.

Note: Step 7 uses results from the basic model, not the ODP Residual model, so the triangle results are identical.

Note: Step 8 uses the ODP Residual steps to run in parallel. The future incremental pattern is a combination of possible outcomes and expected values.

The result from the *X* simulations is an estimate of the distribution of possible outcomes for the first N diagonals and point estimates beyond the N diagonals. From this we can calculate the mean, standard deviation, percentiles, etc. To calculate the Claim Development Result, the average from the basic ODP Bootstrap model is subtracted from each iteration of this model. It is important to note that for this algorithm the first N diagonals will be different from the results for the simple incurred chain ladder algorithm. It is important to note that for this algorithm the sample triangles will be identical to the results for the basic incurred chain ladder algorithm, but all of the future values will differ.

A Bornhuetter-Ferguson Simulation

The adjustments for the paid and incurred Bornhuetter-Ferguson (BF) methods for the ODP Residual algorithm are similar to the adjustments to the basic chain ladder methods. The BF bootstrap model requires an additional set of parameters for the ultimate losses (i.e., the mean or *a priori*) per year, as noted above.

Adjusting Step 5 of the ODP Residual chain ladder model, the expected values for N future diagonals will be calculated using the expected values of the BF method (i.e., without randomness). The squaring of the trapezoid in Step 6 will similarly be replaced with the BF method. The BF method gives the total remaining unpaid amount by year, as noted above. To maintain internal consistency, the sampled *a priori* loss ratios by year used for the basic BF model are also used for the ODP Residual algorithm without resampling them.

The total unpaid needs to be divided into incremental losses. For the ODP Residual algorithm, only the Deterministic Option noted above can be used.

The incurred BF model uses the same extra step to convert the total IBNR to total unpaid like the incurred chain ladder model, except that it will use a parallel paid BF model instead of a paid chain ladder model in Step 8 and the simulated loss ratios used in the incurred BF model will also be used in the parallel paid BF model. It is important to note that for these algorithms the sample triangles will be identical to the results, but all of the future values will differ from results for the basic paid and incurred BF algorithms, respectively. Keeping with the simplicity theme, we only illustrate the paid BF model in Appendix A.

A Cape Cod Simulation

Similarly, we have extended both the paid and incurred chain ladder models by incorporating the Cape Cod method into the model steps. The Cape Cod bootstrap model requires these additional parameters, as noted above.

Step 5 of the ODP Residual chain ladder model will use the Cape Cod method to calculate the expected values for N future diagonals. The squaring of the trapezoid in Step 6 will similarly be replaced with the CC method. The calculation of the incremental values for the next step can only be done using the Deterministic option for the ODP Process algorithm.

The incurred Cape Cod model uses the same extra step to convert the total IBNR to total unpaid as the incurred chain ladder model, except that it will use a parallel paid Cape Cod model instead of a paid chain ladder model in Step 8. It is important to note that for these algorithms the sample triangles will be identical to the results, but all of the future values will differ from results for the basic paid and incurred CC algorithms, respectively. Keeping with the simplicity theme, we will illustrate only the paid Cape Cod model in Appendix A.

A BASIC MACK BOOTSTRAP ULTIMATE PAID CHAIN LADDER SIMULATION

For purposes of providing a simple review of the Mack Bootstrap model we will follow the most basic form, a simulation of possible future outcomes based on the paid loss triangle and the chain ladder approach.

In very simple terms, the model performs the following steps, most of which are different than the basic ODP Bootstrap chain ladder simulation noted earlier:

1. Use a triangle of cumulative paid losses as input. Calculate a triangle of age-to-age development factors and the average age-to-age factors (using volume weighted averages).
2. Calculate a triangle of residuals using the age-to-age factors and averages. These are called "unscaled residuals" in the model.
3. Calculate the standard deviations of the residuals and standardize the residuals so that they are (i.i.d.).
4. Create a new sample triangle of age-to-age factors by selecting randomly with replacement from among the triangle of standardized residuals[22] and calculate the volume weighted average of the sample age-to-age factors.
5. Use the average factors from Step 4 to project one diagonal of expected cumulative values, adding tail factor, as appropriate.
6. Add process variance to the future cumulative values along the diagonal from Step 5 (which will change the point estimate to a possible outcome).
7. Add the diagonal from Step 6 to the original data triangle and repeat steps 5 and 6 until the all future diagonals are complete. (This is an iterative process which adds one diagonal at a time until all the diagonals are complete.)
8. Repeat the random selection, new diagonal creation, and resulting unpaid calculations in Steps 4 through 7, *X* times.

The result from the *X* simulations is an estimate of a distribution of possible outcomes for the future diagonals. From this we can calculate the mean, standard deviation, percentiles, etc. It is important to note that this algorithm is COMPLETELY different than all other models described above. This model is only available for paid data and this chain ladder approach.

USING THE MACK TIME HORIZON ALGORITHM

A Paid Loss Chain Ladder Simulation

For purposes of providing a simple review of the Mack Horizon algorithm's steps for calculating risk on an N-Year time horizon, we will follow the most basic form, a simulation of possible future outcomes based on the paid loss triangle and the chain ladder approach.

In very simple terms, the model performs the following steps, most of which are different than the simple chain ladder simulation noted earlier:

1. Use a triangle of cumulative paid losses as input. Calculate a triangle of age-to-age development factors and the average age-to-age factors (using volume weighted averages).

22 Other options for simulating the sample triangle do not apply to this model.

2. Calculate a triangle of residuals using the age-to-age factors and averages. These are called "unscaled residuals" in the model.
3. Calculate the standard deviations of the residuals and standardize the residuals so that they are (i.i.d.).
4. Create a new sample triangle of age-to-age factors by selecting randomly with replacement from among the triangle of standardized residuals and calculate the volume weighted average of the sample age-to-age factors.
5. Use the average factors from Step 4 to project one diagonal of expected cumulative values, adding tail factor, as appropriate.
6. Add process variance to the future cumulative values along the diagonal from Step 5 (which will change the point estimate to a possible outcome).
7. Add the diagonal from Step 6 to the original data triangle and repeat steps 5 and 6 until the first N future diagonals are complete. (This is an iterative process which adds one diagonal at a time until the first N diagonals are complete.)
8. Recalculate the age-to-age factors up to the Nth diagonal and use these factors to project the remaining expected unpaid.
9. Repeat the random selection, new diagonal creation, and resulting unpaid calculations in Steps 4 through 8, *X* times.

The result from the *X* simulations is an estimate of a distribution of possible outcomes for the first N diagonals and point estimates beyond the N diagonals. From this we can calculate the mean, standard deviation, percentiles, etc. To calculate the Claim Development Result, the average from the basic Mack Bootstrap model is subtracted from each iteration of this model. It is important to note that this algorithm is COMPLETELY different than all other models described above. This model is only available for paid data and this chain ladder approach.

A BASIC HAYNE MLE INCREMENTAL FREQUENCY SIMULATION

For purposes of providing a simple review of the Hayne MLE model we will follow the most basic form, a simulation of possible future outcomes based on the reported claim count triangle and the ultimate exposures.

In very simple terms, the model performs the following steps, all of which are different than the basic ODP Bootstrap chain ladder simulation noted earlier. While there are 4 different variations of this model (Berquist Sherman, Cape Cod, Chain Ladder and Hoerl Curve), the basic steps are the same for each and only the details of Steps 2 through 5 vary depending on the specific model variation chosen.

1. Use a triangle of cumulative reported claim counts and ultimate exposures as input. Calculate a triangle of incremental claim frequencies by dividing the triangle of claim counts by the ultimate exposures by period, then by taking the differences of the cumulative values to get incremental values.
2. Use maximum likelihood to fit the selected model to the incremental frequency triangle. This will result in a mean and standard deviation for all the parameters, as well as a variance–covariance matrix.
3. Calculate the predicted means and standard deviations of all incremental cells for the entire square. The historical predictions (triangle) can be used to calculate residuals, etc. to examine the goodness of fit of the model. The future predictions can be used to approximate the future distributions.

4. Create a new sample of the model parameters using the multivariate normal distribution and the parameters from Step 2.
5. Use the sample parameters from Step 4 to calculate sample means and standard deviations for all incremental cells in the entire square. [Note: Multiplying these mean values times the ultimate exposure by period would result in a point estimate.]
6. Add process variance to the values from Step 5 by sampling each cell from the normal distribution using the mean and standard deviation values by cell (which will change the point estimate to a possible outcome).
7. Calculate the sample claim counts by multiplying each incremental cell times the ultimate exposures by period. Use just the future incremental claim counts to derive an estimate of the unreported claims by period.
8. Repeat the random parameter sampling, sample square creation, random process variance and resulting unreported calculations in Steps 4 through 7, *X* times.

The result from the *X* simulations is an estimate of a distribution of possible outcomes for the future unreported claim counts. From this we can calculate the mean, standard deviation, percentiles, etc. It is important to note that this algorithm is COMPLETELY different than all other models described above. This model is only available for reported claim count data. Normally, the averages of the unreported claim counts is used to select the ultimate claim counts for use with the Hayne MLE Incremental Severity models. Since the Hayne MLE Incremental Severity model algorithm is similar to the Hayne MLE Incremental Frequency model, only the Incremental Severity models are illustrated in Appendix A.

A BASIC HAYNE MLE INCREMENTAL SEVERITY SIMULATION

For purposes of providing a simple review of the Hayne MLE model we will follow the most basic form, a simulation of possible future outcomes based on the paid loss triangle and the ultimate claim counts.

In very simple terms, the model performs the following steps, all of which are different than the basic ODP Bootstrap chain ladder simulation noted earlier. While there are 4 different variations of this model (Berquist Sherman, Cape Cod, Chain Ladder and Hoerl Curve), the basic steps are the same for each and only the details of Steps 2 through 5 vary depending on the specific model variation chosen.

1. Use a triangle of cumulative paid losses and ultimate claim counts as input. Calculate a triangle of incremental claim severities by dividing the triangle of paid losses by the ultimate claim count by period, then by taking the differences of the cumulative values to get incremental values.
2. Use maximum likelihood to fit the selected model to the incremental severity triangle. This will result in a mean and standard deviation for all the parameters, as well as a variance–covariance matrix.
3. Calculate the predicted means and standard deviations of all incremental cells for the entire square. The historical predictions (triangle) can be used to calculate residuals, etc. to examine the goodness of fit of the model. The future predictions can be used to approximate the future distributions.
4. Create a new sample of the model parameters using the multivariate normal distribution and the parameters from Step 2.
5. Use the sample parameters from Step 4 to calculate sample means and standard deviations for all incremental cells in the entire square. [Note: Multiplying these mean values times the ultimate claim count by period would result in a point estimate.]

6. Add process variance to the values from Step 5 by sampling each cell from the normal distribution using the mean and standard deviation values by cell (which will change the point estimate to a possible outcome).

7. Calculate the sample paid losses by multiplying each incremental cell times the ultimate claim count by period. Use just the future incremental paid losses to derive an estimate of the unpaid losses by period.

8. Repeat the random parameter sampling, sample square creation, random process variance and resulting unpaid calculations in Steps 4 through 7, *X* times.

The result from the *X* simulations is an estimate of a distribution of possible outcomes for the future unpaid losses. From this we can calculate the mean, standard deviation, percentiles, etc. It is important to note that this algorithm is COMPLETELY different than all other models described above. This model is only available for paid loss data.

Appendix A has a numerical example of each of the four variations (Berquist Sherman, Cape Cod, Chain Ladder and Hoerl Curve) of the Hayne MLE Incremental Severity model.

[THIS PAGE INTENTIONALLY LEFT BLANK]

4. Setting up a New Model

OVERVIEW

The basic analysis process looks like this:

1. Open the Arius software and use File | New to create a new project file.

2. Provide basic information about the data you will be entering and working with in the Project Settings dialog:

   - Use the Data Structure tab to change the size and structure of the triangles you will enter into the model;
   - Use the General tab to enter project information and notes; and
   - Use the Segments tab to add all of the lines of business you will work with in this model; each line of business or reserving segment will have its own data, models and assumptions.
   - Click **OK** on the bottom of the Project Settings dialog to finish creating a new file.

Note:
The Company/Business Unit entry will be used in all of your table & graph headers.

Note:
The Abbreviation is used in the system navigation, but the Description will be used in all of your table & graph headers.

3. Use File | Save As to save your file.

4. From the **Home** ribbon, click on the Model Options icon in the Stochastic area to open the Model Options dialog.

   - Use the Options tab to change any of the Global Options used with all stochastic models or either of the ODP Bootstrap Options which are only used with the ODP Bootstrap models for all segments;
   - If you select the "*Yes, Term*" Enable Discount Rate option, then use the Term Discount tab to either manually enter, or select from a lookup table, a discount rate yield curve; and
   - Use the Default Model Selection tab to select the models you expect to use for every segment.
   - Click **OK** on the bottom of the Model Options dialog to save your changes.

5. Below the **Home** ribbon, use the Segment drop-down list to select one of the lines of business.

6. In the Data | Inputs | All Inputs area of the **Navigation Pane**:

   - Enter Paid Loss and/or Incurred Loss data.
   - Enter Earned Premiums, Ultimate Premiums and/or Exposure data.
   - Enter Closed Claims and/or Reported Claims data.

Note:
Remember that throughout the system, data entry areas are white and areas that are calculated or which do not require user input are signified by a tan background.

7. In the Stochastic | ODP Bootstrap | Model Assumptions area of the **Navigation Pane**:

   - Select your General Model Options for this line of business.
   - Enter Bornhuetter-Ferguson and/or Cape Cod assumptions.

8. From the **Home** ribbon, click on the Run Diagnostics icon and select the Run Diagnostics for *SegmentAbbr* option to fill the exhibits and graphs with diagnostics.

9. In the Stochastic | ODP Bootstrap | Paid Loss | Diagnostics and/or the Stochastic | ODP Bootstrap | Incurred Loss | Diagnostics areas of the **Navigation Pane**, respectively:

- Review the patterns in the RESIDUAL GRAPHS window and adjust for heteroscedasticity as necessary, using the graphs in the RESIDUAL RELATIVITY window and/or using the SUGGEST HETERO GROUPS icon and select the SUGGEST HETERO GROUPS FOR *SEGMENTABBR* option from the **HOME** ribbon, if desired.
- If using the icon to find suggested hetero groups, you will need to enter values in, or copy and paste into, the Group Number row in the HETEROSCEDASTICITY table. Alternatively, you can click on the **Select Hetero Groups Graphically** button in the RESIDUAL GRAPHS window to select the groups.
- Review the Normality (Q-Q) Plot and Box-Whisker Plot in the NORMALITY window to determine if you need to exclude any outliers.
- To remove an outlier, you can either click on the appropriate dot in the RESIDUAL GRAPHS window (the dot will turn red once selected as an outlier) or you can identify the correct cell with a one ("1") in the OUTLIERS table.
- After selecting (or changing) hetero groups and/or outliers, you will need to use the RUN DIAGNOSTICS icon again to recalculate all of the diagnostic statistics.
- Use the TAIL FACTOR tool to enter tail factor assumptions.

10. Run the simulations for this segment using RUN SIMULATIONS icon and selecting the RUN SIMULATIONS FOR *SEGMENTABBR* option from the **HOME** ribbon.
11. In the STOCHASTIC | ODP BOOTSTRAP | PAID LOSS | *MODEL NAME* and/or the STOCHASTIC | ODP BOOTSTRAP | INCURRED LOSS | *MODEL NAME* areas of the **NAVIGATION PANE**, respectively:
    - Review the simulation results in all the tables and graphs for each model; and
    - Interactively, adjust model options and re-run the diagnostics and/or simulations until you are satisfied with the model fit and simulation results for each model.
12. Repeat steps #7 to 11 above for the Mack Bootstrap and Hayne MLE models, as desired. To activate these models for each segment, you must have selected them as part of your Default Model Selection options in step #4 above, or you can use the CHOOSE MODELS icon from the **HOME** ribbon to customize which models you use for each segment.
13. In the STOCHASTIC | ODP BOOTSTRAP | ODP SUMMARY | ASSUMPTIONS area of the **NAVIGATION PANE**, enter weights by accident period for each model in the MODEL WEIGHTS table and simulate again to get the initial "best estimate."
14. In the STOCHASTIC | ODP BOOTSTRAP | ODP SUMMARY | SUMMARY RESULTS area of the **Navigation Pane**:
    - Review the simulation results in all the tables and graphs for "best estimate" of the distribution;
    - Optionally, change the weights entered by accident period in step #13 above, re-simulate the "best estimate", and review the simulation results again; and
    - Compare stochastic and deterministic "best estimates" and, optionally, enter selected total unpaid in the last column in the DETERMINISTIC CALCULATIONS table, click to check the "Use Selected Unpaid as Mean" checkbox, and re-simulate to shift results to match selected reserves.
15. Repeat steps #5 to 14 above for each line of business in the model.
16. From the **HOME** ribbon, click on the RUN DIAGNOSTICS icon and select the RUN DIAGNOSTICS FOR ALL SEGMENTS & CORRELATION option which will not only update all the diagnostic results for all the

segments, but it will also generate the correlation matrix tables on the ODP BOOTSTRAP AGGREGATION area of the **NAVIGATION PANE**.

17. In the ODP BOOTSTRAP AGGREGATION | ASSUMPTIONS | CORRELATION area of the **Navigation Pane**:

    - Review the various correlation matrices that are calculated for you in the Calculated table.
    - Use the User Selected object to enter correlation coefficients for each pair of segments, or use one of the **Quick Fill** buttons to either fill the correlation matrix with the same value for each pair or to fill the correlation matrix with values from one of the calculated tables.
    - You may also change the Degrees of Freedom for the T-distribution to be used in the correlation process in the USER SELECTED window.

18. From the **HOME** ribbon, click on the RUN SIMULATIONS icon and select the RUN SIMULATIONS FOR ALL SEGMENTS & AGGREGATION option to run simulations for all segments and generate a final overall distribution taking into account the effect of correlation between the segments.

STEP 1: CREATE A NEW MODEL FILE

1. In Windows, go to **Start** | **All Programs** | **Arius | Arius** which will open the Arius software.
2. In Arius, use **File | New** to create a new project file.

STEP 2: DEFINE THE SIZE AND STRUCTURE OF YOUR DATA SETS

1. When using **File | New** as noted in Step 1, the Project Settings dialog will automatically open. You can also open the Project Settings dialog using the **Project Settings** icon on the **Home** ribbon.
2. Use the Data Structure tab (as illustrated in Image 4-1).

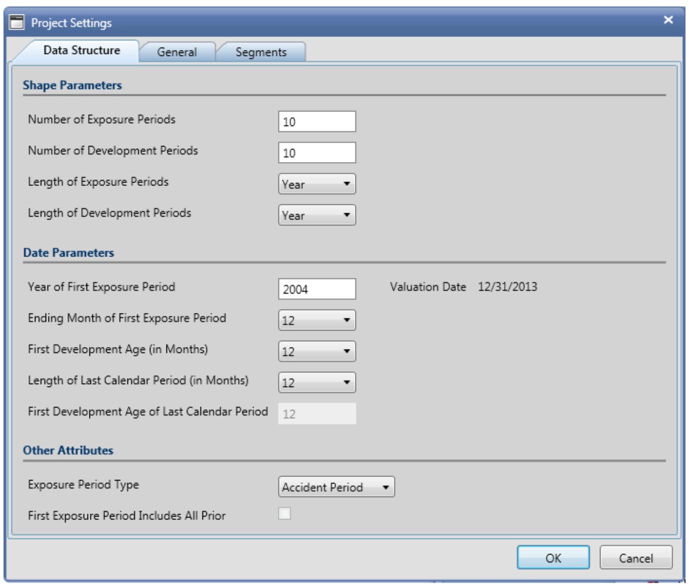


Image 4-1:
Data Structure tab in the Project Settings Dialog Box

Here you define the basic attributes of all triangles of data in the project file. If your project is for many lines of business with triangles of different sizes, then you should fill in the specifications for the largest triangle that you expect to work with in this file. You will specify:

- **Number of Exposure Periods** – the maximum number rows in your triangles
- **Number of Development Periods** – the maximum number of columns in your triangles
- **Length of Exposure Periods** – whether the rows are annual, semi-annual, quarterly, or monthly.
- **Length of Development Periods** – whether the columns are annual, semi-annual, quarterly, or monthly.
- **Year of First Exposure Period** – the year associated with the first row of data in your largest triangle
- **Ending Month of First Exposure Period** – the month that represents the end of the first row's data (June = 6, December = 12, etc.)

Note:
For the Stochastic models the length of the Exposure Periods and Development Periods must be the same (i.e., symmetrical). For the Deterministic methods they can be different (i.e., asymmetrical).

- **First Development Age (in Months)** – how many months of development there are in the first column (in most annual x annual triangles this will be 12, but this first column *can* be less than the other columns, as in 3/15/27/39… triangles)
- **Length of Last Calendar Period (in Months)** – how many months are in your most recent diagonal (for example in annual x annual triangles evaluated after only 6 months, this would be reflected by a "6")
- **Exposure Period Type** –This can be either Accident or Policy period. Your selection affects the labels on your triangles, and it also affects the model calculations in some situations (e.g., for uneven exposures or "stub" period exposures).[23]
- **First Exposure Period Includes All Prior** – If checked, this signifies that the first row is different from the rest and will be ignored in the simulation models.

As a check, the dialog calculates the **First Development Age of the Last Calendar Period** and the data's **Valuation Date** based on your input. If this date is not as expected, review your other selections on this dialog before pressing **OK**.

STEP 3: ADD GENERAL INFORMATION ABOUT YOUR PROJECT

1. When using **FILE | NEW** as noted in Step 1, the PROJECT SETTINGS dialog will automatically open. You can also open the PROJECT SETTINGS dialog using the **PROJECT SETTINGS** icon on the **HOME** ribbon.
2. Use the GENERAL tab (as illustrated in Image 4-2).

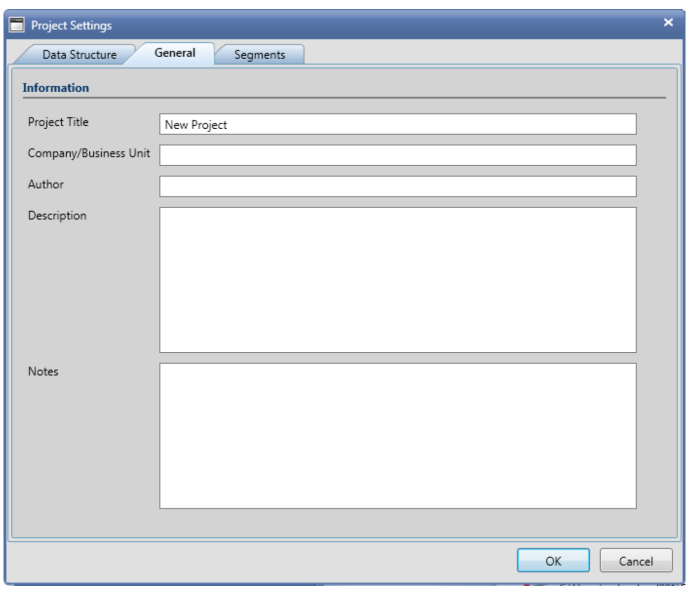


Image 4-2: GENERAL tab in the Project Settings Dialog Box

Here you include general information about your project. You will specify:

23 From this point forward the manual will generally refer to either accident year or accident period, but policy year or policy period can usually be considered interchangeable.

- **Project Title** – this can be the same as the filename, but it can be different. Whatever you include in this field will be part of filenames when exporting results to help you identify the files.
- **Company/Business Unit** – this should be a full company or business unit name. Whatever you include in this field will be part of the headers for all tables and graphs, including when exported for creating reports.
- **Author** – this is the name of the person(s) primarily responsible for this project.
- **Description** – this field can be any size. It is available to allow you to describe this project in detail for a peer reviewer or anyone else that might use the project file.
- **Notes** – this field can be any size. It is available to allow you to save notes on your analysis to be shared with a peer reviewer or anyone else that might use the project file.
- Press **OK** to save the changes in the GENERAL tab.

STEP 4: IDENTIFY THE LINES OF BUSINESS TO BE ANALYZED IN THIS PROJECT

Decide how many LOBs you want in your project. Typically, these will be related groups of data, lines that are managed by the same management unit, or data that is otherwise reviewed for reserving purposes. You should consider three things when deciding which and how many lines to include in a project file:

- All input areas must have the same “shape” (e.g., if one triangle is developing 12-24-36 with a 6 month last diagonal by accident year, all triangles in the same project must have this shape) and evaluation date.
- It may be better to keep the number at a “manageable” level (say, for a 10x10 triangle, approximately 10-15 LOBs) as the model could run more slowly with more LOBs. This could also depend on how you split up the workload.
- On the other hand, you can correlate and aggregate different sub-groups within the same project file (e.g., aggregating different sub-lines or hazard types within a line of business, geographic areas, management or business units, etc. into subtotals and/or a corporate total) – See Section 7. Therefore, you can fit all the LOBs into one project file.

1. When using **FILE | NEW** as noted in Step 1, the PROJECT SETTINGS dialog will automatically open. You can also open the PROJECT SETTINGS dialog using the **PROJECT SETTINGS** icon on the **HOME** ribbon.
2. Use the SEGMENTS tab (as illustrated in Image 4-3).

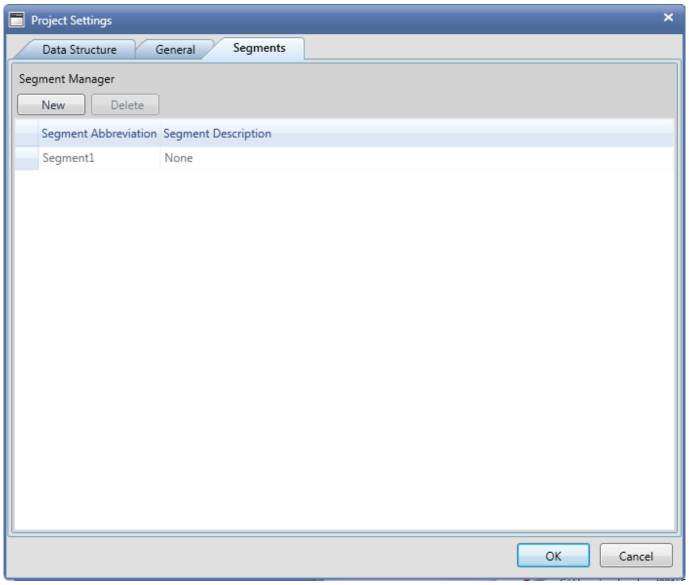


Image 4-3:
SEGMENTS tab in the Project Settings Dialog Box

Here you include an abbreviation and full name for each segment in your project. You will specify:

- **Segment Abbreviation** – this is normally a shorter name or code used to identify a segment (e.g., WC-xCA). The abbreviation will be used in all navigation areas for easy identification of a segment.
- **Segment Description** – this is normally a longer description of the segment (e.g., Workers Compensation - Countrywide excluding CA). Whatever you include in this field will be part of the headers for all tables and graphs, including when exported for creating reports.

3. Use the **New** button (shown in Image 4-3) to create additional lines of business or reserving segments you want to work with in the file. For each segment added to the file you will need to include the abbreviation and description as noted just above.
4. Select an existing segment and use the **Delete** button (shown in Image 4-3) to remove any lines of business or reserving segments you no longer want in the file.
5. Press **OK** to save the changes in the SEGMENTS tab.

STEP 5: REVIEW THE STOCHASTIC MODEL OPTIONS

From the **HOME** ribbon you can click on the Stochastic MODEL OPTIONS icon to review and control settings that affect all the segments in the model, as shown in Image 4-4.

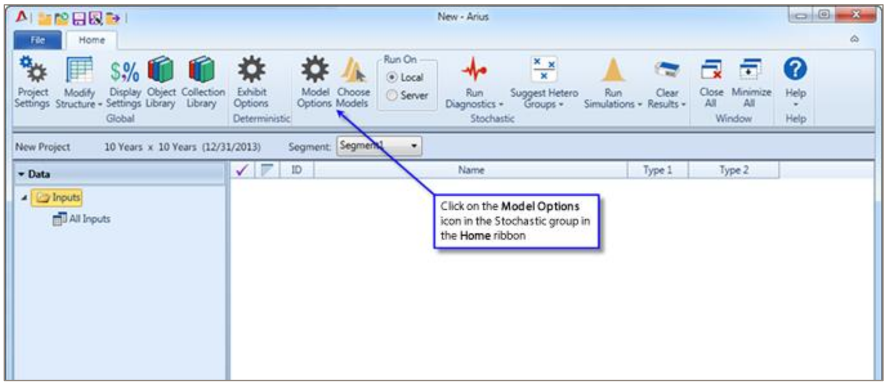


Image 4-4:
Model Options icon in the **Home** ribbon

This will open the Model Options dialog box which will be focused on the Options tab, as shown in Image 4-5.

The **Global Options** area in the Options tab of the Model Options dialog (illustrated in Image 4-5) allows you to change options that apply to all segments when running the models or tasks that will be used later. It is good, however, to have an understanding of the options here, so you will know where to find them and how they are used.

1. **Iterations** – The default is 10,000, but you can reduce this to a low of 1 or increase it to a high of 50,000. This will, of course, affect the speed of the model (e.g., you could use a lower number while testing models), but you should set the number of iterations high enough to get a relatively stable result from one run to the next.

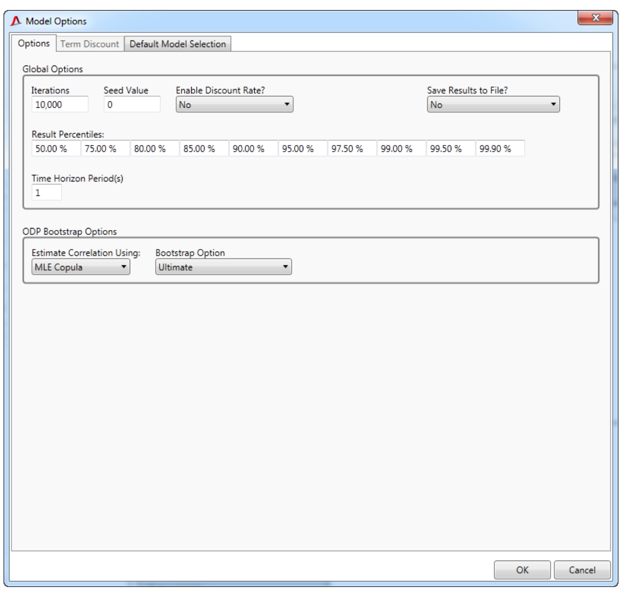


Image 4-5:
Options tab in the Model Options dialog box

Note:
A Seed Value other than zero will only guarantee the results will stay the same on each model run if all parameters and other LOB specific model settings are the same.

2. **Seed Value** – The default is zero, which means that a new set of random values will be simulated each time the model is run. If you would like to replicate the results exactly each time (e.g., to test

the impact of a specific model parameter change) you should enter a positive integer in this field between 1 and 2,146,483,648 (= $2^{31} - 1$).

3. **Result Percentiles** – In addition to the Mean, Standard Error, Coefficient of Variation, Minimum and Maximum values, you can specify ten percentiles that will be included in the output tables when the simulations are run.

4. **Time Horizon Period(s)** – When using the ODP Bootstrap or Mack Bootstrap models, this is the number of periods in the time horizon.

5. **Enable Discount Rate** – The default is **No**, but the other options can be used to generate discounted simulation results. Selecting the **Yes, Annual** option will open the **Discount Rate per Annum** box so you can enter a single discount rate that will be used for all discount factors, as illustrated in Image 4-6. The discount rate is calculated as follows:

   Discounted incremental = incremental x (1 + discount rate / 12) ^ months in exposure period

Note:
By using a Seed Value other than zero, you can specify different percentiles each time you run the model and obtain more than ten percentile outputs or turn discounting on and off to assure that the difference is only due to the discount factors.

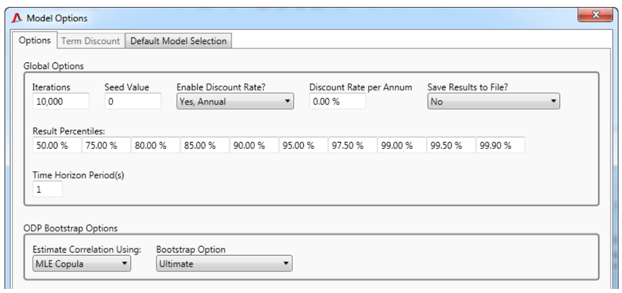


Image 4-6:
OPTIONS tab in the MODEL OPTIONS dialog box, with **Enable Discount Rate** option set to **Yes, Annual**

Selecting the **Yes, Term** option hides the **Discount Rate per Annum** box, but enables the TERM DISCOUNT tab in the MODEL OPTIONS dialog, as illustrated in Image 4-7. With the **Yes, Term** option selected, you can now select the TERM DISCOUNT tab in the MODEL OPTIONS dialog and use one of two options for using a yield curve for the discount factors instead of a single discount rate. The first option is to select the **Enter/Edit Rate Manually** radio button and then enter a **Month** and **Rate** for as many discount values as you need in the **User entered values** area, as illustrated in Image 4-8. When using this option you can change the **Manual Increment** option to either **Annual**, **Semi-Annual**, **Quarterly** or **Monthly**, in which case the values in the **Month** column will increment automatically by 12, 6, 3 or 1, respectively, as you enter a new **Rate**. This is only intended to ease the data entry of a discount yield curve, so you can modify any entered **Month** or **Rate**.

Note:
When entering rates manually, keep in mind that the values in the **Month** column must be increasing.

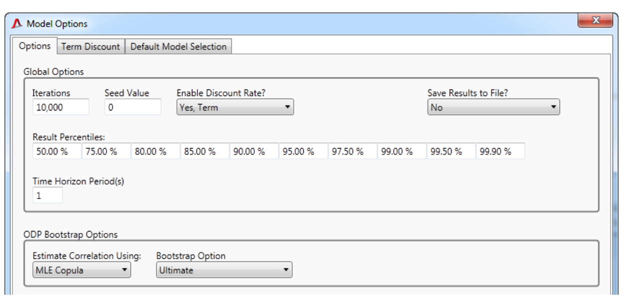


Image 4-7:
OPTIONS tab in the MODEL OPTIONS dialog box, with **Enable Discount Rate** option set to **Yes, Term**

When the **User entered values** are interpolated in order to calculate discount factors, the **Month** value will be interpreted as final month that the prior rate will be used. Thus, the first **Month** will always be zero since there are no rates prior to the first **Period**. In the example in Image 4-8, the second **Month** of 12 means that the first **Rate** will be used for 12 months and then the second **Rate** will start being used until the **Month** shown in the next **Period**. For the last **Period** entered, the final **Rate** will be used indefinitely (e.g., in the example in Image 4-8 the second rate of 5.00% will be used after 12 months for as many months needed to discount all cash flows).

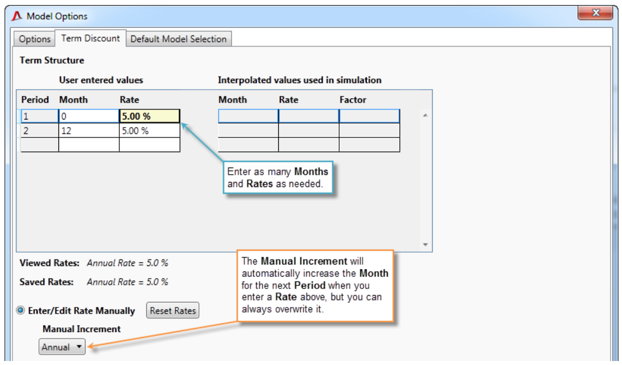


Image 4-8:
Term Discount tab in the Model Options dialog box, with **Enter/Edit Rate Manually** selected

The second option is to select the **Get Rates for File** radio button, which will allow you to import rates from saved rate files, as illustrated in Image 4-9. To import rates, first use the **Select Table** drop down list to select a file, then use the **Select Rates** drop down list to select a specific rate curve saved in that file. Once the rates you want are selected, click the **Get Rates** button to import those rates into the **User entered values** table.

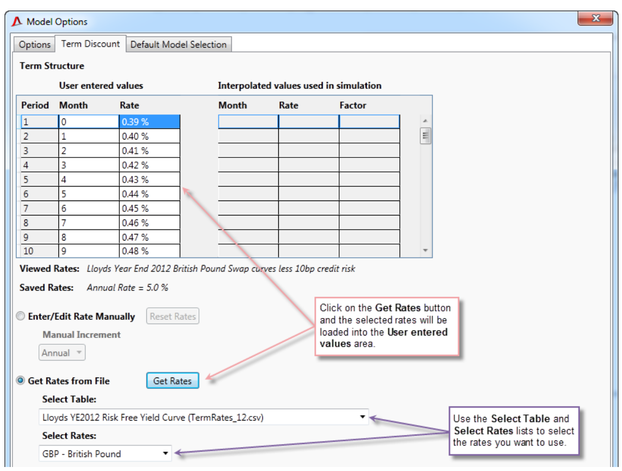


Image 4-9:
TERM DISCOUNT tab in the MODEL OPTIONS dialog box, with **Get Rates from File** selected

Note:
The **Interpolated values used in simulation** table will be populated when you run simulations. You can reopen the MODEL OPTIONS dialog after simulation to see the discount factors.

After you import rates from a file, you can always switch to the **Enter/Edit Rate Manually** option and edit any values in the **User entered values** table.

6. **Save Results to File** – The default is **No**, but the other options might be useful if you wish to calculate output statistics that are not already available in the standard output or combine the Arius results with non-Arius simulations. Selecting **Exposure Totals** means that the accident year unpaid, accident year ultimate losses and calendar year cash flow results for year in total and for all years combined will be saved for each iteration in the simulation, as illustrated in Image 4-10.

The location of the saved data file will be found in the *C:\Users\username\Documents\Milliman\Arius\Sim_Results* directory, where the *username* is your Windows user name.

Row							
1		Unpaids					
2	Iteration	2009	2010	2011	2012	2013	Total
3	1	0	56	127	501	656	1,340
4	2	0	87	156	287	429	958
5	3	0	52	177	223	839	1,291
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
10,002	10,000	0	61	172	353	768	1,353
10,003		Ultimate Losses					
10,004	Iteration	2009	2010	2011	2012	2013	Total
10,005	1	1,288	720	970	1,360	994	5,333
10,006	2	1,133	927	669	974	728	4,431
10,007	3	1,447	913	861	816	1,091	5,129
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
20,004	10,000	1,162	598	836	1,017	1,139	4,752
20,005		Premiums					
20,006	Iteration	2009	2010	2011	2012	2013	Total
20,007	N	2,000	2,000	2,000	2,000	2,000	10,000
20,008		CashFlows					
20,009	Iteration	2014	2015	2016	2017	Totals	
20,010	1	627	388	265	61	1,340	
20,011	2	521	236	110	92	958	
20,012	3	744	287	156	105	1,291	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	
30,009	10,000	762	270	220	101	1,353	

Image 4-10:
Exposure Totals Data. The row column shown in ***bold italic*** was added for illustration purposes and is not part of the saved data file.

Selecting **All Incrementals, by Iteration** means that the results for each incremental cell (both historical and future) for each iteration will be saved, as illustrated in Image 4-11.

Row							
1	Iteration	Accident Year	12	24	36	48	60
2	1	2009	466	463	178	103	78
3	1	2010	206	262	118	79	56
4	1	2011	284	405	155	83	44
5	1	2012	380	479	216	118	167
6	1	2013	338	272	226	98	61
7	2	2009	279	451	223	113	68
8	2	2010	293	309	155	83	87
9	2	2011	220	214	79	65	91
10	2	2012	257	430	152	92	43
11	2	2013	299	217	53	66	92
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
49,997	10,000	2009	331	397	242	85	107
49,998	10,000	2010	171	239	108	20	61
49,999	10,000	2011	191	306	167	135	36
50,000	10,000	2012	233	431	154	96	102
50,001	10,000	2013	371	411	138	117	101

Image 4-11:
All Incremental Data. The row column shown in ***bold italic*** and lines were added for illustration purposes and are not part of the saved data file.

Selecting **All Incrementals, by Year** means that the results for each incremental cell (both historical and future) for each iteration will be saved, as illustrated in Image 4-12.

Row							
1	Iteration	Accident Year	12	24	36	48	60
2	1	2009	466	463	178	103	78
3	2	2009	279	451	223	113	68
4	3	2009	466	511	279	74	117
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
10,001	10000	2009	331	397	242	85	107
10,002	1	2010	206	262	118	79	56
10,003	2	2010	293	309	155	83	87
10,004	3	2010	210	382	186	83	52
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
20,001	10000	2010	171	239	108	20	61
20,002	1	2011	284	405	155	83	44
20,003	2	2011	220	214	79	65	91
20,004	3	2011	222	326	135	66	111
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
50,001	10,000	2013	371	411	138	117	101

Image 4-12:
All Incrementals Data. The row column shown in ***bold italic*** and lines were added for illustration purposes and are not part of the saved data file.

Selecting **Correlation Ranks** means that the ranks used to correlate the segment results into the aggregate results will be saved. This is illustrated in Image 4-13. In this example, we have 10,000 iterations and 3 lines of business. To use these ranks, you would sort your 10,000 total unpaid results according to these rank values. For example, the aggregate unpaid for the first iteration would be the 1,117th largest total unpaid result for the first line of business, plus the 3,561st largest total unpaid result for the second line of business, plus the 8,601st largest unpaid result for the third line of business. The model does this automatically for you when producing aggregate results.

Row				
1	Iterations	LOB001	LOB002	LOB003
2	1	1,117	3,561	8,601
3	2	387	1,859	715
4	3	5,221	2,899	2,226
⋮	⋮	⋮	⋮	⋮
10,001	10,000	2,915	950	2,750

Image 4-13:
Correlation Ranks Data. The row column shown in ***bold italic*** was added for illustration purposes and is not part of the saved data file.

The **ODP Bootstrap Options** area in the OPTIONS tab of the MODEL OPTIONS dialog (illustrated in Image 4-5) allows you to change options that apply only to the ODP Bootstrap models, but to all segments when running the models or tasks that will be used later.

1. **Estimate Correlation Using** – The default is **MLE Copula**, which uses a maximum likelihood estimation copula to solve for all correlations at once. The **Pairwise** option calculates the correlation between each pair of LOBs.

   The **MLE Copula** is a more mathematically rigorous methodology. However, since it solves for all correlations at once, it can only calculate the correlations between the data points that exist in the triangles in all the segments. If you have, say, some lines of business that you started to write in the last 5 years, and some that were put in run-off 5 years ago, they will have no common data points in their triangles, and this correlation calculation will fail and provide an all-zero matrix.

   In these types of situations, the **Pairwise** option is recommended. This measures correlations between each pair of segments individually. Those pairs of LOBs that do not have any overlapping data points will still show a correlation of zero, but correlations will still be calculated for the remaining pairs of segments.

2. **Bootstrap Option** – The default is **Ultimate**, which activates the Basic models described in Section 3 and Appendix A. Selecting the **Time Horizon – ODP Process** option will activate the ODP Process

models described in Section 3 and Appendix A. Selecting the **Time Horizon – ODP Residual** option will activate the ODP Residual models described in Section 3 and Appendix A.

The DEFAULT MODEL SELECTION tab (illustrated in Image 4-14) allows you to define which models will automatically be selected when you create a new segment or when you have “Use Default Models” checked in the segment CHOOSE MODELS dialog (see Image 4-16).

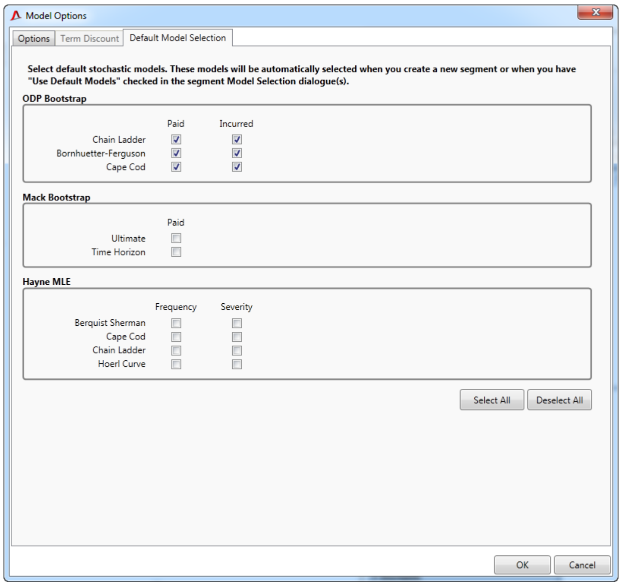


Image 4-14:
Default Model Selection tab in the Model Options dialog box

From the **HOME** ribbon you can click on the Stochastic CHOOSE MODELS icon to change the models used in each segment, as illustrated in Images 4-15 and 4-16.

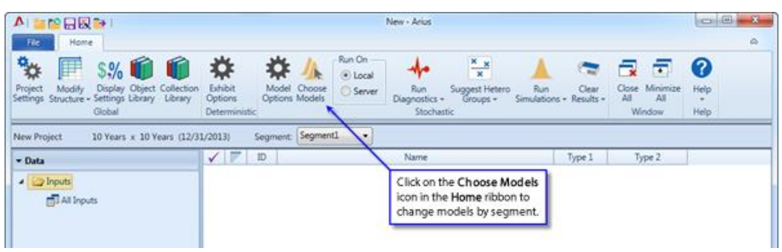
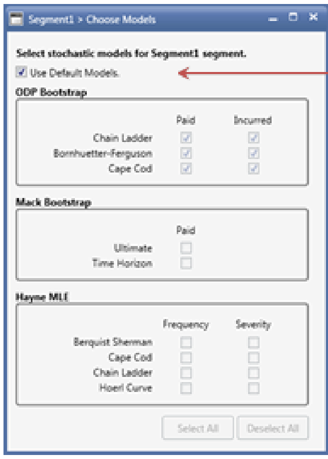


Image 4-15:
CHOOSE MODELS icon in the **HOME** ribbon

After the CHOOSE MODELS dialog is open, you can leave the **Use Default Models** option checked (or recheck the option) to use the models selected in the DEFAULT MODEL SELECTION tab of the MODEL OPTIONS dialog OR you can uncheck the **Use Default Models** option and customize the models used for a specific segment, as illustrated in Image 4-16.



Check **Use Default Models** option to use the models from the DEFAULT MODEL SELECTION tab of the MODEL OPTIONS dialog OR uncheck **Use Default Models** option to customize the model(s) selected for a Segment.

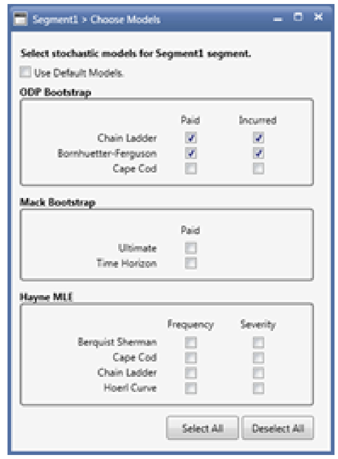


Image 4-16:
Using the CHOOSE MODELS dialog box

5. Using the ODP Bootstrap Models

Even though the Arius system has numerous options to help you obtain the best model possible for your data, you can obtain valuable diagnostic information and even initial distribution estimates for a line of business with only a few steps, which can be summarized as:

- enter the data to be modeled,
- run the model diagnostics to populate the necessary statistics and fields, and
- run the simulation to estimate future results (i.e., use the default model settings).

Of course, the diagnostics and model results can be used to evaluate and improve how your model fits your data. Understanding the purpose and use of the diagnostic tools requires some prior statistical knowledge so we direct the interested reader to Appendix B, which provides a general overview of the diagnostic process. Therefore, this section assumes prior knowledge of statistics, and starts with the basics of running a model and builds on that foundation by exploring all of the different models, model options, diagnostics, and model output.

REQUIRED DATA: PAID MODEL

Inputs for the paid model are relatively simple. You can start with nothing more than a triangle of paid loss data, but if:

IN ADDITION TO PAID LOSS DATA, IF YOU PROVIDE:	THE SYSTEM CAN:
▪ a vector of earned premium data	▪ provide loss ratios by accident period at various percentiles
▪ a triangle of incurred or reported loss data	▪ reconcile Ultimate Losses using Paid, Case Reserves, and IBNR
▪ a vector of ultimate exposure data	▪ simulate based on exposure-adjusted losses rather than only the raw data

Note:
If you have a partial last exposure period, then you should enter the earned premium in the appropriate column, but the ultimate premium and ultimate exposure are for the **full period**. For example, if you have an annual triangle but a 6 month last diagonal, then you should enter the premiums earned for the first 6 months in the earned premium column and the fully annualized premium and/or exposure in the ultimate premium and ultimate exposure columns, respectively. For more details see Section 9.

There are certain limitations that are imposed on the data by the mathematics involved in the model. Specifically:

- The triangle shape must be symmetrical in terms of row and column periods – i.e., it must be annual x annual or quarter x quarter;
    - The system *will* work with triangles that contain a stub period (e.g., annual x annual with most recent diagonal evaluated at 6 months)
    - The system *will* work with triangles where the first development period is different from the rest (e.g., development columns of 6/18/30/42… or 3/15/27/39…)
    - The system *will not* work with truly asymmetrical triangles, such as annual accident periods x quarterly development.
- There must be at least 3 diagonals of data.
- Blank cells are acceptable anywhere in the triangle except on the most recent two diagonals, unless a whole row is blank (i.e., a triangle in run-off is OK)
- Individual negative age-to-age factors are acceptable, and the average for a column can be negative.

- Do not enter “0” values where the values are unknown. The model will treat cells with “0” values as information (that is, no losses occurred in this period) and blank cells as unknown.

REQUIRED DATA: INCURRED MODEL

Inputs for the incurred model are also relatively simple. You can start with nothing more than a triangle of paid loss data and a triangle of incurred loss data, but if:

IN ADDITION TO DATA TRIANGLES, IF YOU PROVIDE:	THE SYSTEM CAN:
▪ a vector of earned premium data	▪ provide loss ratios by accident period at various percentiles
▪ a vector of ultimate exposure data	▪ simulate based on exposure-adjusted losses rather than only the raw data

All of the limitations that are imposed on the data by the mathematics involved in the model for the paid data also apply to the incurred data. In addition:

- the paid and incurred triangles must be identical in terms of underlying shape and size (i.e., the triangle properties apply to both paid and incurred triangles);
- the paid and incurred triangles must be virtually identical in terms of data:
    - The system *will* usually work if some individual cells are missing (i.e., blank) in one triangle but not the other;
    - The system *will not* work with an entire row missing in one triangle and not the other.

REQUIRED DATA: BORNHUETTER-FERGUSON MODEL

Inputs for the Bornhuetter-Ferguson model are also relatively simple. You can start with nothing more than a triangle of paid loss data and/or a triangle of incurred loss data as well as the a priori ultimate loss assumption by year, but if:

IN ADDITION TO BASIC DATA IF YOU PROVIDE:	THE SYSTEM CAN:
▪ a vector of ultimate premium data	▪ provide loss ratios as the a priori ultimate assumption
▪ a vector of ultimate exposure data	▪ simulate based on exposure-adjusted losses rather than only the raw data and can use pure premiums as the a priori ultimate assumption.
▪ a vector of coefficients of variation for the a priori assumptions	▪ include uncertainty in the a priori assumption by simulating a different assumption for each iteration

All of the limitations that are imposed on the data by the mathematics involved in the model for the paid and incurred data, respectively, also apply to the Bornhuetter-Ferguson model.

REQUIRED DATA: CAPE COD MODEL

Inputs for the Cape Cod model are also relatively simple. You can start with nothing more than a triangle of paid loss data and/or a triangle of incurred loss data as well as the default Cape Cod assumptions by year, but if:

IN ADDITION TO BASIC DATA IF YOU PROVIDE:	THE SYSTEM CAN:
▪ a vector of ultimate premium data	▪ provide loss ratios as the a priori ultimate assumption
▪ a vector of ultimate exposure data	▪ simulate based on exposure-adjusted losses rather than only the raw data and can use exposures as the Cape Cod calculation basis
▪ vectors of premium index factors, loss trends and weights for the Cape Cod assumptions	▪ use specific assumptions for the Cape Cod model instead of the default assumptions

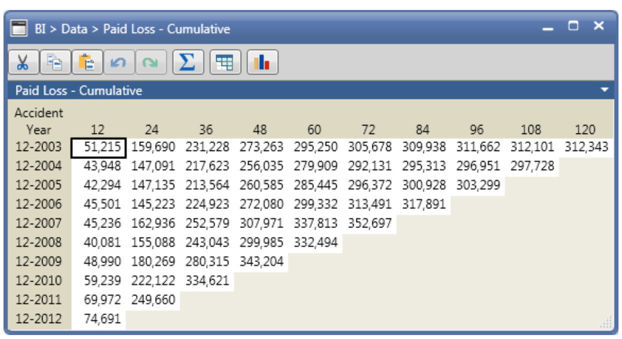
All of the limitations that are imposed on the data by the mathematics involved in the model for the paid and incurred data, respectively, also apply to the Cape Cod model.

Note:
Throughout the system, data entry areas are white and areas that are calculated or which do not require user input are signified by a tan background.

STEP 1: ENTER BASIC MODEL DATA

To get started, select one of your segments using the **Segment** drop down box below the **HOME** ribbon. In the **Navigation Pane**, select the DATA | INPUTS | ALL INPUTS collection. Notice that the first three tables in the collection, **Paid Loss**, **Case Loss Reserves** and **Incurred Loss**, are white; these are the data entry tables. You can fill in any two of these tables and the third will change to tan, which means it will be filled automatically and that you cannot enter data here any longer.

1. Enter data for the **Paid Loss** triangle (as illustrated in Image 5-1) and the **Incurred Loss** triangle, if you have that available. You can either type in data or paste it in from another source.

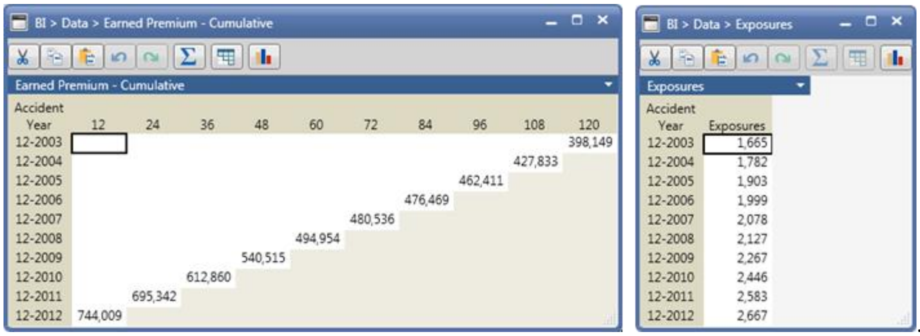


Accident Year	12	24	36	48	60	72	84	96	108	120
12-2003	51,215	159,690	231,228	273,263	295,250	305,678	309,938	311,662	312,101	312,343
12-2004	43,948	147,091	217,623	256,035	279,909	292,131	295,313	296,951	297,728	
12-2005	42,294	147,135	213,564	260,585	285,445	296,372	300,928	303,299		
12-2006	45,501	145,223	224,923	272,080	299,332	313,491	317,891			
12-2007	45,236	162,936	252,579	307,971	337,813	352,697				
12-2008	40,081	155,088	243,043	299,985	332,494					
12-2009	48,990	180,269	280,315	343,204						
12-2010	59,239	222,122	334,621							
12-2011	69,972	249,660								
12-2012	74,691									

Image 5-1:
Paid Loss Data Triangle

Note:
You can use the icon to switch between cumulative and incremental or the icon to switch between accident and calendar views, or both, prior to bringing in the data.

2. Also from the ALL INPUTS collection, you can enter **Earned Premium** and **Exposure** data, if you have that available (as illustrated in Image 5-2). Having this additional data allows the model to provide more information; this is especially true of Premium data, which allows the projection of ultimate loss ratios.



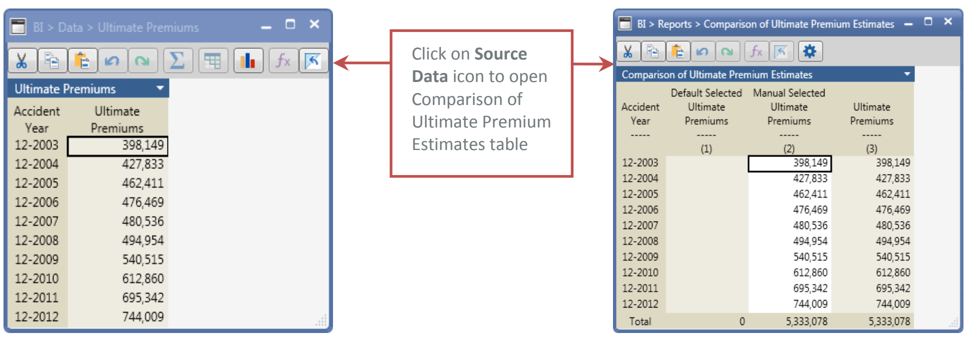
Accident Year	12	24	36	48	60	72	84	96	108	120
12-2003										398,149
12-2004									427,833	
12-2005								462,411		
12-2006							476,469			
12-2007						480,536				
12-2008					494,954					
12-2009				540,515						
12-2010			612,860							
12-2011		695,342								
12-2012	744,009									

Accident Year	Exposures
12-2003	1,665
12-2004	1,782
12-2005	1,903
12-2006	1,999
12-2007	2,078
12-2008	2,127
12-2009	2,267
12-2010	2,446
12-2011	2,583
12-2012	2,667

Image 5-2: Earned Premium and Exposure tables

Note: The earned premiums are entered in a triangle so that they can be developed in the Deterministic portion of the system.

3. In order to enter the **Ultimate Premium** data (again from the ALL INPUTS collection), you must open the table and click on the Source Data icon in order to get to the Deterministic table used to estimate **Ultimate Premium**. This is illustrated in Image 5-3.



Accident Year	Ultimate Premiums
12-2003	398,149
12-2004	427,833
12-2005	462,411
12-2006	476,469
12-2007	480,536
12-2008	494,954
12-2009	540,515
12-2010	612,860
12-2011	695,342
12-2012	744,009

Accident Year	Default Selected Ultimate Premiums	Manual Selected Ultimate Premiums	Ultimate Premiums
-----	-----	-----	-----
	(1)	(2)	(3)
12-2003		398,149	398,149
12-2004		427,833	427,833
12-2005		462,411	462,411
12-2006		476,469	476,469
12-2007		480,536	480,536
12-2008		494,954	494,954
12-2009		540,515	540,515
12-2010		612,860	612,860
12-2011		695,342	695,342
12-2012		744,009	744,009
Total	0	5,333,078	5,333,078

Image 5-3: Ultimate Premiums and Comparison of Ultimate Premiums Estimates tables

STEP 2: REVIEW / ENTER THE MODEL ASSUMPTIONS

Most of the **Model Assumptions** and **Model Options** are equivalent for both paid and incurred data, but they are independently applied. As noted in Section 3, for the incurred chain ladder model both the paid and incurred models are run in parallel and then the paid simulation is adjusted to match the ultimate values by year for the incurred model so that we end up with an apples-to-apples comparison of total unpaid estimates. Thus, the **Model Assumptions** and **Model Options** must be selected for *both* the paid *and* incurred columns whenever applicable.

In the **Navigation Pane**, select the STOCHASTIC | ODP BOOTSTRAP | MODEL ASSUMPTIONS collection. The **General** window (shown in Image 5-4) includes model assumptions that will apply to all of the ODP Bootstrap models.

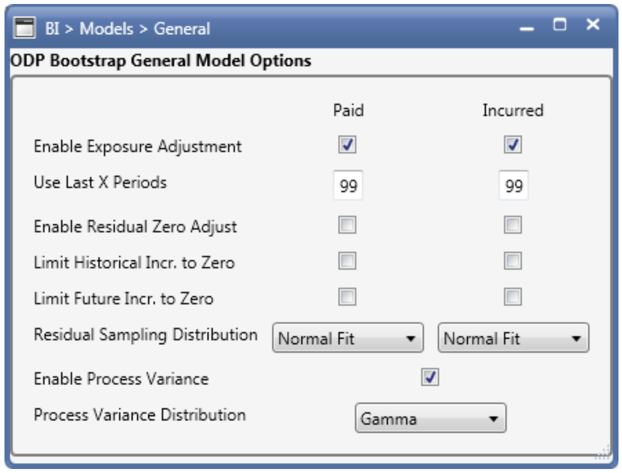


Image 5-4:
General Model Options window

Note:
Reducing the number of periods in the average age-to-age ratios will also reduce the pool of residuals for resampling. Also, when the last diagonal is not a full period, it will be grossed up and included in the weighted average age-to-age factors (See Section 9).

1. **Enable Exposure Adjustment** – If you check this option, the system divides each row in your data triangle by the corresponding row in the **Ultimate Exposures** vector and uses the “exposure-adjusted” data for all further calculations in the model. Values are then multiplied by the **Ultimate Exposures** again after all iteration calculations are complete, returning the modeled results to a “value” basis. This option can be useful when there is a changing exposure volume. By using exposure adjusted data in the model, a better fit could result and the simulation results will be adjusted for the relative exposures by period.

2. **Use Last X Periods** – The default for all periods (“**99**”) is typical because the Generalized Linear Model theory underlying the ODP Bootstrap model is used to derive a volume weighted all-year average for the age-to-age ratios. However, if you feel that perhaps the most recent X years of history are more representative of the payment activity you may expect in the future, you can adjust the model to take this into account by setting **Use Last X Periods** to **X**, which causes the model to use X-period average age-to-age ratios.

3. **Enable Zero Residual Adjustment** – All the residuals in your model should be independent and identically distributed. Theoretically, they should also sum to zero. If you want to, you can *force* the total, prior to any adjustment for heteroscedasticity, to be equal to zero by checking this option. If you do, the system subtracts from each non-zero residual the total of all the residuals divided by the number of non-zero residuals, so that the resulting total of all residuals is adjusted to zero. The model then uses these adjusted residuals.

   While it may be theoretically correct to check this option (set it to “Yes”**)**, the default is unchecked (set to “No”) so that the sum will provide you with information about your data before running the model, and so that you can see how the sum changes as you change model options (e.g., hetero factors) – by setting this to “No” it is another diagnostic tool. In addition, if there is skewness in the residuals (and therefore the underlying data), you may want that to flow through into your projections by leaving it set to “No”.

Note:
Changing any of the following options will change the core parts of the model and will require you to select RUN DIAGNOSTICS to update the residuals and other diagnostics: **Enable Exposure Adjustment**, **Use Last X Periods**, **Enable Zero Residual Adjustment**, **Heteroscedasticity Groups**, or **Outlier Triangle**.

4. **Limit Historical Incrementals to Zero** – The random nature of the simulation process can result in negative amounts in the incremental results. When this option is checked, the system automatically replaces any negative incremental values in the bootstrap sample triangles with zero. Negative incremental values are certainly acceptable in many situations, for example when modeling paid data that includes salvage amounts, or when modeling incurred data; in those cases, negatives are frequently expected, and they should be reflected in the simulated data. Occasionally, however, negative incremental values in the bootstrap sample triangle can also lead

to extreme age-to-age factors which, in turn, lead to unrealistic unpaid values for a few iterations. Limiting historical incremental values to zero in that case can be thought of as adding a constraint to the model which will effectively “adjust” these unrealistic iterations.

5. **Limit Future Incrementals to Zero** – When this option is checked, the system automatically replaces any negative incremental values in the lower right portion (after process variance) of the completed rectangle with zero. This provides a similar ability to effectively constrain unrealistic iterations, but it is a separate constraint since there are times when only the future incremental values need to be constrained and not the historical incremental values, and vice versa. For example, negative historical incremental values in the bootstrap sample data can be reasonable when case reserves for later development ages are expected to be redundant while negative future incremental values could be causing unrealistic results for a few iterations.

6. **Residual Sampling Distribution** – There are six options that will govern how the sample triangles are created.

    - **Residuals** – is the commonly known ODP Bootstrap model (illustrated in Appendix A) which samples the residuals with replacement to create the sample triangles. This is the default option.

    - **Normal Fit** – fits a normal distribution to the residuals and then simulates the residuals from the fitted normal distribution to create the sample triangles. You can choose this option if you believe the normal distribution is a good fit and don’t want the sampling process to be limited to the existing residuals. The parameters used in this option are shown below the Normal and Box Whisker Plots that are part of the DIAGNOSTICS collection.

    - **Normal PV** – The model uses the fitted triangle and assumes each incremental value is the mean and the incremental value times the scale parameter (adjusted by the hetero-factors) is the variance of a normal distribution. Essentially, the historical triangle incremental values are simulated in the same fashion as the process variance for the future incremental values.

    - **Lognormal PV** – Similar to the Normal PV, except with a Lognormal distribution.

    - **Gamma PV** – Similar to the Normal PV, except with a Gamma distribution.

    - **No Resampling** – For this option, the actual data triangle is used for each iteration. While this option should not be used for a final simulation, it is useful for diagnostic purposes to see how much of the difference between the deterministic estimate and the bootstrap mean is due to residual sampling versus process variance.

7. **Enable Process Variance** – A key feature of the ODP Bootstrap model is the simulation of Process Variance. The primary calculation steps in the model focus on parameter risk, but process risk is used to add the final “random fluctuations” to the future incremental values. Checking this option will turn these random fluctuations on, while unchecking turns them off.

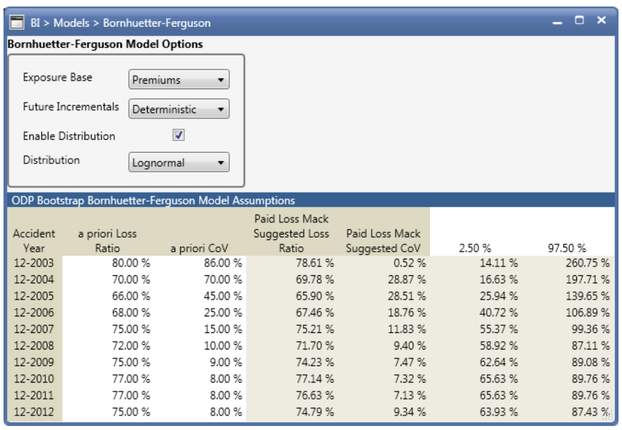
    You can get a measure of the effects of process variance versus residual sampling on your results by running the same model multiple times, with this option and/or residual sampling turned on and off, using the same user-input random seed value each time. The differences in the various simulations will help you diagnose the differences between the deterministic estimate and bootstrap mean for each model.

8. **Process Variance Distribution** – If you check the Enable Process Variance option (above) then the default approach is to add process variance to the projected future development by simulating

from a **Gamma** distribution[24] using each incremental value as the mean and the incremental value times the scale parameter (adjusted by the hetero-factors) as the variance. As an option you can select to use a **Normal** or **Lognormal** distribution instead. For example, if the residuals exhibit little or no skewness then either the normal or lognormal may be a more appropriate distribution for this feature of the data.

In the **Navigation Pane**, select the STOCHASTIC | ODP BOOTSTRAP | MODEL ASSUMPTIONS collection. If you open the **Bornhuetter-Ferguson** window, you can enter the **a priori loss ratio** assumptions, including the **Coefficient of Variation** if you want to include uncertainty for this assumption. After you select RUN DIAGNOSTICS from the **HOME** ribbon, suggested parameters from the **Mack Method** are displayed to provide some additional guidance in making these entries (as illustrated in Image 5-5). Another helpful place to find more information about these inputs is to look at the **Estimated Ultimate Loss Ratio** results that come from simulating the other models.

Based on your selected parameters by period, the percentile columns will show you the 95% confidence interval for the sampled a priori loss ratios.



Accident Year	a priori Loss Ratio	a priori CoV	Paid Loss Mack Suggested Loss Ratio	Paid Loss Mack Suggested CoV	2.50 %	97.50 %
12-2003	80.00 %	86.00 %	78.61 %	0.52 %	14.11 %	260.75 %
12-2004	70.00 %	70.00 %	69.78 %	28.87 %	16.63 %	197.71 %
12-2005	66.00 %	45.00 %	65.90 %	28.51 %	25.94 %	139.65 %
12-2006	68.00 %	25.00 %	67.46 %	18.76 %	40.72 %	106.89 %
12-2007	75.00 %	15.00 %	75.21 %	11.83 %	55.37 %	99.36 %
12-2008	72.00 %	10.00 %	71.70 %	9.40 %	58.92 %	87.11 %
12-2009	75.00 %	9.00 %	74.23 %	7.47 %	62.64 %	89.08 %
12-2010	77.00 %	8.00 %	77.14 %	7.32 %	65.63 %	89.76 %
12-2011	77.00 %	8.00 %	76.63 %	7.13 %	65.63 %	89.76 %
12-2012	75.00 %	8.00 %	74.79 %	9.34 %	63.93 %	87.43 %

Image 5-5: Bornhuetter-Ferguson model assumptions

Note: The column headings for the **Percentiles** (i.e., 2.5% and 97.5%) can be changed to show a different range.

In addition to the basic assumptions for the Bornhuetter-Ferguson model, there are some additional options that can be adjusted as needed.

Note: When you change the exposure base, the a priori column heading and column formatting will change to adjust to the new assumption requirements.

1. **Exposure Base** – When this option is set to **Premiums**, the a priori ultimate losses are calculated by multiplying the a priori loss ratio times the Ultimate Premiums. When this option is set to **Exposures**, the a priori ultimate losses are calculated by multiplying the a priori pure premium by the Ultimate Exposures. When this option is set to **None**, the a priori ultimate losses are entered directly as the a priori assumption.

2. **Future Incrementals** – When this option is set to **Deterministic**, the total Bornhuetter-Ferguson unpaid amounts are converted to the incremental values using the sequential unpaid factors as described Section 3 and illustrated in Appendix A. When this option is set to **Statistical**, the total Bornhuetter-Ferguson unpaid amounts are converted to the incremental values using a Bayesian weighting of the column sums and row sums.

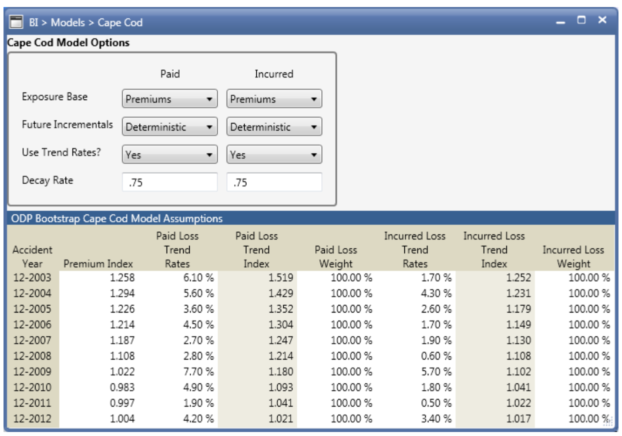
24 The Generalize Linear Model is based on the over-dispersed Poisson distribution, but the gamma distribution is used as a close approximation and simulates significantly faster.

3. **Enable Distribution** – A key feature of the deterministic Bornhuetter-Ferguson method is the ability to “weight” the deterministic chain ladder method with other knowledge about the expected outcome. Moving from a deterministic to a stochastic framework allows the user to also include uncertainty with respect to the expected outcome. By checking this option, the model will incorporate uncertainty in the a priori assumption by simulating a different expected outcome for each iteration. If this option is unchecked, the a priori assumption that you entered into the assumption dialog is used for every iteration.

4. **Distribution –** If you checked Enable Distribution (above) then you can choose to simulate the a priori ultimate from a **Lognormal** distribution or a **Normal** distribution. The selection of the distribution will be based on your understanding of the skewness of the ultimate loss ratios.

Note:
If you select the normal distribution for the simulation of a priori ultimate values it is possible to simulate a negative value. Negative iterations are possible as no zero limitation is applied to this part of the iteration process!

Also in the Model Assumptions collection is the Cape Cod window, which you can use to enter the Premium Index factors, the annual **Loss Trend** (separately for Paid and Incurred) and **Weight** (separately for Paid and Incurred) (as illustrated in Image 5-6). The default values are 1.000, 0% and 100%, respectively, if no data-specific assumptions are entered for the Cape Cod model.


BI > Models > Cape Cod

Cape Cod Model Options

	Paid	Incurred
Exposure Base	Premiums	Premiums
Future Incrementals	Deterministic	Deterministic
Use Trend Rates?	Yes	Yes
Decay Rate	.75	.75

ODP Bootstrap Cape Cod Model Assumptions

Accident Year	Premium Index	Paid Loss Trend Rates	Paid Loss Trend Index	Paid Loss Weight	Incurred Loss Trend Rates	Incurred Loss Trend Index	Incurred Loss Weight
12-2003	1.258	6.10 %	1.519	100.00 %	1.70 %	1.252	100.00 %
12-2004	1.294	5.60 %	1.429	100.00 %	4.30 %	1.231	100.00 %
12-2005	1.226	3.60 %	1.352	100.00 %	2.60 %	1.179	100.00 %
12-2006	1.214	4.50 %	1.304	100.00 %	1.70 %	1.149	100.00 %
12-2007	1.187	2.70 %	1.247	100.00 %	1.90 %	1.130	100.00 %
12-2008	1.108	2.80 %	1.214	100.00 %	0.60 %	1.108	100.00 %
12-2009	1.022	7.70 %	1.180	100.00 %	5.70 %	1.102	100.00 %
12-2010	0.983	4.90 %	1.093	100.00 %	1.80 %	1.041	100.00 %
12-2011	0.997	1.90 %	1.041	100.00 %	0.50 %	1.022	100.00 %
12-2012	1.004	4.20 %	1.021	100.00 %	3.40 %	1.017	100.00 %

Image 5-6:

Cape Cod model assumptions

The **Premium Index** factors will “adjust” the premiums to the current rate level so that they are “re-stated” as if they were written in the latest exposure period. The calculation of the adjusted premium is simply Ultimate Premium x Premium Index. If another **Exposure Base** is selected (see below) then the **Premium Index** factors are not used. The **Loss Trends** are described in more detail below.

The **Weight**s are usually either 100% or 0%. If the **Weight** is 100%, then this year is included when calculating the weighted averages. If the Weight is 0%, then this year will be excluded from the weighted averages. The **Weight**s have the effect of overriding or adjusting the **Decay Rate** (described below) for a particular year(s) as a weight between 0% and 100% is also possible.

In addition to the basic assumptions for the Cape Cod model, there are some additional options that can be adjusted as needed.

1. **Exposure Base** – When this option is set to **Premiums**, the Cape Cod methodology is calculated using the ultimate premiums as the basis, which has the effect of using loss ratios. When this option is set to **Exposures**, the methodology uses ultimate exposures as the basis, which has the effect of using pure premiums. When this option is set to **None**, the methodology uses one as the basis, which has the effect of using ultimate value.

2. **Future Incrementals** – When this option is set to **Deterministic**, the total Cape Cod unpaid amounts are converted to the incremental values using the sequential unpaid factors as described in Section 3 and illustrated in Appendix A. When this option is set to **Statistical**, the total Cape Cod unpaid amounts are converted to the incremental values using a Bayesian weighting of the column sums and row sums.

3. **Use Trend Rates** – A key feature of the deterministic Cape Cod method is the ability to "trend" the losses to remove the effect of inflation and adjust the losses to a common level. By selecting **Yes** for this option, the model will convert the annual **Loss Trend Rates** into "ultimate earned trend factors" and apply them within the algorithm of the Cape Cod model. If **No** is selected for this option, the trend factors are all set to 1.000 (in effect setting the Loss Trend to zero). To trend the losses, each factor is a cumulative multiplication of the later years (e.g., the factor for 2010 is 2010 x 2011 x 2012 x etc.) with the calculation of each individual year factor, except the current year, being:

   (1 + Loss Trend Factor / 12) ^ months in exposure period

   The current year factor is calculated using:

   (1 + Loss Trend Factor / 12) ^ (months in exposure period / 2)

   EXAMPLE

ACCIDENT YEAR	LOSS TREND	TREND INDEX	CUMULATIVE
2009	7.0%	1.0723	1.2961
2010	5.0%	1.0512	1.2087
2011	5.0%	1.0512	1.1499
2012	6.0%	1.0617	1.0939
2013	6.0%	1.0304	1.0304

4. **Decay Rate** – The Decay Rate is used to "credibility weight" the adjusted loss ratio for each accident year with the other years within the Cape Cod methodology and must be between 0% and 100%. Each year's weighted loss ratio is calculated by weighting the loss ratios of that year and the years around it. The weight for each year is calculated using the Decay Rate raised to the power of N, with N being equal to the absolute difference in number of years between the primary accident year and secondary accident year. Loss ratios in closer years get a higher weight than those in years further away. A Decay Rate of 100% effectively gives every accident year equal weight, while a Decay Rate of 0% means that each accident year effectively stands on its own with all other accident years given zero weight.

   If you have not done so, save your file at this point.

Note:
The Paid model options should be set up to achieve the desired Paid model results and **should not** be "ignored" when setting up an Incurred model. For example, if the Paid tail factor should be 1.10 and you simply "ignore" it and set it to 1.00, then not enough of the paid incremental values will be included in the "unpaid incurred" amounts and the model results will diverge.

STEP 3: EVALUATE YOUR DATA WITH THE MODEL'S DIAGNOSTICS

The standard ODP bootstrap model is essentially based on a traditional chain ladder development method. In order to increase the model's predictive power, the data must be consistent with the assumptions that are inherent in the deterministic form of the model (or the model should be adjusted to be consistent with the data). Specifically:

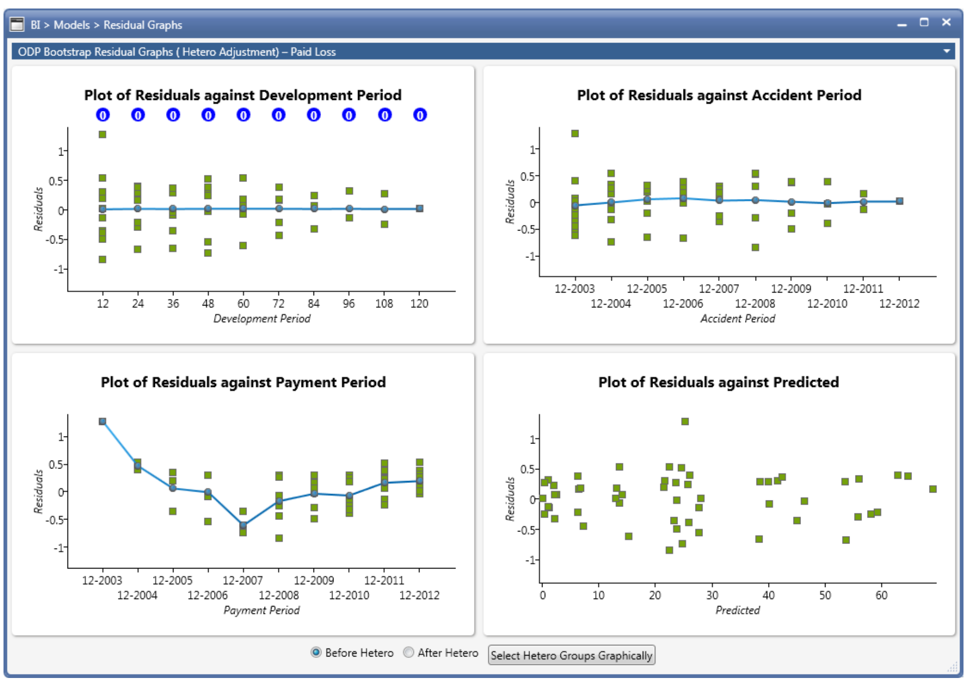
- the expected value of the incremental losses to emerge in the next period is proportional to the total losses emerged to date, by accident year;
- the columns of incremental losses are independent (except for observations in the same accident year); and

- the variance of the next incremental observation is a function of the age and the cumulative losses to date.[25]

The diagnostic output includes a variety of tables and graphs to help the user test these assumptions and then to adjust the model options to improve the statistical fit of the model to the data.

First, from the **Home** ribbon, click on the Run Diagnostics icon to populate the tables and graphs. In an iterative process, you will now want to analyze the diagnostic output, make adjustments to the model options (described above), and then Run Diagnostics again to update the diagnostics results. An additional part of this iterative process is to click on the Run Simulations icon from the **Home** ribbon to run the simulations for the segment you are analyzing. This will allow you to review the model output for the segment, make adjustments to the model options and then either run diagnostics or simulations again until you have optimized the model.

In the **Navigation Pane**, select the Stochastic | ODP Bootstrap | Paid Loss | Diagnostics collection. The Diagnostics collection includes a **Residual Graphs** window (as illustrated in Graph 5-1 prior to heteroscedasticity adjustment). These graphics show plots of the residuals (from Image 5-12) against the development, accident, and payment periods, as well as a plot of the residuals vs. the fitted (i.e., predicted) values. These will help you identify trends or other features in your data that may not be completely modeled by the chain ladder approach, thus indicating that the ODP bootstrap predictions from the data may be less than optimal. Particularly important are the identification of heteroscedasticity and outliers.



Graph 5-1:
Plots of Residuals Prior to Heteroscedasticity Adjustment

25 For some forms of the chain-ladder model, another assumption is that the error terms (residuals) are normally distributed. An advantage of the ODP bootstrap model is that this assumption is not a requirement since the model will simulate using the actual distribution of the residuals, whether they are normally distributed or not. However, since the ODP bootstrap simulations are based on the chain-ladder, this still means that the results may not be optimal.

STEP 3A: IDENTIFY AND ADJUST FOR POTENTIAL HETEROSCEDASTICITY

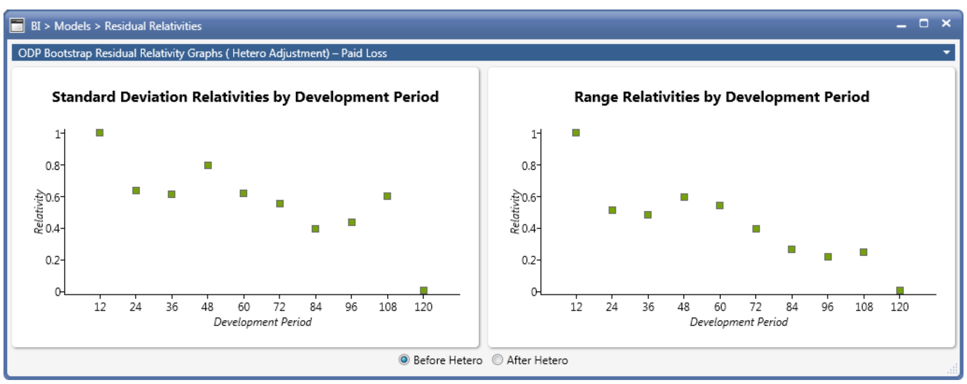
For illustration purposes, we are using the BI data in the ODP_Mack_Hayne.apj file that is included with the system files in the *C:\Users\username\Documents\Milliman\Arius\DemoFiles* directory, where the *username* is your Windows user name.

In the ODP bootstrap model, residuals are resampled with replacement–that is, they are taken from any location in the residual triangle, and placed in another random location to form a sample triangle. Therefore, the residuals should all be independent, identically distributed random numbers. Heteroscedasticity occurs when the residuals are not identically distributed. Usually, we see that some groups of residuals have different standard deviations from other groups. Arius allows you to adjust for this.

The adjustment for heteroscedasticity is made by focusing on the plot by development period in Graph 5-1. Looking at the **Plot of Residuals against Development Period** and the **Residual Relativities** table (illustrated in Graph 5-2 using the factors shown in Image 5-9) you can identify columns with a similar dispersion of residuals. By grouping similar columns together into (potentially) several groups ("heteroscedasticity groups"), you help the model adjust and account for this heteroscedasticity.

Using the **Heteroscedasticity** table (illustrated in Image 5-9) you can "manually" identify the various groups, then use Run Diagnostics and the system will return a new adjusted set of plots and statistics throughout the Diagnostics collection. Alternatively, you can use Suggest Hetero Groups from the ribbon to run the system algorithms for finding groups. Either way, after you group similar residuals together, the modeling goal is to adjust the residuals to a common standard deviation so that they are identically distributed.

Note:
The final development period, which has no residual on the scatter plot, must be included in one of the groups for the model to run. *This final period can never be its own hetero group.* As a general rule, including it in the group with the narrowest dispersion (group 1 in this example) will be a logical choice, but this is not mandatory. The final development period group is also important since it will also be used for all tail factor periods.



Graph 5-2:
Plots of Residual Relativities Prior to Heteroscedasticity Adjustment

BI > Models > Heteroscedasticity

ODP Bootstrap Heteroscedasticity Groups – Paid Loss

Period	1	2	3	4	5	6	7	8	9	10
Age	12	24	36	48	60	72	84	96	108	120
StDev Rel Prior	1.000	0.631	0.612	0.794	0.618	0.550	0.389	0.432	0.596	0.000
Range Rel Prior	1.000	0.506	0.480	0.591	0.541	0.388	0.261	0.216	0.241	0.000
StDev Rel Post	0.775	0.795	0.771	1.000	0.778	0.692	0.490	0.544	0.462	0.000
Range Rel Post	1.000	0.822	0.780	0.960	0.878	0.630	0.424	0.350	0.241	0.000
Group Number	1	0	0	0	0	0	0	0	1	0
Suggested Group	1	0	0	0	0	0	0	0	1	0
Adj. Factor	0.706	1.147	1.147	1.147	1.147	1.147	1.147	1.147	0.706	1.147

Select Hetero Groups Graphically

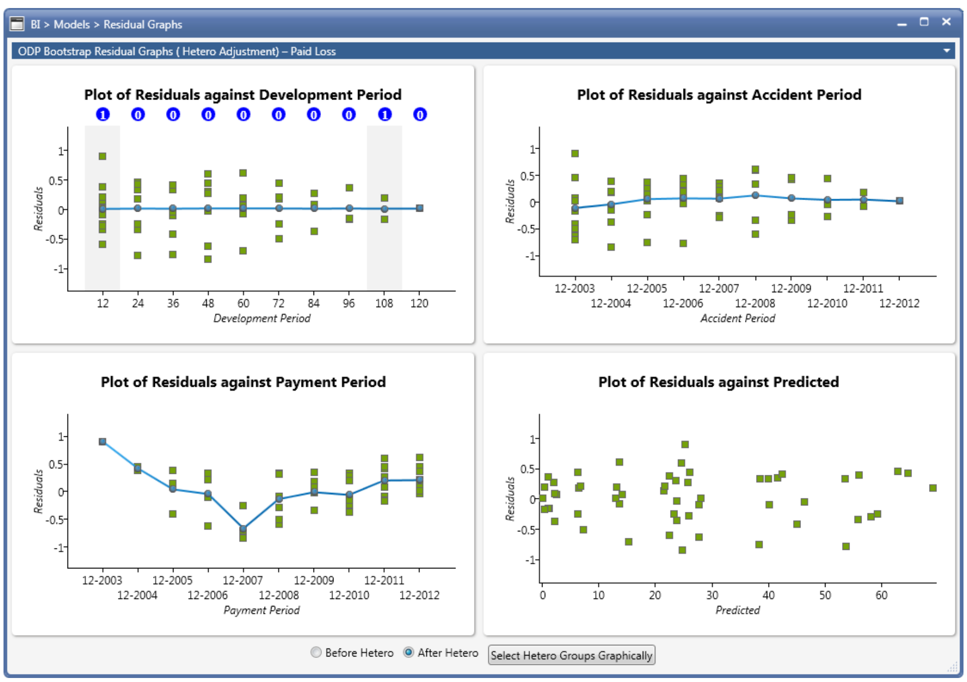
Image 5-9:
Illustration of Suggested Heteroscedasticity Groups

For example, the optimization algorithm used the data from Graph 5-1 to select the hetero groups shown in Image 5-9. While this example shows the "optimal" groups[26], the process of finding the optimal groupings involves trying different groups and comparing the other diagnostics. Indeed, comparing the two sets of relativities in Graph 5-2 shows how different groups are possible and the solution is not obvious. The goal of either the manual iterative process or optimization algorithm is to find the fewest number of groups that result in the "best" diagnostics.[27]

Once the groups are entered in the **Group Number** row of the **Heteroscedasticity** table (as illustrated in Image 5-9), the numbers in the blue circles above the development columns in the **Plot of Residuals against Development Period** will change to match the numbers in the **Heteroscedasticity** table (as illustrated in Graph 5-3).

Rather than enter the hetero group numbers manually, you can click on the **Select Hetero Groups Graphically** buttons at the bottom of either the **Residual Graphs** window or the **Heteroscedasticity** table.

After the hetero groups are entered, use Run Diagnostics and the system will return an adjusted set of plots in the Diagnostics collection. For example, the plots in Graph 5-3 take into account the effect of the groupings made in Image 5-9. Compare the adjusted data plot in Graph 5-3 to the unadjusted data in Graph 5-1 and note how the adjusted residuals are more "consistently" and randomly dispersed.

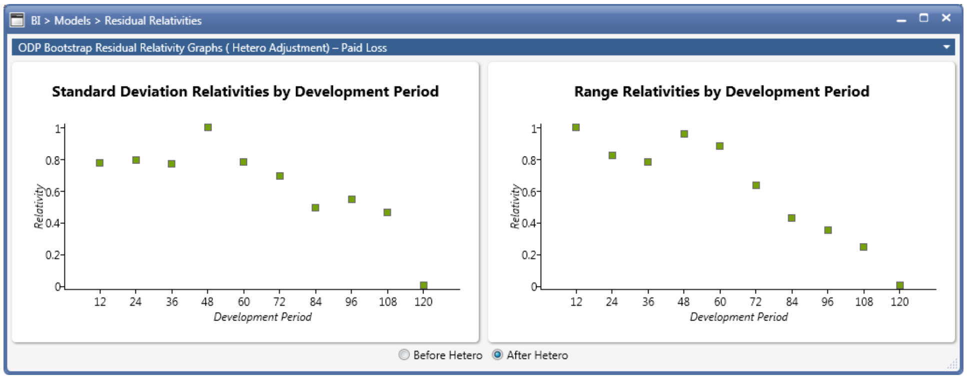


Graph 5-3:
Plots of Residuals after Heteroscedasticity Adjustment

26 For smaller triangles the Suggest Hetero Groups algorithm can search through all of the possible combinations of groups to find the one with the best statistics and, thus, it could be considered optimal. For larger triangle sizes the number of combinations become exponentially too large to check in a reasonable time so other algorithms are used to suggest a solution in a reasonable amount of time.

27 Statistically, the principle of parsimony suggests that the simpler the model (i.e. fewer parameters) the better, all else being equal.

In addition to reviewing the **Residual Graphs** before and after the heteroscedasticity adjustment, the plots of **Residual Relativities** can also be similarly compared. Indeed, comparing Graph 5-4 to Graph 5-2 shows more consistency in the residual relativities.



Graph 5-4:
Plots of Residual Relativities after Heteroscedasticity Adjustment

While it is tempting to use more hetero groups to force even more consistency of the residuals in Graph 5-3 and relativities in Graph 5-4, this will generally be done at the expense of adding more groups (more model parameters) than are optimal. This is not to say that trying other hetero groups is never justified, just that additional groups may not be quite as "optimal" statistically as those in the **Suggested Groups** row.[28]

Before moving on, it is useful to note the values in the rows of the **Heteroscedasticity** table shown in Image 5-9. More specifically:

- **StDev Rel Prior** – this is the standard deviation of the residuals (illustrated in Graphs 5-1 and 5-2), by development period, divided by the largest standard deviation of all the development periods. That is, the period with the largest standard deviation will show a 1.000 here.
- **Range Rel Prior** – this is the difference between the maximum and minimum of the residuals (illustrated in Graphs 5-1 and 5-2), by development period, divided by the largest range of all the development periods. That is, the period with the largest range will show a 1.000 here.
- **StDev Rel Post** – this is the same as **StDev Rel Prior** (illustrated in Graphs 5-3 and 5-4), except the values are calculated using residuals after the heteroscedasticity adjustment.
- **Range Rel Post** – this is the same as **Range Rel Prior** (illustrated in Graphs 5-3 and 5-4), except the values are calculated using residuals after the heteroscedasticity adjustment.
- **Group Number -** you can manually identify groups of development periods that appear to have similar standard deviations, by using a different number for each group. *This is the set of groupings that the model will actually use to adjust for heteroscedasticity in its simulations.* The numbers must be a continuous set of positive integers which includes zero. As a default, a zero

28 For larger triangle sizes it is possible that the algorithm did not find the true optimal solution, although it should be very close. For all triangle sizes, different groupings might make more sense based on the analyst's understanding of the data and what might be driving differences in variance by development period, so statistics from "logical" groups can be compared to statistics from "optimal" groups.

will be entered in each cell so that the model will not use any hetero groups in the initial diagnostic calculations (shown in Image 5-9).[29]

- **Suggested Group** – these are the hetero groups returned when using SUGGEST HETERO GROUPS. This goes through an optimization routine that finds the optimal balance between getting identically distributed random variables, and not having too many hetero group parameters.[30] If you elect to use these recommendations in your model, you can simply copy the figures in this row and paste them into the Group Number row directly above.
- **Adj. Factor** – these are the resulting hetero group adjustment factors that will be used in the simulation process.

The remaining DIAGNOSTIC collection windows include the **Adjusted Triangle**, **Residuals** and **Age-to-Age Factors** tables. If you have checked the **Enable Exposure Adjustment** option, then the incremental values will be divided by the exposures in each period. If you have stub period data (see Section 8) then the last diagonal will be grossed up to a full period. Otherwise, this will simply be the difference in the cumulative values you entered (as illustrated in Image 5-1).

The next diagnostic output is the **Age-to-Age Factors** table (as illustrated in Image 5-10). These factors are essentially calculated as if from a deterministic analysis, except that exposure adjustments and/or stub period adjustments for the last diagonal are also included. In other words, the factors are calculated after cumulating the adjusted incremental values.

BI > Models > Age-To-Age Factors

ODP Bootstrap Age-To-Age Factors (Exposure Adjusted) – Paid Loss

Accident Year	12-24	24-36	36-48	48-60	60-72	72-84	84-96	96-108	108-120	120-ult.
12-2003	3.118	1.448	1.182	1.080	1.035	1.014	1.006	1.001	1.001	
12-2004	3.347	1.480	1.177	1.093	1.044	1.011	1.006	1.003		
12-2005	3.479	1.451	1.220	1.095	1.038	1.015	1.008			
12-2006	3.192	1.549	1.210	1.100	1.047	1.014				
12-2007	3.602	1.550	1.219	1.097	1.044					
12-2008	3.869	1.567	1.234	1.108						
12-2009	3.680	1.555	1.224							
12-2010	3.750	1.506								
12-2011	3.568									
12-2012										
Averages - Paid Loss										
Volume Wtd										
All Periods	3.490	1.511	1.209	1.095	1.042	1.014	1.006	1.002	1.001	
Latest 9	3.490	1.511	1.209	1.095	1.042	1.014	1.006	1.002	1.001	
Latest 7	3.585	1.521	1.209	1.095	1.042	1.014	1.006	1.002	1.001	
Latest 5	3.685	1.544	1.222	1.099	1.042	1.014	1.006	1.002	1.001	
Latest 3	3.661	1.541	1.226	1.102	1.043	1.013	1.006	1.002	1.001	
Selected	3.490	1.511	1.209	1.095	1.042	1.014	1.006	1.002	1.001	

Image 5-10:
Age-to-Age Factor Triangle

In addition to the age-to-age factors, this table also includes the **Averages** for various volume weighted averages of the age-to-age factors (including exposure and/or last diagonal adjustments if appropriate). The **Selected** average factors are based on the **Use Last X Periods** model option, which defaults to a value of 99 periods (as illustrated in Image 5-4 above).

With the average ratio parameters selected, the next diagnostic output is the standardized Pearson residuals shown in the **Residuals** table. These residuals are the basis for the model's simulations. The calculations for the residuals are described in Appendix A, although the residuals will be based on the data adjusted for number of periods in the average, exposures and/or stub periods (as illustrated in Image 5-11). The residuals shown in this table will be *prior to* any hetero group adjustments.

29 More technically, a zero for each development column is required so that the model will group all data into a single group.

30 For a more detailed discussion of the issues related to testing hetero groups and a comparison of using manual adjustments compared to the optimization algorithm see Appendix B.

BI > Models > Residuals

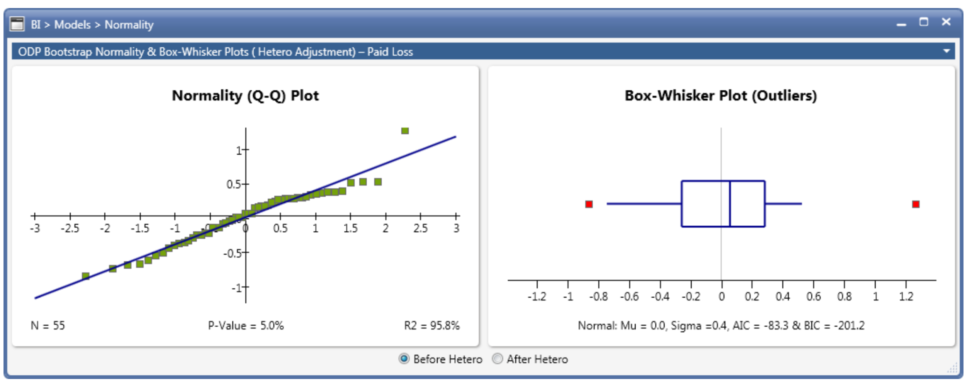
ODP Bootstrap Standardized Pearson Residuals – Paid Loss

Accident Year	12	24	36	48	60	72	84	96	108	120
12-2003	1.264	0.386	(0.370)	(0.560)	(0.626)	(0.451)	0.056	(0.153)	(0.255)	0.000
12-2004	0.522	0.328	(0.097)	(0.746)	(0.082)	0.150	(0.334)	(0.145)	0.255	
12-2005	0.179	0.278	(0.673)	0.257	0.002	(0.224)	0.220	0.305		
12-2006	0.289	(0.689)	0.277	(0.034)	0.164	0.373	0.060			
12-2007	(0.374)	(0.260)	0.292	0.228	0.054	0.172				
12-2008	(0.859)	(0.303)	0.281	0.507	0.521					
12-2009	(0.510)	(0.224)	0.347	0.377						
12-2010	(0.402)	0.363	(0.049)							
12-2011	(0.145)	0.145								
12-2012	0.000									

Image 5-11: Standardized Pearson Residuals

STEP 3B: IDENTIFY AND EXCLUDE OUTLIERS

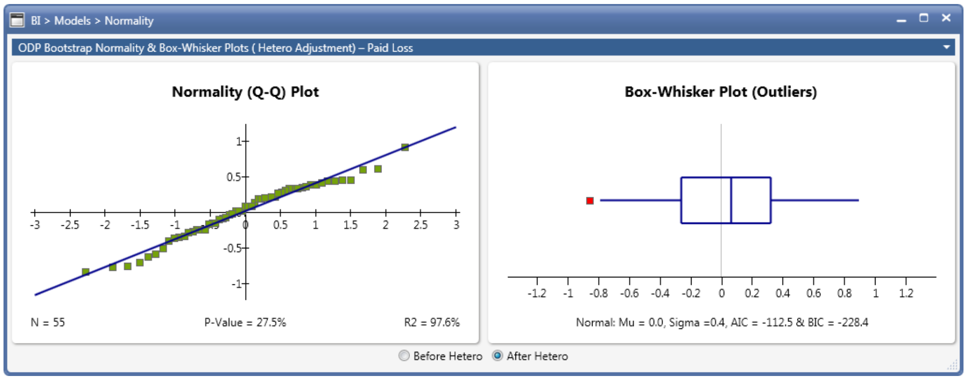
The next DIAGNOSTICS window, **Normality**, will help you judge the general improvement in the model as you change the model options. For example, look at Graph 5-5 below which corresponds to the graphs shown above in Graph 5-1.



Graph 5-5: Normality & Box-Whisker Plots before Heteroscedasticity Adjustment

As noted in Appendix B, the changes in the P-Value, R^2, AIC and BIC values under the Normality (Q-Q) Plot and Box-Whisker Plot are a useful guide. In Graph 5-5 you can see all of these values prior to adjusting for heteroscedasticity. You can also review these graphs before and after other changes to the model options, but the hetero adjustment will usually have the most significant impact on these values.

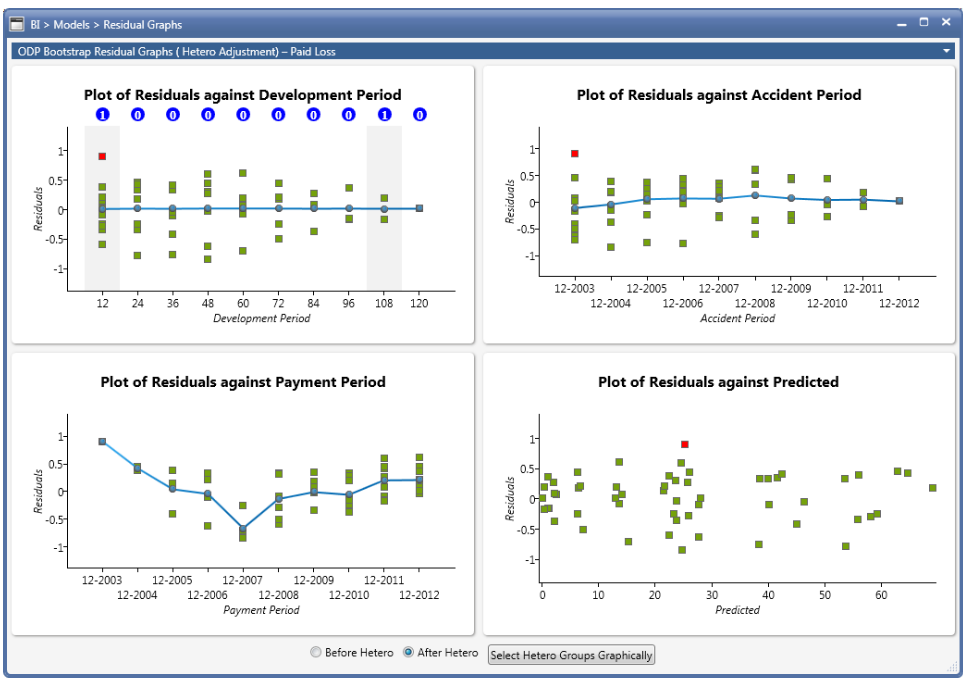
In Graph 5-6 below the Normality (Q-Q) Plot and Box-Whisker Plot are shown after the heteroscedasticity adjustment, and you can see the improvement in the test values, which indicate that the model fit has been improved. These plots are also designed to help you identify possible outliers. For example, one of the outliers was "removed" by the hetero adjustment in Graph 5-3 but, more importantly, the plot is more symmetrical.



Graph 5-6:
Normality & Box-Whisker Plots after Heteroscedasticity Adjustment

Note:
Removing outliers should be done with caution as this will usually reduce the "extremes" of the resulting bootstrap model distribution.

As noted earlier, it might be reasonable to remove outliers before or after adjusting for heteroscedasticity. However, quite often you may find that there are multiple outliers identified in the Box-Whiskers plot, which generally indicates either skewness in the data or a heteroscedasticity problem. In the latter case, adjusting for heteroscedasticity will often "fix" the outlier problem so finding the optimal hetero groups before you remove any outliers will provide better results. On the other hand, skewness may be a feature of the data that you might not want to remove by eliminating outliers.



Graph 5-7:
Illustration of Outlier Selection, *prior to* using RUN DIAGNOSTICS

When you do want to "remove" an outlier from the data, the procedure for doing so is to click on a dot(s) in any of the plots in the **Residual Graphs** window (as illustrated in Graph 5-7) and this will automatically identify it (them) with a one ("1") in the corresponding cell(s) in the **Outliers** triangle (as illustrated in Image 5-12).

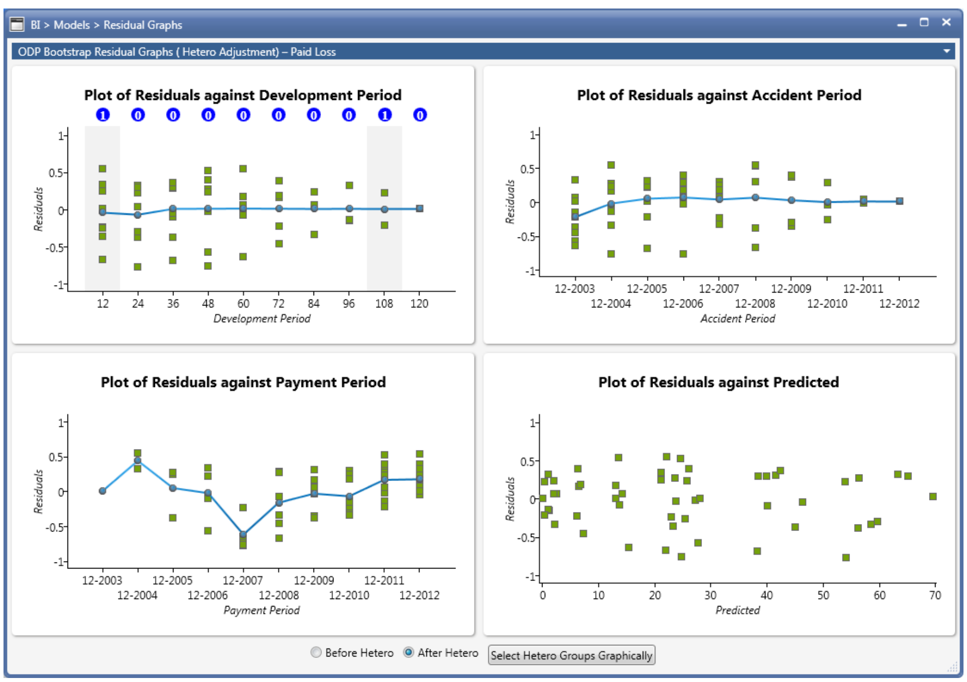
BI > Models > Outliers

ODP Bootstrap Outliers – Paid Loss

Accident Year	12	24	36	48	60	72	84	96	108	120
12-2003	1	0	0	0	0	0	0	0	0	0
12-2004	0	0	0	0	0	0	0	0	0	
12-2005	0	0	0	0	0	0	0	0		
12-2006	0	0	0	0	0	0	0			
12-2007	0	0	0	0	0	0				
12-2008	0	0	0	0	0					
12-2009	0	0	0	0						
12-2010	0	0	0							
12-2011	0	0								
12-2012	0									

Image 5-12:
Outliers triangle with one outlier selected

After the outlier(s) have been identified in this manner, use RUN DIAGNOSTICS again to update the tables and graphs. After the tables and graphs have been updated, the selected outlier(s) will no longer be visible in any of the graphs (as illustrated in Graph 5-8), but you can still see which cell(s) have been eliminated from the simulation (i.e., given no weight in the model) by opening the **Outliers** table (Image 5-12). To restore an outlier to inclusion in the model, you must change the 1 in the **Outliers** table to a zero ("0") and use RUN DIAGNOSTICS again.



Graph 5-8:
Illustration of Outlier Selection, *after* using RUN DIAGNOSTICS

STEP 3C: DEFINE HOW TO HANDLE TAIL FACTORS

After you a better sense from the diagnostics about how your data fits the requirements of the model, and before you run your model simulations, the last consideration that is common to all deterministic methods is the potential inclusion of a tail factor. In the **Navigation Pane**, select the STOCHASTIC | ODP BOOTSTRAP | PAID LOSS | DIAGNOSTICS collection. You can then open the **Tail Factor** window as illustrated in Image 5-13.

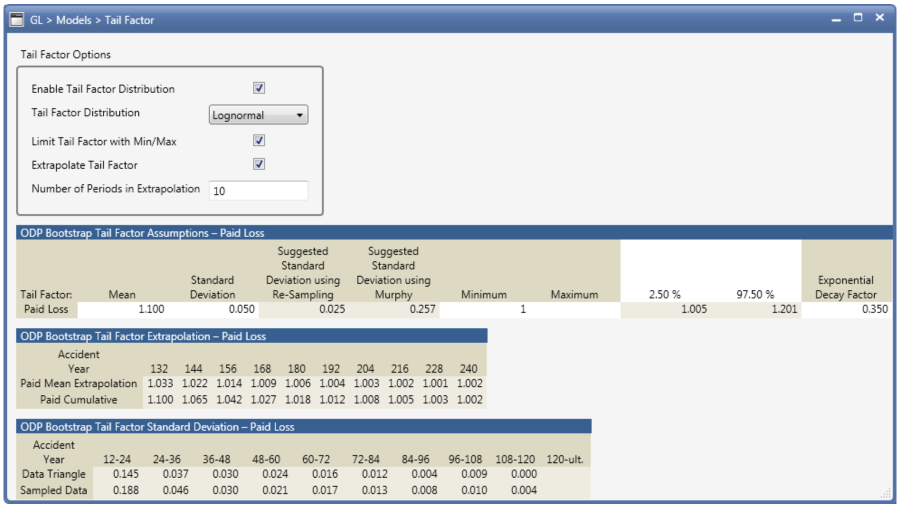


Image 5-13:
Tail Factor assumptions

1. **Enable Tail Factor Distribution** – The system provides two options:
   - Check this option and the system will select a tail for each iteration based on your supplied mean, standard deviation and distribution type; or
   - Uncheck this option and the system will use your **Mean** tail factor amount in each simulation.
2. **Tail Factor Distribution** – If you check **Enable Tail Factor Distribution** (above), you can select from a lognormal or normal distribution from which to simulate the tail factor.
3. **Limit Tail Factor with Min/Max** – If you choose to randomly select a tail factor, you can also provide specific minimums and/or maximum amounts for the model to use. If you provide min/max levels, and check this option, any random amounts outside these levels will be limited to these levels. An example might be to limit factors to a minimum of 1.00.
4. **Extrapolate Tail Factor** – One of the outputs of the simulation is an estimate of the cash flows resulting from the estimated unpaid amounts. These can be presented two ways:
   - If you check this option, the future payments related to the tail will be extended out beyond the development of the triangle itself; or
   - If you uncheck this option, the future payments related to the tail are all accumulated into one final period in the Estimated Cash Flow exhibit.

   The future cash flows related to the tail are extrapolated into the future based on the **Number of Periods in Extrapolation** field and using the value entered into the **Exponential Decay Factor** field. This is important if you want a meaningful **Estimated Cash Flow** table and will also affect the discounted results.

5. **Number of Periods in Extrapolation** – This is an estimate of how many future periods are assumed to be in the tail factor, used as noted above to extrapolate the Cash Flows to future periods.

   The incurred selection will effectively be converted to the number of periods for the paid selection. For example, if the paid model extrapolates 5 years and the incurred model is set to not extrapolate, the simulated paid values will be adjusted to sum to the same ultimate values as the incurred values and the extrapolation for an additional 5 years will be included. Alternatively, if the incurred model extrapolates 5 years and the paid model is set to not extrapolate, the adjustment of the paid simulations will include zeroes beyond the end of the triangle since no payment pattern is simulated beyond the end of the triangle.

6. **Tail Factor** – The parameterization of the tail factor has several related parts (as illustrated in Image 5-13):

   - **Mean** – Enter your best estimate of a tail factor.
   - **Standard Deviation** – If you have checked the **Enable Tail Factor Distribution** option, enter an estimate of the standard deviation of the tail factor.
   - **Suggested Std Deviation: Tail Factor** – After you RUN DIAGNOSTICS**,** suggested parameters for the standard deviation of the tail factor will be shown here based on two different methods:
     - **Resampling** – this method extends the residual resampling that is used for the body of the triangles into the tail of the triangle. 10,000 resampling iterations are done and the implied standard deviation is shown here.
     - **Murphy** – this method uses your a priori loss ratio input (from Model Assumptions | Bornhuetter-Ferguson) and Ultimate Premiums (or Exposures) and calculates the selected ultimate loss for each accident year. The implied tail factor is derived from the difference between the chain ladder ultimate, excluding the tail factor, and the user's selected ultimate for each year. The standard deviation from this set of implied tail factors is shown here.
   - **Min / Max** – If you have checked the **Limit Tail Factor with Min/Max** option, select a minimum and/or maximum for your tail factor.
   - **Percentile** – The 95% confidence interval for the tail factor distribution is shown. Similar to the confidence interval for the Bornhuetter-Ferguson a prior assumption, you can change the percentiles in the heading to see a different interval.
   - **Exponential Decay Factor** – When extrapolation is turned on, one minus this factor is multiplied times the tail factor for each period in the extrapolation. Since the tail factor is a factor to ultimate, each successive factor is a new factor **to ultimate one period** later and dividing each factor by the next factor results in incremental age-to-age factors for the tail (as illustrated in Image 5-14, with a Tail Factor = 1.1, Decay Factor = 35.0% and Number of Years = 10).

Tail Factor Extrapolation					
Period	TF	TF + 1	TF + 2	...	TF + n-1
Extrapolation	TF / TF + 1	TF + 1 / TF + 2	TF + 2 / TF + 3	...	TF + n-1
Cumulative	TF(Mean)	1 + [("TF" - 1) x (1 - Decay)]	1 + [("TF + 1" - 1) x (1 - Decay)]	...	1 + [("TF + n-2" - 1) x (1 - Decay)]
Period	TF	TF + 1	TF + 2	...	TF + 10-1
Extrapolation	1.1000 / 1.0650	1.0650 / 1.0423	1.0423 / 1.0275	...	1.0013
Cumulative	1.1000	1 + [(1.1000 - 1) x (1 - 0.35)]	1 + [(1.0650 - 1) x (1 - 0.35)]	...	1 + [(1.0021 - 1) x (1 - 0.35)]
Period	TF	TF + 1	TF + 2	...	TF + 9
Extrapolation	1.0329	1.0218	1.0144	...	1.0013
Cumulative	1.1000	1.0650	1.0423	...	1.0013

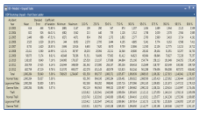
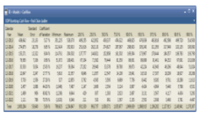
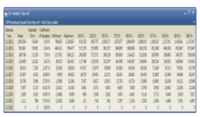
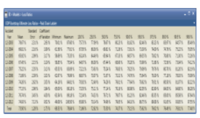
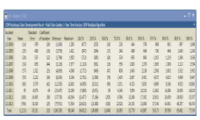
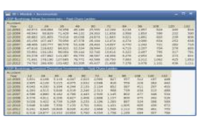
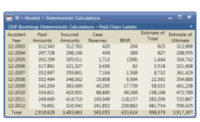
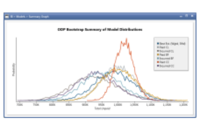
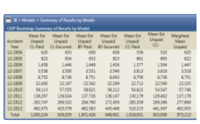
Image 5-14:
Tail Factor Extrapolation Example with Formulas

7. **Tail Factor Standard Deviations** – In addition to the average age-to-age factors (illustrated in Image 5-10 above), the standard deviations of the Age-to-Age factors are shown in the Data Triangle row (illustrated in Image 5-13 above). After you run the simulations, the **Sampled Data** row of this table is also shown, which is based on all of the simulated data. Both of the standard deviation rows can be used to help you select a standard deviation for the tail factors.

In addition to the model diagnostics described above, the results output also has diagnostic features. Thus, running the model using RUN SIMULATIONS, reviewing the model output and adjusting model parameters and assumptions is part of the diagnostic process. Reviewing the model output is discussed in more detail in the remainder of this Section.

SUMMARY OF OUTPUT

The results for each model are shown in their own collection. For example, in the **Navigation Pane**, select the STOCHASTIC | ODP BOOTSTRAP | PAID LOSS | CHAIN LADDER collection to view all of the simulation results for the ODP Paid Chain Ladder model. For the weighted results, there is an additional table and graph which summarize the individual models.

	Estimated Unpaid	Mean, Standard Error, Coefficient of Variation, Min, Max and Percentiles. Total Distributions and TVaRs.
	Total Unpaid Distribution	Histogram and kernel density of total unpaid.
	Estimated Cash Flow	Future calendar period payments.
	Estimated Run-off	Total unpaid as future calendar periods are removed.
	Estimated Loss Ratios	Time zero to ultimate loss ratios.
	Estimated CDR	Claim Development Results (for Time Horizon options only).
	Incremental Values	Mean and standard deviation values for each incremental cell, historical and future.
	Deterministic Calculations	Deterministic Unpaid, Reconciliation and Selected Mean Unpaid (portions only with Weighted Results)
	Summary of Distributions	Histogram distribution for each model and weighted results. (only with Weighted results)
	Summary of Results by Model	Estimated Unpaid, Coefficient of Variation and Mean Loss Ratios for all models. (only with Weighted results)

The flow of the analysis of the results will usually start by reviewing the results of each model individually, then selecting a weighted “best estimate.” Thus, we will discuss how to review results for a model first, before examining how to weight the results of different models.

STEP 4: EVALUATE THE OUTPUT FOR EACH MODEL

After the model diagnostics have been set up and reviewed, the next step in the evaluation of each model is to use RUN SIMULATIONS to run the simulations for the segment you are analyzing. To illustrate the diagnostic elements of the simulation output, we will review the results for the paid chain ladder model.

Estimated Unpaid Results

The **Unpaid Table** illustrated in Table 5-1 was simulated prior to any hetero adjustment. The first diagnostic element of the **Unpaid Table** can be seen by reviewing the Standard Error and Coefficient of Variation columns. As general rules, the standard error should go up as you move from the oldest years to the most recent years and the standard error for the total of all years should be larger than any individual year. In Table 5-1, the standard errors follow these general rules. For the coefficients of variation, they should go down when moving from the oldest years to the more recent years and the coefficient of variation for all years combined should be less than for any individual year.[31] Except for the 2013 year, the coefficients of variation in Table 5-1 also follow the general rules.

BI > Models > Unpaid Table

ODP Bootstrap Unpaid – Paid Chain Ladder

Accident Year	Mean	Standard Error	Coefficient of Variation	Minimum	Maximum	10.0 %	25.0 %	50.0 %	75.0 %	90.0 %	95.0 %	97.5 %	99.0 %	99.5 %	99.9 %
12-2004	620	524	84.47 %	(654)	4,682	71	229	507	894	1,321	1,615	1,931	2,252	2,564	3,364
12-2005	823	637	77.34 %	(1,027)	6,575	126	361	718	1,167	1,686	2,018	2,330	2,788	3,071	3,732
12-2006	1,458	850	58.30 %	(863)	5,744	484	851	1,326	1,946	2,604	3,061	3,452	3,949	4,295	4,923
12-2007	3,538	1,289	36.42 %	302	11,007	1,993	2,626	3,395	4,324	5,267	5,890	6,382	7,000	7,471	8,495
12-2008	8,752	2,020	23.08 %	2,978	17,715	6,239	7,349	8,619	10,043	11,384	12,299	13,066	13,981	14,745	15,991
12-2009	22,400	3,257	14.54 %	10,588	37,666	18,309	20,143	22,239	24,484	26,593	28,006	29,219	30,631	31,661	34,185
12-2010	58,113	5,543	9.54 %	40,602	82,489	51,043	54,298	58,054	61,760	65,341	67,517	69,184	71,281	72,620	75,962
12-2011	138,297	9,659	6.98 %	106,779	180,590	126,111	131,687	138,186	144,664	150,681	154,654	158,089	161,625	164,028	169,155
12-2012	283,747	17,141	6.04 %	227,792	351,727	261,888	272,006	283,270	295,217	306,020	312,542	318,392	324,511	329,069	338,571
12-2013	482,475	41,765	8.66 %	326,786	651,411	428,733	454,339	482,482	510,488	536,057	551,251	565,823	580,228	588,871	614,895
Total	1,000,224	49,068	4.91 %	839,493	1,181,470	937,367	966,873	999,733	1,033,692	1,063,190	1,081,925	1,097,404	1,117,184	1,127,999	1,154,542
Normal %iles	1,000,224	49,065	4.91 %			937,344	967,130	1,000,224	1,033,318	1,063,103	1,080,929	1,096,390	1,114,366	1,126,607	1,151,847
Lognormal %iles	1,000,225	49,164	4.92 %			938,066	966,461	999,019	1,032,674	1,063,933	1,083,091	1,099,987	1,119,964	1,133,774	1,162,788
Gamma %iles	1,000,224	49,092	4.91 %			937,849	966,689	999,421	1,032,883	1,063,631	1,082,319	1,098,703	1,117,959	1,131,197	1,158,819
TVaR						1,009,632	1,020,922	1,039,482	1,063,010	1,087,420	1,103,286	1,117,188	1,132,840	1,143,482	1,162,284
Normal TVaR						1,009,791	1,021,013	1,039,372	1,062,591	1,086,332	1,101,431	1,114,929	1,130,993	1,142,118	1,165,431
Lognormal TVaR						1,009,507	1,020,694	1,039,413	1,063,701	1,089,191	1,105,739	1,120,753	1,138,897	1,151,637	1,178,807
Gamma TVaR						1,009,594	1,020,788	1,039,380	1,063,294	1,088,166	1,104,195	1,118,661	1,136,046	1,148,191	1,173,924

Table 5-1: Estimated Unpaid Model Output

Note: For policy period data or incomplete accident period data, the unpaid data in the last row(s) will be reduced to only include earned exposures.

The reason the standard errors (value scale) tend to go up is that they tend to follow the magnitude of the mean or expected value estimates. The reason the coefficients of variation (percent scale) tend to go down has more to do with the independence in the incremental claim payment stream. For the oldest accident year, there is typically only one (or a few) incremental payment(s) left so the variability of that payment(s) is (almost) fully reflected in the coefficient. For the most current accident year, the "up and down" variations in the future incremental payment stream can offset each other thus causing the total variation to be a function of the correlation between each incremental payment for that accident year (i.e., the incremental payments are assumed independent).

Note: Caution should be exercised in the interpretation and adjustments for increases in the coefficient of variation in recent years. While keeping the theory in mind is appropriate, this must be balanced with the need to keep from underestimating the uncertainty of the more recent years.

The coefficient of variation rules noted above are a reflection of the step 7's described in Section 3 (and Appendix A), in the sense that they describe the process variance in the model. While the coefficients of variation should go down, if they do start going back up in the most recent year(s), as illustrated in Table 5-1 for 2013, then this could be the result of the following issues:

31 These standard error and coefficient of variation rules are based on the independence of the incremental process risk and assume that the underlying exposures are relatively stable from year to year – i.e., no radical changes. In practice, random changes do occur from one year to the next which could cause the actual standard errors to deviate from these rules somewhat. In other words, these rules should generally hold true, but are not considered hard and fast rules in every case. Strictly speaking, the total all years rules assume that the individual years are not positively correlated.

1. The parameter uncertainty tends to increase when moving from the oldest years to the more recent years as more and more parameters are used in the model. In the most recent year(s), the parameter uncertainty could be “overpowering” the process uncertainty, causing the coefficient of variation to start going back up. At the very least, the increasing parameter uncertainty will cause the rate of decrease in the coefficient of variation to slow down.
2. If the increase in the most recent year(s) is significant, then this could indicate that the model is overestimating the uncertainty in those years. If this is the case, then an adjustment to the model parameters may be needed (e.g., limit incrementals to zero, etc.) or you may need to use a Bornhuetter-Ferguson or Cape Cod model instead of a Chain Ladder model.

While we mentioned the rules for the standard error and coefficient of variation for the total of all years, it is also worth noting that in addition to the correlation (independence) within each accident year the total of all years also includes the impact of the correlation (independence) between accident years. In essence, when one or more accident years are “bad” we do not expect all accident years to be “bad.” To see this impact, you can add the accident year standard errors and note that they will not sum to the standard error for all years combined.[32]

The next diagnostic element in the **Unpaid Table** is the **Minimum** and **Maximum** columns. In these columns, the smallest and largest values, respectively, from among all iterations of the simulation are displayed. These values can be reviewed judgmentally to make sure that they are not outside the “realm of possibility.” If they do seem a bit unrealistic then they could indicate the need to review the model options. For example, the presence of negative numbers might lead to changing one or both of the options which limit incremental values to zero. Sometimes “extreme” outliers in the results will show up in these columns and may also distort the histogram (discussed later in this section).

Risk Measures

Also included in Table 5-1, notice that there are three rows of “Percentile” numbers and then four rows of TVaR numbers at the bottom of these tables under each of the percentile columns. For the three “Percentile” rows, the normal, lognormal and gamma distributions, respectively, have been fit to the Total unpaid claim distribution. The fitted mean, standard deviation and selected percentiles are shown under the Mean, Standard Error and Percentile columns, respectively, so that the smoothed results can be used to judge the quality of fit for each distribution or other purposes such as parameterizing a DFA model or using smoothed results in the tail of the distribution.

The Tail Value at Risk (TVaR)[33] is the average of all of the simulated values equal to or greater than the percentile value. For example, in Table 5-1 the 75th percentile value for the total unpaid for all accident years combined is 1,033,692 and the average of all simulated values that are greater than or equal to 1,033,692 is 1,063,010. The “Normal TVAR,” “Lognormal TVaR” and “Gamma TVaR” rows are calculated the same way, except that instead of using the actual simulated values from the model the respective fitted distributions are used in the calculations.

To interpret the TVaR numbers, the question we are trying to answer with a TVaR number is “if the actual outcome does exceed the X percentile value, on average how much might it exceed that value by?” This is an important question related to risk based capital calculations and other technical aspects of enterprise risk management, although a more complete discussion is beyond the scope of this manual. It is worth noting, however, that the purpose of the normal, lognormal and gamma TVaR

32 Likewise, the minimum, maximum and each of the percentile columns will not sum to the total for all years combined. In contrast, adding the mean values for each accident year will sum to the total for all years combined.

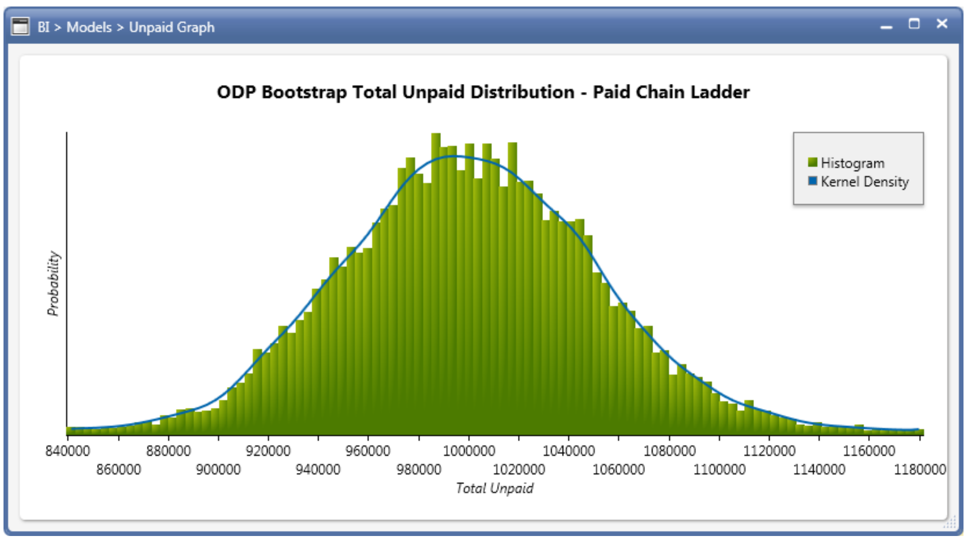
33 The Tail Value at Risk is sometimes referred to as the Conditional Tail Expectation.

numbers is to provide “smoothed” values in the sense that some of the random noise is kept from distorting the calculations.

Total Unpaid Distribution Graph

The final model output from the simulations for each model is a histogram of the estimated unpaid amounts for the total of all accident years combined, as illustrated in Graph 5-9. The **Unpaid Graph**, or histogram, is created by dividing the range of all values from the simulation (using the maximum and minimum values) into one hundred “buckets” of equal size and counting the number of simulations that fall within each “bucket.” Dividing by the total number of simulations (10,000 in this case) results in the frequency or probability for each “bucket” in the graph.

Since the simulation results often look “jagged” (as they do in Graph 5-9) a kernel density function is also used to calculate a “smoothed” line fit to the histogram values. The kernel density distribution is represented by the blue line in Graph 5-9.[34]



Graph 5-9:
Total Unpaid Distribution

When you initially parameterize and run the model, you may find the resulting graph to be extremely narrow – almost a straight line. This is normally caused by a handful of extreme iterations. Many of the percentile results in the **Unpaid Table** may still appear reasonable, but it is still important to remove these extreme iterations since they will unduly affect your mean result. One of the most common causes of the extreme iteration is negative incremental values which can sometimes also result in an unrealistically high age-to-age factor. Thus, checking the **Limit Incremental to Zero** constraints may help remove these extreme iterations.

Estimated Cash Flow Results

In addition to the results by accident year, we can also review the model output by calendar year (or by future diagonal) in the **Cash Flow** table as illustrated in Table 5-2. Comparing Table 5-2 to 5-1, notice

34 In simple terms, a kernel density function can be thought of as a weighted average of values “close” to each point in the “jagged” distribution with progressively less weight being given to values the further they are from the point being evaluated. For a more detailed discussion of kernel density functions, see Wand & Jones, “Kernel Smoothing,” Chapman & Hall. 1995.

that the Total row is identical since the total is the same whether you add the parts horizontally or diagonally. Similar diagnostic issues can be reviewed in this table, except that the relative values of the standard errors and coefficients of variation move in the opposite direction for calendar years compared to accident years. This should make intuitive sense as the "final" payments projected the farthest out into the future should be the smallest yet relatively most uncertain.

BI > Models > Cashflow

ODP Bootstrap Cash Flow – Paid Chain Ladder

Calendar Year	Mean	Standard Error	Coefficient of Variation	Minimum	Maximum	10.0 %	25.0 %	50.0 %	75.0 %	90.0 %	95.0 %	97.5 %	99.0 %	99.5 %	99.9 %
12-2014	438,547	21,545	4.91 %	359,069	523,367	410,793	423,920	438,362	452,901	466,193	474,165	481,465	490,272	495,640	506,804
12-2015	274,790	16,433	5.98 %	212,935	335,317	253,581	263,547	274,614	285,842	296,296	302,121	307,124	312,305	317,609	327,670
12-2016	153,280	11,434	7.46 %	113,266	201,348	138,794	145,462	153,059	160,798	168,023	172,614	176,679	181,189	184,339	190,624
12-2017	76,781	7,525	9.80 %	52,306	107,903	67,355	71,551	76,536	81,731	86,730	89,441	92,036	94,711	97,206	101,979
12-2018	33,426	4,618	13.82 %	18,778	52,330	27,603	30,151	33,302	36,456	39,367	41,139	42,804	44,781	46,706	49,502
12-2019	12,664	2,805	22.15 %	4,512	27,911	9,190	10,662	12,504	14,484	16,360	17,533	18,565	19,639	20,775	22,793
12-2020	5,697	1,929	33.87 %	256	14,814	3,382	4,313	5,531	6,891	8,231	9,139	10,006	10,971	11,681	13,046
12-2021	2,453	1,334	54.39 %	(1,402)	10,533	909	1,498	2,295	3,224	4,218	4,903	5,515	6,283	6,848	8,121
12-2022	1,483	1,070	72.15 %	(1,564)	8,296	288	724	1,326	2,081	2,893	3,450	3,978	4,664	5,139	6,799
12-2023	1,102	896	81.28 %	(1,694)	6,904	137	434	931	1,576	2,310	2,803	3,267	3,903	4,420	5,548
Total	1,000,224	49,068	4.91 %	839,493	1,181,470	937,367	966,873	999,733	1,033,692	1,063,190	1,081,925	1,097,404	1,117,184	1,127,999	1,154,542

Table 5-2:
Estimated Cash Flow Model Output

Estimated Unpaid Claim Runoff Results

Another report similar to the **Cash Flow** table is the **Run-off** table. Rather than looking at individual diagonal results, the **Run-off** table starts with the total unpaid results and then looks at how the total unpaid will decrease over time as successive diagonals are removed, as illustrated in Table 5-3. Comparing Table 5-3 to 5-1 & 5-2, notice that the first row of Table 5-3 is identical to the Total rows in Tables 5-1 and 5-2. Each successive row in Table 5-3 is then the total of the remaining diagonals.

BI > Models > Run-off

ODP Bootstrap Unpaid Claim Run-off – Paid Chain Ladder

Calendar Year	Mean	Standard Error	Coefficient of Variation	Minimum	Maximum	10.0 %	25.0 %	50.0 %	75.0 %	90.0 %	95.0 %	97.5 %	99.0 %	99.5 %	99.9 %
12-2013	1,000,224	49,068	4.91 %	839,493	1,181,470	937,367	966,873	999,733	1,033,692	1,063,190	1,081,925	1,097,404	1,117,184	1,127,999	1,154,542
12-2014	561,677	32,309	5.75 %	452,728	685,219	520,623	539,663	561,170	583,334	603,097	614,853	625,963	638,948	648,359	666,619
12-2015	286,886	20,105	7.01 %	217,470	362,819	261,287	273,124	286,448	300,236	312,338	320,763	327,846	336,371	340,871	353,318
12-2016	133,607	12,277	9.19 %	89,265	184,705	117,876	125,354	133,248	141,823	149,497	154,117	158,409	163,577	166,979	172,407
12-2017	56,826	7,592	13.36 %	31,129	86,295	47,191	51,576	56,663	61,858	66,657	69,555	72,280	75,385	77,558	81,259
12-2018	23,400	4,991	21.33 %	5,500	42,917	17,154	19,939	23,219	26,677	29,991	31,851	33,654	35,599	37,019	40,486
12-2019	10,736	3,474	32.36 %	(2,050)	25,265	6,420	8,306	10,571	12,922	15,292	16,803	18,104	19,586	20,639	23,022
12-2020	5,039	2,391	47.45 %	(3,055)	15,842	2,133	3,379	4,851	6,506	8,187	9,192	10,195	11,317	12,454	14,826
12-2021	2,585	1,597	61.76 %	(3,100)	12,340	707	1,444	2,420	3,529	4,668	5,432	6,156	7,035	7,968	9,616
12-2022	1,102	896	81.28 %	(1,694)	6,904	137	434	931	1,576	2,310	2,803	3,267	3,903	4,420	5,548
12-2023	0	0	0.00 %	0	0	0	0	0	0	0	0	0	0	0	0

Table 5-3:
Estimated Unpaid Claim Run-Off Model Output

Estimated Ultimate Loss Ratio Results

The next collection table shows the ultimate **Loss Ratios** by accident year as illustrated in Table 5-4. If there are no earned premiums or ultimate premiums input into the model, then this table will not be filled in since the model cannot calculate a loss ratio without the premium information.[35]

BI > Models > Loss Ratios

ODP Bootstrap Ultimate Loss Ratios – Paid Chain Ladder

Accident Year	Mean	Standard Error	Coefficient of Variation	Minimum	Maximum	10.0 %	25.0 %	50.0 %	75.0 %	90.0 %	95.0 %	97.5 %	99.0 %	99.5 %	99.9 %
12-2004	78.67 %	2.30 %	2.92 %	70.05 %	88.21 %	75.75 %	77.10 %	78.67 %	80.21 %	81.62 %	82.47 %	83.15 %	83.99 %	84.42 %	85.51 %
12-2005	69.81 %	2.15 %	3.08 %	61.99 %	77.49 %	67.05 %	68.37 %	69.79 %	71.26 %	72.60 %	73.33 %	74.10 %	74.74 %	75.21 %	76.14 %
12-2006	65.91 %	2.08 %	3.16 %	58.11 %	72.94 %	63.24 %	64.50 %	65.93 %	67.31 %	68.55 %	69.30 %	70.00 %	70.74 %	71.28 %	72.16 %
12-2007	67.45 %	2.14 %	3.17 %	59.48 %	74.86 %	64.69 %	66.00 %	67.47 %	68.89 %	70.23 %	70.96 %	71.68 %	72.49 %	72.99 %	74.21 %
12-2008	75.22 %	2.34 %	3.11 %	65.69 %	83.43 %	72.26 %	73.63 %	75.21 %	76.80 %	78.22 %	79.10 %	79.84 %	80.66 %	81.22 %	82.13 %
12-2009	71.69 %	2.39 %	3.34 %	62.57 %	80.02 %	68.60 %	70.07 %	71.70 %	73.33 %	74.76 %	75.61 %	76.35 %	77.26 %	77.83 %	79.09 %
12-2010	74.26 %	2.59 %	3.49 %	65.01 %	83.53 %	70.94 %	72.48 %	74.27 %	76.01 %	77.61 %	78.53 %	79.31 %	80.35 %	80.91 %	82.17 %
12-2011	77.16 %	2.96 %	3.84 %	66.16 %	89.46 %	73.41 %	75.17 %	77.14 %	79.16 %	80.96 %	82.01 %	83.01 %	84.17 %	84.93 %	86.73 %
12-2012	76.74 %	3.60 %	4.69 %	64.32 %	91.00 %	72.16 %	74.26 %	76.66 %	79.19 %	81.47 %	82.77 %	84.02 %	85.31 %	86.22 %	87.90 %
12-2013	74.89 %	6.29 %	8.39 %	51.32 %	99.94 %	66.77 %	70.69 %	74.92 %	79.13 %	82.92 %	85.21 %	87.43 %	89.60 %	90.90 %	94.85 %
Total	73.49 %	1.21 %	1.65 %	69.17 %	77.71 %	71.95 %	72.66 %	73.50 %	74.31 %	75.05 %	75.51 %	75.89 %	76.34 %	76.61 %	77.39 %

Table 5-4:
Estimated Loss Ratio Model Output

Note:
For policy period data or incomplete accident period data, the unpaid data in the last row(s) will be reduced to only include earned exposures. However, since the earned exposures are divided by the Earned Premium to calculate the loss ratios we have a match of losses to premium.

35 Earned premiums are used as the denominator of the loss ratios. However, if earned premiums are not input then earned premiums are estimated from the ultimate premiums.

Unlike the **Unpaid**, **Cash Flow** and **Run-off** tables, the values in the **Loss Ratios** table are calculated from all simulated values, not just the values beyond the end of the triangles. In other words, since the simulated sample triangles represent other possibilities of what could have happened in the past and the "squaring of the triangle" and process variance represent what could happen as those other possible past values play out into the future, we have enough information to estimate the complete variability in the loss ratio from day one in each accident year until all claims are completely paid and settled.[36]

Because we are using all simulated values, the standard errors in Table 5-4 should be proportionate to the means while the coefficients of variation should be relatively constant by accident year. Diagnostically, the increases in standard error and coefficient of variation for the latest few years are consistent with the reasons cited earlier for the **Unpaid Table**. For the Chain Ladder and Cape Cod models, this table can also help parameterize the a priori means and coefficients of variation for the Bornhuetter-Ferguson models.

Estimated Claim Development Result

The next collection table shows the **CDR** or Claim Development Result. When the **Bootstrap Option** in the OPTIONS tab of the MODEL OPTIONS dialog is set to **Ultimate** then this table will be blank, since it is only calculated when one of the time horizon options is used. Thus, when either the **Time Horizon – ODP Process** or **Time Horizon – ODP Residual** option is selected (which will activate the algorithms described in Section 3 and Appendix A) then this table will be created as illustrated in Table 5-5 for the **ODP Residual** option.

BI > Models > CDR

ODP Bootstrap Claim Development Result – Paid Chain Ladder, 1-Year Time Horizon, ODP Residual Algorithm

Accident Year	Mean	Standard Error	Coefficient of Variation	Minimum	Maximum	10.0 %	25.0 %	50.0 %	75.0 %	90.0 %	95.0 %	97.5 %	99.0 %	99.5 %	99.9 %
12-2004	(1)	403	(422)	(1,589)	1,459	(522)	(270)	3	268	513	664	788	915	1,023	1,236
12-2005	(3)	459	(136)	(1,750)	1,607	(591)	(318)	6	295	576	753	915	1,091	1,201	1,421
12-2006	(21)	603	(28)	(2,014)	2,154	(798)	(427)	(22)	381	755	981	1,152	1,380	1,561	1,844
12-2007	(19)	950	(50)	(3,463)	3,999	(1,251)	(655)	(15)	629	1,216	1,541	1,804	2,128	2,380	2,867
12-2008	(5)	1,483	(314)	(5,395)	5,208	(1,923)	(1,004)	3	989	1,913	2,466	2,882	3,465	3,786	4,470
12-2009	26	2,399	93	(9,168)	9,252	(3,018)	(1,608)	0	1,670	3,075	3,968	4,693	5,515	6,262	7,919
12-2010	(67)	4,034	(60)	(13,964)	14,515	(5,165)	(2,800)	(76)	2,669	5,131	6,600	7,803	9,256	10,253	12,299
12-2011	(1)	6,850	(10,513)	(26,625)	26,218	(8,759)	(4,589)	(1)	4,557	8,779	11,362	13,311	16,021	17,591	20,851
12-2012	(301)	11,790	(39)	(42,328)	46,600	(15,288)	(8,172)	(325)	7,543	14,853	19,292	23,180	27,500	30,176	36,210
12-2013	(368)	20,953	(57)	(82,957)	78,743	(26,833)	(14,328)	(575)	13,693	27,106	34,384	41,198	48,256	53,138	63,064
Total	(760)	29,016	(38)	(117,891)	109,610	(38,230)	(20,141)	(779)	18,580	36,733	46,976	56,141	66,642	73,804	86,738

Table 5-5:
Estimated Claim Development Result Output

The output for this table is calculated by subtracting the mean of the **Unpaid Table** when the Ultimate option is selected from each of the iterations when one of the time horizon options is selected. For example, the **Unpaid Table** for the **ODP Residual** option is shown in Table 5-6. Subtracting the Total **Mean** from Table 5-1 of 1,000,224 from the Total **Mean** from Table 5-6 of 999,463, results in the Total **Mean** for Table 5-5 of (760). The CDR is used to calculate the required capital for Solvency II regulations in Europe.

36 If we are only interested in the "remaining" volatility in the loss ratio, then the values in the **Estimated Unpaid** table can be added to the cumulative values in the data input table and divided by the premiums.

BI > Models > Unpaid Table

ODP Bootstrap Unpaid – Paid Chain Ladder, 1-Year Time Horizon, ODP Residual Algorithm

Accident Year	Mean	Standard Error	Coefficient of Variation	Minimum	Maximum	10.0 %	25.0 %	50.0 %	75.0 %	90.0 %	95.0 %	97.5 %	99.0 %	99.5 %	99.9 %
12-2004	619	403	65.09 %	(969)	2,079	99	350	623	889	1,134	1,284	1,408	1,535	1,644	1,856
12-2005	820	459	56.00 %	(927)	2,431	233	506	829	1,118	1,399	1,576	1,739	1,914	2,025	2,244
12-2006	1,437	603	41.93 %	(556)	3,613	661	1,031	1,436	1,839	2,213	2,439	2,610	2,838	3,019	3,302
12-2007	3,519	950	26.98 %	75	7,538	2,287	2,883	3,523	4,168	4,755	5,079	5,342	5,666	5,918	6,405
12-2008	8,747	1,483	16.96 %	3,357	13,960	6,829	7,748	8,755	9,741	10,665	11,218	11,634	12,217	12,538	13,222
12-2009	22,426	2,399	10.70 %	13,231	31,652	19,382	20,791	22,400	24,070	25,475	26,368	27,093	27,915	28,662	30,319
12-2010	58,046	4,034	6.95 %	44,150	72,629	52,948	55,313	58,037	60,783	63,244	64,713	65,916	67,369	68,366	70,412
12-2011	138,296	6,850	4.95 %	111,671	164,514	129,538	133,708	138,295	142,854	147,076	149,659	151,608	154,318	155,888	159,147
12-2012	283,446	11,790	4.16 %	241,419	330,347	268,459	275,575	283,422	291,290	298,600	303,039	306,927	311,246	313,922	319,957
12-2013	482,107	20,953	4.35 %	399,518	561,218	455,642	468,147	481,900	496,168	509,581	516,859	523,673	530,731	535,613	545,540
Total	999,463	29,016	2.90 %	882,333	1,109,833	961,994	980,083	999,445	1,018,804	1,036,956	1,047,200	1,056,364	1,066,866	1,074,027	1,086,962
Normal %iles	999,463	29,014	2.90 %			962,280	979,894	999,463	1,019,033	1,036,647	1,047,188	1,056,330	1,066,961	1,074,199	1,089,124
Lognormal %iles	999,464	29,053	2.91 %			962,516	979,649	999,042	1,018,819	1,036,953	1,047,960	1,057,601	1,068,923	1,076,701	1,092,919
Gamma %iles	999,463	29,032	2.90 %			962,442	979,731	999,182	1,018,889	1,036,845	1,047,692	1,057,160	1,068,240	1,075,828	1,091,588
TVaR						1,005,149	1,011,767	1,022,583	1,036,411	1,050,341	1,059,054	1,067,037	1,076,122	1,082,439	1,094,950
Normal TVaR						1,005,121	1,011,757	1,022,613	1,036,344	1,050,383	1,059,312	1,067,293	1,076,793	1,083,371	1,097,157
Lognormal TVaR						1,005,023	1,011,650	1,022,637	1,036,745	1,051,395	1,060,826	1,069,332	1,079,546	1,086,678	1,101,780
Gamma TVaR						1,005,054	1,011,683	1,022,625	1,036,603	1,051,042	1,060,299	1,068,621	1,078,584	1,085,520	1,100,152

Table 5-6:
Estimated Unpaid Model Output for **ODP Residual** option

Estimated Incremental Results

The next collection table is designed to help you take a deeper look at the simulations and to understand the reasons for increases in the coefficients of variation (illustrated in Tables 5-1 and 5-4). They show the mean and standard deviations, respectively, by accident year by incremental period. As illustrated in Table 5-7, both the Mean and Standard Deviation **Incrementals** can be reviewed down each column or across each row to look for any irregularities in the expected patterns.

As you can see by looking down the 24 and 36 month columns in Table 5-7, it appears as though there might be "too much" variability in future incremental values for years 2012 and 2013 – i.e., those "future" values do not appear consistent with the values in the prior years. This does not imply that the "historical" values are correct and that the "future" values are overstated, just that they are not always consistent. These inconsistencies appear to be impacting both the unpaid and loss ratio results for 2012 and 2013.

BI > Models > Incrementals

ODP Bootstrap Mean Incrementals – Paid Chain Ladder

Accident Year	12	24	36	48	60	72	84	96	108	120	132
12-2004	42,062	104,697	74,943	46,179	25,576	12,196	4,133	1,960	627	245	620
12-2005	40,045	99,832	71,425	44,121	24,366	11,653	3,961	1,853	590	231	592
12-2006	40,883	101,781	72,998	45,027	24,879	11,847	4,018	1,890	617	244	597
12-2007	43,152	107,443	76,939	47,383	26,188	12,478	4,281	1,997	652	251	638
12-2008	48,469	120,759	86,559	53,337	29,497	14,092	4,769	2,256	725	288	714
12-2009	47,616	118,626	84,916	52,329	28,943	13,831	4,674	2,192	707	281	715
12-2010	53,944	134,190	95,997	59,151	32,791	15,614	5,301	2,494	812	315	786
12-2011	63,386	157,970	113,258	69,861	38,588	18,385	6,275	2,947	945	364	931
12-2012	71,665	178,199	127,588	78,814	43,607	20,834	7,052	3,318	1,074	413	1,047
12-2013	74,749	186,241	133,356	82,222	45,546	21,620	7,370	3,473	1,109	436	1,102

ODP Bootstrap Standard Deviation Incrementals – Paid Chain Ladder

Accident Year	12	24	36	48	60	72	84	96	108	120	132
12-2004	3,326	5,280	4,444	3,513	2,600	1,814	1,062	716	400	254	524
12-2005	3,391	5,307	4,528	3,538	2,611	1,822	1,049	719	407	371	513
12-2006	3,511	5,565	4,683	3,685	2,744	1,870	1,096	755	538	390	528
12-2007	3,701	5,801	4,902	3,924	2,887	2,009	1,166	930	551	405	563
12-2008	3,982	6,285	5,367	4,167	3,140	2,171	1,434	1,017	618	460	621
12-2009	3,973	6,312	5,388	4,191	3,095	2,480	1,439	1,004	592	448	617
12-2010	4,366	6,969	5,868	4,575	3,964	2,725	1,598	1,112	677	496	679
12-2011	4,902	7,953	6,561	6,260	4,478	3,088	1,828	1,291	768	568	767
12-2012	5,419	8,540	9,614	7,145	5,074	3,445	1,984	1,406	859	630	851
12-2013	5,659	17,694	13,459	9,374	6,178	3,823	2,167	1,479	885	652	896

Table 5-7:
Estimated Incrementals by Accident Year by Development Period

Deterministic Calculations

In addition to the tables displaying results for the estimated distributions, the collection also contains a table showing the deterministic results for the same assumptions for each model. For example, the Deterministic Calculations table is illustrated in Table 5-8 for the Paid Chain Ladder model. The results for this table are calculated based on one iteration of the model with all stochastic elements turned off.

BI > Models > Deterministic Calculations

ODP Bootstrap Deterministic Calculations – Paid Chain Ladder

Accident Year	Paid Amounts	Incurred Amounts	Case Reserves	IBNR	Estimate of Total	Estimate of Ultimate
12-2004	312,343	312,763	420	204	625	312,968
12-2005	297,728	298,166	439	389	827	298,555
12-2006	303,299	305,011	1,713	(269)	1,444	304,742
12-2007	317,891	321,327	3,436	82	3,518	321,408
12-2008	352,697	359,861	7,164	1,568	8,732	361,429
12-2009	332,494	348,302	15,808	6,584	22,392	354,886
12-2010	343,204	383,499	40,295	17,739	58,033	401,238
12-2011	334,621	423,501	88,880	49,288	138,168	472,789
12-2012	249,660	414,710	165,049	118,157	283,206	532,867
12-2013	74,691	316,543	241,852	239,882	481,734	556,425
Total	2,918,628	3,483,683	565,055	433,624	998,679	3,917,307

Table 5-8:
Deterministic Calculations

Results of Other Models

Thus far we have only reviewed results for the Paid Chain Ladder model. By selecting the other models from either the MODEL OPTIONS or CHOOSE MODELS icon and then using RUN SIMULATIONS to run the simulations again, you can complete the iterative process of reviewing the results for each model and adjusting model parameters as necessary. Since the tables and graphs described for Step 4 in this Section are the same for each model there is no need to repeat all of the previous discussion. However, it is useful to look at a couple of the models to compare some aspects of the results.

Starting with the Incurred Chain Ladder, Table 5-9 illustrates how the estimated unpaid claim distribution might differ for an incurred model compared to a paid model. Comparing these results with the results in Table 5-1 for paid data, the first thing you should notice is that the mean unpaid results are lower. This is consistent with differences in deterministic paid and incurred chain ladder methods. It is also worth noting that incurred results will not always be less than paid results – it is only the fact that the results diverge that is consistent with deterministic methods, not the direction of the divergence.

BI > Models > Unpaid Table

ODP Bootstrap Unpaid – Incurred Chain Ladder

Accident Year	Mean	Standard Error	Coefficient of Variation	Minimum	Maximum	10.0 %	25.0 %	50.0 %	75.0 %	90.0 %	95.0 %	97.5 %	99.0 %	99.5 %	99.9 %
12-2004	633	527	83.26 %	(813)	4,328	69	243	520	911	1,338	1,632	1,929	2,319	2,563	3,232
12-2005	834	642	77.04 %	(818)	4,652	129	366	724	1,182	1,697	2,054	2,404	2,812	3,131	3,608
12-2006	1,446	858	59.35 %	(695)	6,481	459	819	1,330	1,933	2,600	3,031	3,443	4,002	4,353	5,148
12-2007	3,500	1,318	37.66 %	(605)	11,109	1,936	2,554	3,368	4,293	5,295	5,835	6,437	7,129	7,701	8,641
12-2008	8,736	2,166	24.80 %	2,598	19,008	6,103	7,192	8,571	10,089	11,613	12,521	13,455	14,400	15,240	16,851
12-2009	22,197	3,752	16.90 %	9,546	38,233	17,637	19,545	21,938	24,508	27,177	28,766	30,290	32,061	33,214	36,344
12-2010	57,555	7,557	13.13 %	33,713	95,357	47,940	52,295	57,329	62,530	67,273	70,448	73,183	76,080	77,925	82,311
12-2011	136,024	15,538	11.42 %	85,494	199,234	116,279	125,337	136,088	146,264	156,044	162,305	167,524	173,041	176,146	185,712
12-2012	269,532	31,369	11.64 %	156,839	397,387	229,905	248,061	269,060	290,009	310,011	322,121	333,742	345,349	352,330	370,735
12-2013	425,578	57,690	13.56 %	226,577	634,133	352,578	385,860	425,317	463,693	500,188	522,253	543,073	563,583	580,295	606,841
Total	926,035	68,775	7.43 %	688,553	1,209,385	838,540	879,257	924,827	972,514	1,014,798	1,041,064	1,063,068	1,087,189	1,105,657	1,141,997
Normal %iles	926,035	68,772	7.43 %			837,901	879,649	926,035	972,421	1,014,170	1,039,155	1,060,825	1,086,022	1,103,179	1,138,556
Lognormal %iles	926,045	69,175	7.47 %			839,279	878,158	923,472	971,125	1,016,112	1,044,026	1,068,857	1,098,473	1,119,107	1,162,884
Gamma %iles	926,035	68,912	7.44 %			838,891	878,670	924,326	971,538	1,015,375	1,042,226	1,065,891	1,093,848	1,113,159	1,153,678
TVaR						939,246	954,872	980,993	1,014,426	1,048,131	1,070,189	1,088,938	1,111,674	1,127,377	1,164,589
Normal TVaR						939,446	955,174	980,907	1,013,451	1,046,728	1,067,892	1,086,810	1,109,327	1,124,920	1,157,596
Lognormal TVaR						938,880	954,564	981,111	1,016,014	1,053,141	1,077,501	1,099,774	1,126,901	1,146,087	1,187,377
Gamma TVaR						939,039	954,725	980,977	1,015,034	1,050,756	1,073,929	1,094,938	1,120,305	1,138,101	1,176,00[illegible]

Table 5-9:
Estimated Unpaid Model Output – Incurred Chain Ladder

The next difference is the standard errors which are larger for the incurred model. The differences in the standard errors are a function of the residuals for each model, which in turn are a function of the underlying data and the reserving philosophy of the company.[37] Thinking about how to measure risk (as described in Section 1), these differences could impact the weights given to each model when determining the “best” distribution. Finally, the coefficient of variation is decreasing and then increases in the latest two years similarly to the paid model.

BI > Models > Unpaid Table

ODP Bootstrap Unpaid – Paid Bornhuetter-Ferguson

Accident Year	Mean	Standard Error	Coefficient of Variation	Minimum	Maximum	10.0 %	25.0 %	50.0 %	75.0 %	90.0 %	95.0 %	97.5 %	99.0 %	99.5 %	99.9 %
12-2004	630	952	151.07 %	(1,359)	20,539	3	64	294	830	1,641	2,354	3,194	4,281	5,291	9,162
12-2005	833	1,005	120.70 %	(3,258)	10,945	25	172	526	1,156	2,045	2,753	3,437	4,834	5,762	8,028
12-2006	1,449	1,138	78.52 %	(963)	11,001	288	625	1,179	1,988	2,966	3,672	4,320	5,263	6,078	7,234
12-2007	3,531	1,623	45.96 %	(58)	13,089	1,685	2,368	3,285	4,440	5,718	6,524	7,322	8,298	9,031	11,268
12-2008	8,751	2,505	28.63 %	1,712	21,798	5,689	6,951	8,550	10,284	12,005	13,271	14,265	15,680	16,626	18,344
12-2009	22,542	3,936	17.46 %	11,412	39,656	17,652	19,787	22,296	25,086	27,752	29,403	30,863	32,451	33,773	36,539
12-2010	58,621	7,573	12.92 %	35,875	91,935	49,153	53,361	58,145	63,533	68,524	71,573	74,447	77,667	80,275	85,862
12-2011	137,726	13,590	9.87 %	92,396	197,162	120,670	128,420	137,199	146,639	155,322	160,900	165,230	170,952	176,557	184,795
12-2012	284,760	25,996	9.13 %	191,249	403,695	252,151	266,415	283,790	301,831	318,507	328,420	337,996	349,810	357,623	378,184
12-2013	482,583	41,926	8.69 %	346,927	693,374	429,951	453,079	481,083	510,127	536,894	553,832	569,268	587,476	602,017	628,338
Total	1,001,426	52,665	5.26 %	811,249	1,213,455	935,089	965,238	1,000,040	1,036,829	1,068,924	1,089,688	1,106,781	1,126,530	1,144,083	1,178,414
Normal %iles	1,001,426	52,663	5.26 %			933,936	965,906	1,001,426	1,036,946	1,068,916	1,088,048	1,104,643	1,123,937	1,137,076	1,164,166
Lognormal %iles	1,001,427	52,708	5.26 %			934,856	965,187	1,000,043	1,036,157	1,069,775	1,090,413	1,108,636	1,130,207	1,145,135	1,176,541
Gamma %iles	1,001,426	52,644	5.26 %			934,580	965,433	1,000,504	1,036,413	1,069,457	1,089,563	1,107,203	1,127,949	1,142,222	1,172,025
TVaR						1,011,386	1,023,409	1,043,524	1,069,232	1,095,884	1,113,491	1,129,438	1,149,670	1,165,303	1,196,546
Normal TVaR						1,011,695	1,023,739	1,043,445	1,068,366	1,093,848	1,110,054	1,124,541	1,141,783	1,153,723	1,178,746
Lognormal TVaR						1,011,355	1,023,343	1,043,433	1,069,548	1,097,007	1,114,858	1,131,072	1,150,687	1,164,474	1,193,911
Gamma TVaR						1,011,458	1,023,460	1,043,414	1,069,110	1,095,866	1,113,126	1,128,713	1,147,457	1,160,559	1,188,341

Table 5-10:
Estimated Unpaid Model Output – Paid Bornhuetter-Ferguson

The Paid Bornhuetter-Ferguson model illustrated in Table 5-10 provides another brief comparison to the Paid Chain Ladder model. For the Paid Bornhuetter-Ferguson model the mean results are slightly higher than the Paid Chain Ladder model results, but the standard error by year is reasonably close to the standard error for the Paid Chain Ladder model except for 2013 which is quite a bit less, causing the total standard error to be lower. The most interesting difference illustrated in Table 5-10 is the shape of the coefficients of variation, which consistently decrease (i.e., they do not increase in the last two years). This is actually a more common feature of both the Bornhuetter-Ferguson and Cape Cod models since the processes in step 7 (see Section 3 and Appendix A) tend to weigh more “stable” expected results with more “volatile” loss development factor results by year.

STEP 5: WEIGHT THE OUTPUT FOR EACH MODEL

Once the results for each model have been reviewed and “finalized” in the iterative process involving the diagnostics and model output, they can be “combined” by assigning a weight to the results of each model. Similar to the process of weighting the results of different deterministic methods to arrive at an actuarial “best estimate,” the process of weighting the results of different stochastic models will result in an actuarial “best estimate of the distribution.”

In the **Navigation Pane**, select the Stochastic | ODP Bootstrap | ODP Summary | Assumptions collection. The Assumptions collection includes a **Model Weights** table, as illustrated in Table 5-11, in which a weight can be selected for each model by accident year. When the model is run, the weighted results will be compiled and output in the Summary Results collection. In general, you will want to focus on the individual models before deciding on how to weight them together to arrive at your “best estimate.” Therefore, you should select options in the Model Assumptions and Diagnostics collections first and the **Model Weights** table can be left blank until all models are complete.

37 Like for the means, whether the paid standard errors are less or greater than the incurred standard errors depends on the underlying data.

BI > Models > Model Weights

ODP Bootstrap Weights by Model

Accident Year	Chain Ladder - Paid	Chain Ladder - Incurred	Bornhuett... Ferguson - Paid	Bornhuett... Ferguson - Incurred	Cape Cod - Paid	Cape Cod - Incurred
12-2004						
12-2005						
12-2006						
12-2007						
12-2008						
12-2009						
12-2010						
12-2011						
12-2012						
12-2013						

Table 5-11:
Model Weights Table

With all of the ODP Bootstrap models selected, and no weights entered for the moment, a summary of the results for each model is provided as the **Summary of Results by Model** table (illustrated in Table 5-12) in the SUMMARY RESULTS collection. For example, consider the summarized results for all six models in Table 5-12 and note that the estimated unpaid is consistent with the results in Tables 5-1, 5-9 and 5-10.[38]

BI > Models > Summary of Results by Model

ODP Bootstrap Summary of Results by Model

Accident Year	Mean Est Unpaid CL-Paid	Mean Est Unpaid CL-Incurred	Mean Est Unpaid BF-Paid	Mean Est Unpaid BF-Incurred	Mean Est Unpaid CC-Paid	Mean Est Unpaid CC-	Weighted Mean Unpaid
12-2004	620	633	630	636	536	532	0
12-2005	823	834	833	823	861	865	0
12-2006	1,458	1,446	1,449	1,436	1,577	1,564	0
12-2007	3,538	3,500	3,531	3,546	3,813	3,816	0
12-2008	8,752	8,736	8,751	8,693	8,756	8,736	0
12-2009	22,400	22,197	22,542	22,269	22,712	22,590	0
12-2010	58,113	57,555	58,621	58,212	54,622	54,547	0
12-2011	138,297	136,024	137,726	136,167	130,179	129,402	0
12-2012	283,747	269,532	284,760	272,459	285,359	269,386	0
12-2013	482,475	425,578	482,583	445,448	510,215	441,567	0
Total	1,000,224	926,035	1,001,426	949,691	1,018,631	933,006	0

Table 5-12:
Summary of Results by Model (in part)

By comparing the results for all six models (or fewer if fewer are used) a qualitative assessment of the relative merits of each model can be determined. This can be determined separately for each year so that different weights can be used for each year. The assessment of the weights can then be entered into the **Model Weights** table in the ASSUMPTIONS collection. For example, Table 5-13 illustrates an example of weights for the data used in this section. While the weights used in this example are between 0% and 100% and add to 100% for each accident year, this is not a requirement. Similar to the Deterministic method weights, any positive value can be used as a weight and the total for each accident year will be adjusted to calculate the weights so they do add up to 100%.

38 The **Summary of Results by Model** illustrated in Tables 5-12, 5-14 and 5-17 have been reduced to save space. The actual table also includes similar summaries for the coefficients of variation and loss ratios by model.

BI > Models > Model Weights

ODP Bootstrap Weights by Model

Accident Year	Chain Ladder - Paid	Chain Ladder - Incurred	Bornhuett... Ferguson - Paid	Bornhuett... Ferguson - Incurred	Cape Cod - Paid	Cape Cod - Incurred
12-2004	50.00 %	50.00 %				
12-2005	50.00 %	50.00 %				
12-2006	50.00 %	50.00 %				
12-2007	50.00 %	50.00 %				
12-2008	50.00 %	50.00 %				
12-2009	50.00 %	50.00 %				
12-2010	50.00 %	50.00 %				
12-2011	50.00 %	50.00 %				
12-2012	25.00 %		25.00 %	25.00 %		25.00 %
12-2013	25.00 %		25.00 %	25.00 %		25.00 %

Table 5-13:
Model Weights by Accident Year

With the weights entered, use the RUN SIMULATIONS icon to run the simulations again. The weighted results will now be displayed in the Best Estimate column of **Summary of Results by Model** table. Since we are concerned with the entire distribution and not just a single point estimate, the weights by year are used to randomly sample the specified percentage of the iterations from each model. An example of the "blended" results is illustrated in Table 5-14.

With weights entered in the **Model Weights** table, the model will also populate the **Unpaid Table**, **Cash Flow**, **Run-off**, **Loss Ratios**, **CDR** (if one of the Time Horizon options is used) and **Incrementals** tables and the **Unpaid Graph** for the weighted (Best Estimate) results, similar to what are shown in Tables 5-1, 5-2, 5-3, 5-4, 5-5 and 5-7 and Graph 5-9, respectively. The "Best Estimate" results can be used as is or further adjusted to account for information outside of these models or other actuarial judgments.

BI > Models > Summary of Results by Model

ODP Bootstrap Summary of Results by Model

Accident Year	Mean Est Unpaid CL-Paid	Mean Est Unpaid CL-Incurred	Mean Est Unpaid BF-Paid	Mean Est Unpaid BF-Incurred	Mean Est Unpaid CC-Paid	Mean Est Unpaid CC-	Weighted Mean Unpaid
12-2004	620	633	630	636	536	532	625
12-2005	823	834	833	823	861	865	822
12-2006	1,458	1,446	1,449	1,436	1,577	1,564	1,447
12-2007	3,538	3,500	3,531	3,546	3,813	3,816	3,518
12-2008	8,752	8,736	8,751	8,693	8,756	8,736	8,751
12-2009	22,400	22,197	22,542	22,269	22,712	22,590	22,325
12-2010	58,113	57,555	58,621	58,212	54,622	54,547	57,746
12-2011	138,297	136,024	137,726	136,167	130,179	129,402	137,178
12-2012	283,747	269,532	284,760	272,459	285,359	269,386	277,869
12-2013	482,475	425,578	482,583	445,448	510,215	441,567	462,933
Total	1,000,224	926,035	1,001,426	949,691	1,018,631	933,006	973,213

Table 5-14:
Summary of Results by Model (in part) with Weighted Best Estimate

In addition to the tables displaying results for the estimated distributions, the output also contains a table showing the deterministic results for the same assumptions for each model. Similar to the **Summary of Results by Model** table, the **Deterministic Calculations** table (illustrated in Table 5-15) shows the total unpaid estimate for each of the six methods, as well as the "Best Estimate" using the same weights used for the models (in this example, the weights shown in Table 5-13).

Comparing Table 5-15 with Table 5-14, note that all of the deterministic estimates in Table 5-15 are not equal to the mean estimates in Table 5-14. Some sources characterize the deterministic estimates as the "true mean" and infer that with enough simulations the mean of the simulated distributions will converge to the deterministic estimate. As long as the stochastic and deterministic assumptions are identical, then this may in fact be true, but other sources and research would infer that the skewness of the assumptions used in the ODP bootstrap model may be different than the "implied symmetric" assumptions of the deterministic methods. If so, then it would be appropriate to assume that the deterministic estimates could represent a mode or median instead of a mean, except when all assumptions are symmetrical (e.g., assuming a normal distribution).[39]

BI > Models > Deterministic Calculations

☐ Use Selected Unpaid as Mean

ODP Bootstrap Deterministic Calculations – Best Estimate (Weighted)

Accident Year	Chain Ladder-Paid	Chain Ladder-Incurred	BF- Paid	BF-Incurred	Cape Cod-Paid	Cape Cod-Incurred	Weighted Estimate	Reconciled-Paid Amts.	Reconciled-Incurred Amts.	Reconciled-Case Reserv.	Reconciled-IBNR	Reconciled-Est. of Ult.	Reconciled-Est. of Total Unpaid.	Selected Total Unpaid
12-2004	625	624	636	635	538	538	624	312,343	312,763	420	204	312,968	624	
12-2005	827	827	830	829	857	856	827	297,728	298,166	439	388	298,555	827	
12-2006	1,444	1,446	1,446	1,448	1,565	1,566	1,445	303,299	305,011	1,713	(268)	304,743	1,445	
12-2007	3,518	3,523	3,546	3,551	3,811	3,813	3,520	317,891	321,327	3,436	84	321,411	3,520	
12-2008	8,732	8,732	8,707	8,707	8,730	8,729	8,732	352,697	359,861	7,164	1,568	361,429	8,732	
12-2009	22,392	22,312	22,485	22,402	22,693	22,595	22,352	332,494	348,302	15,808	6,544	354,847	22,352	
12-2010	58,033	57,835	58,633	58,379	54,459	54,569	57,934	343,204	383,499	40,295	17,639	401,139	57,934	
12-2011	138,168	136,323	137,909	136,288	130,137	129,551	137,246	334,621	423,501	88,880	48,366	471,866	137,246	
12-2012	283,206	269,651	284,560	272,981	285,402	269,541	277,572	249,660	414,710	165,049	112,523	527,233	277,572	
12-2013	481,734	425,839	483,103	446,420	509,960	441,935	463,298	74,691	316,543	241,852	221,447	537,989	463,298	
Total	998,679	927,111	1,001,856	951,641	1,018,153	933,694	973,551	2,918,628	3,483,683	565,055	408,496	3,892,179	973,551	985,580

Table 5-15:
Deterministic Estimates by Method

In addition to these methods, you could also use other methods (e.g., Berquist-Sherman, frequency/severity, etc.) to inform your judgment about the "best estimate" of the mean. Thus, using all of the methods and models at your disposal you can make a final selection of the mean unpaid claim estimate and enter it in the **Deterministic Calculations** table as illustrated in Table 5-16.

BI > Models > Deterministic Calculations

☑ Use Selected Unpaid as Mean

ODP Bootstrap Deterministic Calculations – Best Estimate (Weighted, Unshifted)

Accident Year	Chain Ladder-Paid	Chain Ladder-Incurred	BF- Paid	BF-Incurred	Cape Cod-Paid	Cape Cod-Incurred	Weighted Estimate	Reconciled-Paid Amts.	Reconciled-Incurred Amts.	Reconciled-Case Reserv.	Reconciled-IBNR	Reconciled-Est. of Ult.	Reconciled-Est. of Total Unpaid.	Selected Total Unpaid
12-2004	625	624	636	635	538	538	624	312,343	312,763	420	204	312,968	624	625
12-2005	827	827	830	829	857	856	827	297,728	298,166	439	388	298,555	827	830
12-2006	1,444	1,446	1,446	1,448	1,565	1,566	1,445	303,299	305,011	1,713	(268)	304,743	1,445	1,445
12-2007	3,518	3,523	3,546	3,551	3,811	3,813	3,520	317,891	321,327	3,436	84	321,411	3,520	3,520
12-2008	8,732	8,732	8,707	8,707	8,730	8,729	8,732	352,697	359,861	7,164	1,568	361,429	8,732	8,735
12-2009	22,392	22,312	22,485	22,402	22,693	22,595	22,352	332,494	348,302	15,808	6,544	354,847	22,352	22,375
12-2010	58,033	57,835	58,633	58,379	54,459	54,569	57,934	343,204	383,499	40,295	17,639	401,139	57,934	58,000
12-2011	138,168	136,323	137,909	136,288	130,137	129,551	137,246	334,621	423,501	88,880	48,366	471,866	137,246	137,250
12-2012	283,206	269,651	284,560	272,981	285,402	269,541	277,572	249,660	414,710	165,049	112,523	527,233	277,572	281,800
12-2013	481,734	425,839	483,103	446,420	509,960	441,935	463,298	74,691	316,543	241,852	221,447	537,989	463,298	471,000
Total	998,679	927,111	1,001,856	951,641	1,018,153	933,694	973,551	2,918,628	3,483,683	565,055	408,496	3,892,179	973,551	985,580

Table 5-16:
Deterministic Estimates by Method

After the selected best estimate has been entered in the **Deterministic Calculations** table, you can check the **Use Selected Unpaid as Mean** option and use RUN SIMULATIONS to run the simulations again. The model will then "shift" the mean of the **Best Estimate** (i.e., weighted) results so that they match your selected unpaid by year. The "shift" is done in an additive fashion by adding the difference between the selected unpaid and the weighted mean unpaid by accident period to each iteration. The results for this example are illustrated in Table 5-17.

39 Actuarial Standard of Practice No.43 – Property/Casualty Unpaid Claim Estimates, Appendix 3 states in part "As to the definition of the term [Actuarial Central Estimate], it is generally agreed that most traditional actuarial methods are meant to produce some measure of central tendency. But what measure? There are several different measures of central tendency, including (for example) mean, median, mode…"

BI > Models > Summary of Results by Model

ODP Bootstrap Summary of Results by Model

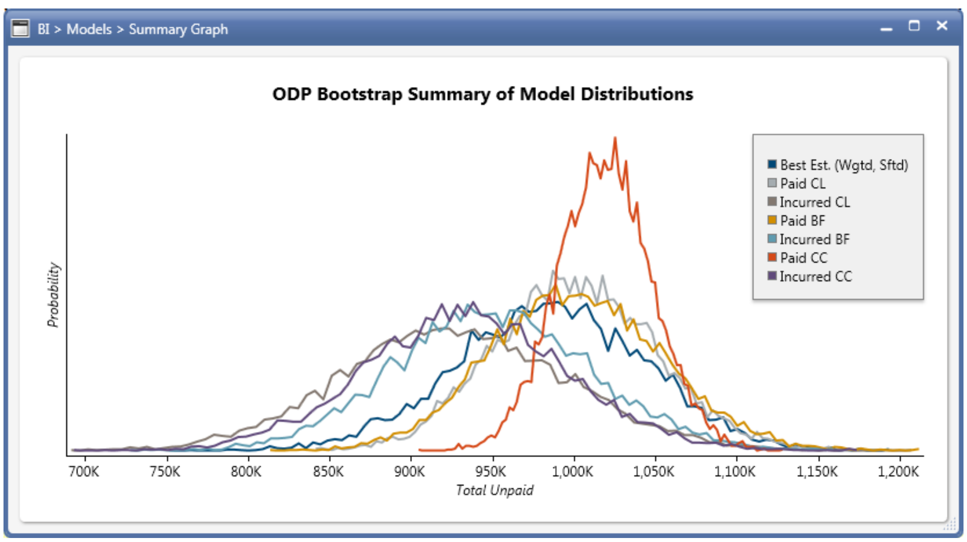
Accident Year	Mean Est Unpaid CL (Unshifted)-Paid	Mean Est Unpaid CL (Unshifted)-Incurred	Mean Est Unpaid BF (Unshifted)-Paid	Mean Est Unpaid BF (Unshifted)-Incurred	Mean Est Unpaid CC (Unshifted)-Paid	Mean Est Unpaid CC (Unshifted)-Incurred	Best Est (Shifted) Unpaid
12-2004	620	633	630	636	536	532	625
12-2005	823	834	833	823	861	865	830
12-2006	1,458	1,446	1,449	1,436	1,577	1,564	1,445
12-2007	3,538	3,500	3,531	3,546	3,813	3,816	3,520
12-2008	8,752	8,736	8,751	8,693	8,756	8,736	8,735
12-2009	22,400	22,197	22,542	22,269	22,712	22,590	22,375
12-2010	58,113	57,555	58,621	58,212	54,622	54,547	58,000
12-2011	138,297	136,024	137,726	136,167	130,179	129,402	137,250
12-2012	283,747	269,532	284,760	272,459	285,359	269,386	281,800
12-2013	482,475	425,578	482,583	445,448	510,215	441,567	471,000
Total	1,000,224	926,035	1,001,426	949,691	1,018,631	933,006	985,580

Table 5-17: Summary of Results by Model (in part) with Selected Best Estimate

Note: Only the Best Estimate (Weighted) results will be shifted. The results from the individual models will not be shifted in any of the results tables or graphs.

With weights entered in the **Model Weights** table and the **Use Selected Unpaid as Mean** option checked, the model will also populate the **Unpaid Table**, **Cash Flow**, **Run-off**, **Loss Ratios** and **CDR** (if one of the Time Horizon options is used) tables and the **Unpaid Graph** for the selected ("shifted" Best Estimate) results, similar to what are shown in Tables 5-1, 5-2, 5-3, 5-4 and 5-5 and Graph 5-9, respectively. Note, however, that the **Incrementals** tables will continue to show the weighted results prior to shifting.

For the weighted results in the SUMMARY RESULTS collection one additional graph is included which summarizes the distribution graphs for all models. An example of the **Summary Graph** is illustrated in Graph 5-10.



Graph 5-10: Summary of Model Distributions

[THIS PAGE INTENTIONALLY LEFT BLANK]

6. Using the Mack Bootstrap Models

Even though the Arius system has numerous options to help you obtain the best model possible for your data, you can obtain valuable diagnostic information and even initial distribution estimates for a line of business with only a few steps, which can be summarized as:

- enter the data to be modeled,
- run the model diagnostics to populate the necessary statistics and fields, and
- run the simulation to estimate future results (i.e., use the default model settings).

Of course, the diagnostics and model results can be used to evaluate and improve how your model fits your data. Understanding the purpose and use of the diagnostic tools requires some prior statistical knowledge so we direct the interested reader to Appendix B, which provides a general overview of the diagnostic process. Therefore, this section assumes prior knowledge of statistics, and starts with the basics of running a model and builds on that foundation by exploring all of the different models, model options, diagnostics, and model output. Note however, that Appendix B is based on using the diagnostic tools for the ODP Bootstrap model, so the differences when using the Mack Bootstrap models will be discussed here.

REQUIRED DATA: MACK BOOTSTRAP PAID MODEL

Inputs for the paid model are relatively simple. You can start with nothing more than a triangle of paid loss data, but if:

IN ADDITION TO PAID LOSS DATA, IF YOU PROVIDE:	THE SYSTEM CAN:
▪ a vector of earned premium data	▪ provide loss ratios by accident period at various percentiles
▪ a vector of ultimate exposure data	▪ simulate based on exposure-adjusted losses rather than only the raw data

Note:
If you have a partial last exposure period, then you should enter the earned premium in the appropriate column, but the ultimate premium and ultimate exposure are for the **full period**. For example, if you have an annual triangle but a 6 month last diagonal, then you should enter the premiums earned for the first 6 months in the earned premium column and the fully annualized premium and/or exposure in the ultimate premium and ultimate exposure columns, respectively. For more details see Section 9.

There are certain limitations that are imposed on the data by the mathematics involved in the model. Specifically:

- the triangle shape must be symmetrical in terms of row and column periods – i.e., it must be annual x annual or quarter x quarter;
  - The system *will* work with triangles that contain a stub period (e.g., annual x annual with most recent diagonal evaluated at 6 months)
  - The system *will* work with triangles where the first development period is different from the rest (e.g., development columns of 6/18/30/42… or 3/15/27/39…)
  - The system *will not* work with truly asymmetrical triangles, such as annual accident periods x quarterly development.
- there must be at least 3 diagonals of data; and
- blank cells are acceptable anywhere in the triangle except on the most recent two diagonals, unless a whole row is blank (i.e., a triangle in run-off is OK)
- Individual negative age-to-age factors are acceptable, and the average for a column can be negative.

- Do not enter "0" values where the values are unknown. The model will treat cells with "0" values as information (that is, no losses occurred in this period), and blank cells as unknown.

STEP 1: ENTER BASIC MODEL DATA

To get started, select one of your segments using the **Segment** drop down box below the **HOME** ribbon. In the **Navigation Pane**, select the DATA | INPUTS | ALL INPUTS collection. Notice that the first three tables in the collection, **Paid Loss**, **Case Loss Reserves** and **Incurred Loss**, are white; these are the data entry tables. You can fill in any two of these tables and the third will change to tan, which means it will be filled automatically and that you cannot enter data here any longer.

1. Enter data for the **Paid Loss** triangle (as illustrated in Image 6-1) and the **Incurred Loss** triangle, if you have that available. You can either type in data or paste it in from another source.

BI > Data > Paid Loss - Cumulative

Paid Loss - Cumulative

Accident Year	12	24	36	48	60	72	84	96	108	120
12-2003	51,215	159,690	231,228	273,263	295,250	305,678	309,938	311,662	312,101	312,343
12-2004	43,948	147,091	217,623	256,035	279,909	292,131	295,313	296,951	297,728	
12-2005	42,294	147,135	213,564	260,585	285,445	296,372	300,928	303,299		
12-2006	45,501	145,223	224,923	272,080	299,332	313,491	317,891			
12-2007	45,236	162,936	252,579	307,971	337,813	352,697				
12-2008	40,081	155,088	243,043	299,985	332,494					
12-2009	48,990	180,269	280,315	343,204						
12-2010	59,239	222,122	334,621							
12-2011	69,972	249,660								
12-2012	74,691									

Image 6-1:
Paid Loss Data Triangle

Note:
You can use the icon to switch between cumulative and incremental or the icon to switch between accident and calendar views, or both, prior to bringing in the data.

2. Also from the ALL INPUTS collection, you can enter **Earned Premium** and **Exposure** data, if you have that available (as illustrated in Image 6-2). Having this additional data allows the model to provide more information; this is especially true of Premium data, which allows the projection of ultimate loss ratios.

BI > Data > Earned Premium - Cumulative

Earned Premium - Cumulative

Accident Year	12	24	36	48	60	72	84	96	108	120
12-2003										398,149
12-2004									427,833	
12-2005								462,411		
12-2006							476,469			
12-2007						480,536				
12-2008					494,954					
12-2009				540,515						
12-2010			612,860							
12-2011		695,342								
12-2012	744,009									

BI > Data > Exposures

Exposures

Accident Year	Exposures
12-2003	1,665
12-2004	1,782
12-2005	1,903
12-2006	1,999
12-2007	2,078
12-2008	2,127
12-2009	2,267
12-2010	2,446
12-2011	2,583
12-2012	2,667

Image 6-2:
Earned Premium and Exposure tables

Note:
The earned premiums are entered in a triangle so that they can be developed in the Deterministic portion of the system.

3. In order to enter the **Ultimate Premium** data (again from the ALL INPUTS collection), you must open the table and click on the Source Data icon in order to get to the Deterministic table used to estimate **Ultimate Premium**. This is illustrated in Image 6-3.

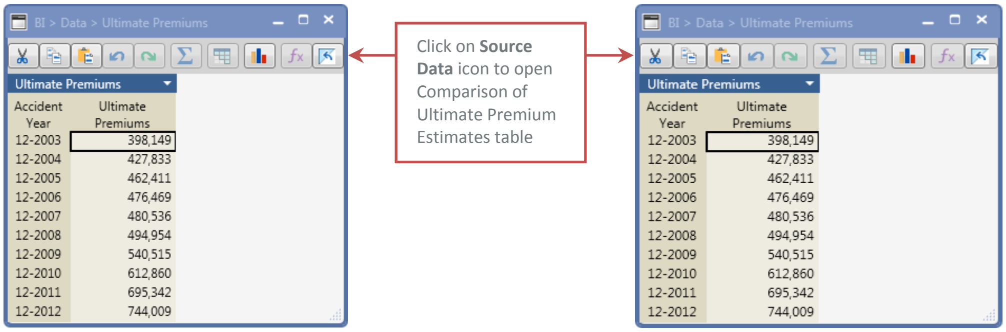


Image 6-3:
Ultimate Premiums and Comparison of Ultimate Premiums Estimates tables

STEP 2: REVIEW / ENTER THE MODEL ASSUMPTIONS

In the **Navigation Pane**, select the STOCHASTIC | MACK BOOTSTRAP | MODEL ASSUMPTIONS collection. The **General** window (shown in Image 6-4) includes model assumptions that will apply to the Mack Bootstrap models.

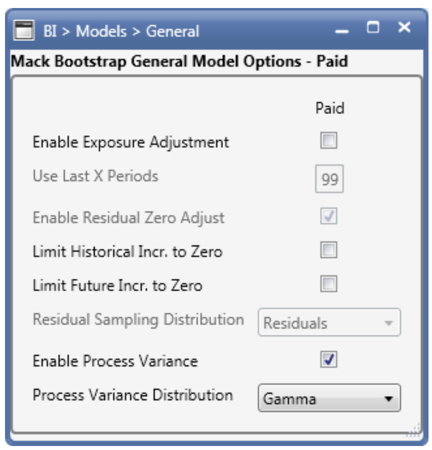


Image 6-4:
General Model Options object

4. **Enable Exposure Adjustment** – If you check this option, the system divides each row in your data triangle by the corresponding row in the **Ultimate Exposures** vector and uses the “exposure-adjusted” data for all further calculations in the model. Values are then multiplied by the **Ultimate Exposures** again after all iteration calculations are complete, returning the modeled results to a “value” basis. This option can be useful when there is a changing exposure volume. By using exposure adjusted data in the model, a better fit could result and the simulation results will be “adjusted” for the relative exposures by period.

5. **Use Last X Periods** – [This is not currently a user-editable option. The system directly follows the underlying theory, using a volume-weighted all-year average for the age-to-age ratios. Additional options for this will be available in an upcoming release of Arius.]

6. **Enable Zero Residual Adjustment** – [This option is not yet active. All the residuals in your model should be independent and identically distributed. Theoretically, they should also sum to zero. The ability to *force* the total to be equal to zero will be available in an upcoming release of the system.]

7. **Limit Historical Incrementals to Zero** – The random nature of the simulation process can result in negative amounts in the incremental results. When this option is checked, the system automatically replaces any negative incremental values in the bootstrap sample triangles with zero. Negative incremental values are certainly acceptable in many situations, for example when modeling paid data that includes salvage amounts; in those cases, negatives are frequently expected, and they should be reflected in the simulated data. Occasionally, however, negative incremental values in the bootstrap sample triangle can also lead to extreme age-to-age factors which, in turn, lead to unrealistic unpaid values for a few iterations. Limiting historical incremental values to zero in that case can be thought of as adding a constraint to the model which will effectively "adjust" these unrealistic iterations.

8. **Limit Future Incrementals to Zero** – When this option is checked, the system automatically replaces any negative incremental values in the lower right portion (after process variance) of the completed rectangle with zero. This provides a similar ability to effectively constrain unrealistic iterations, but it is a separate constraint since there are times when only the future incremental values need to be constrained and not the historical incremental values, and vice versa. For example, negative historical incremental values in the bootstrap sample data can be reasonable when case reserves for later development ages are expected to be redundant while negative future incremental values could be causing unrealistic results for a few iterations.

9. **Residual Sampling Distribution** – The system currently samples the residuals with replacement to create the sample triangles of age-to-age factors.

10. **Enable Process Variance** – A key feature of the Mack Bootstrap model is the simulation of Process Variance. The primary calculation steps in the model focus on parameter risk, but process risk is used to add the final "random fluctuations" to the future incremental values. Checking this option will turn these random fluctuations on, while unchecking turns them off.

    You can get a measure of the effects of process variance versus residual sampling on your results by running the same model multiple times, with this option and/or residual sampling turned on and off, using the same user-input random seed value each time. The differences in the various simulations will help you diagnose the differences between the deterministic estimate and bootstrap mean for each model.

11. **Process Variance Distribution** – If you check the **Enable Process Variance** option (above) then the default approach is to add process variance to the projected future development by simulating from a **Gamma** distribution using each incremental value as the mean and the development factor standard deviation squared as the variance. As an option you can select to use a **Normal** or **Lognormal** distribution instead. For example, if the residuals exhibit little or no skewness then either the normal or lognormal may be a more appropriate distribution for this feature of the data.

If you have not done so, save your file at this point.

STEP 3: EVALUATE YOUR DATA WITH THE MODEL'S DIAGNOSTICS

The standard Mack bootstrap model is essentially based on a traditional chain ladder development method. In order to increase the model's predictive power, the data must be consistent with the assumptions that are inherent in the deterministic form of the model (or the model should be adjusted to be consistent with the data). Specifically:

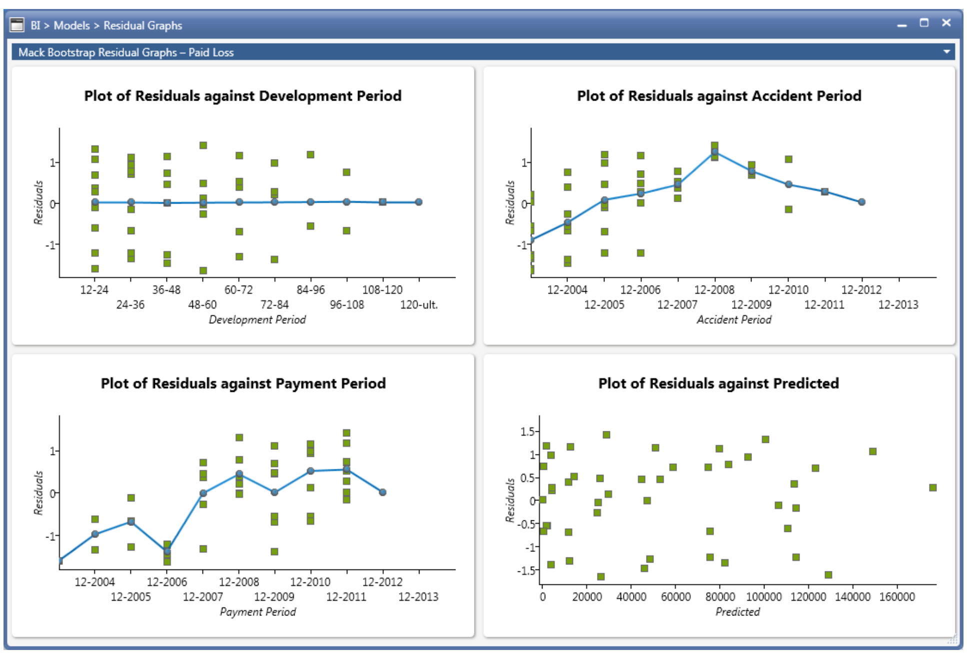
- the expected value of the incremental losses to emerge in the next period is proportional to the total losses emerged to date, by accident year;

- the columns of incremental losses are independent (except for observations in the same accident year); and
- the variance of the next incremental observation is a function of the age and the cumulative losses to date.

The diagnostic output includes a variety of tables and graphs to help the user test these assumptions and then to adjust the model options to improve the statistical fit of the model to the data.

First, from the **Home** ribbon, click on the Run Diagnostics icon to populate the tables and graphs. In an iterative process, you will now want to analyze the diagnostic output, make adjustments to the model options (described above), and then Run Diagnostics again to update the diagnostics results. An additional part of this iterative process is to click on the Run Simulations icon from the **Home** ribbon to run the simulations for the segment you are analyzing. This will allow you to review the model output for the segment, make adjustments to the model options and then either run diagnostics or simulations again until you have optimized the model.

In the **Navigation Pane**, select the Stochastic | Mack Bootstrap | Paid Loss | Diagnostics collection. The Diagnostics collection includes a **Residual Graphs** window (as illustrated in Graph 6-1). These graphics show plots of the residuals (from Image 6-7) against the development, accident, and payment periods, as well as a plot of the residuals vs. the fitted (i.e., predicted) values. These will help you identify trends or other features in your data that may not be completely modeled by the chain ladder approach, thus indicating that the Mack bootstrap predictions from the data may be less than optimal.

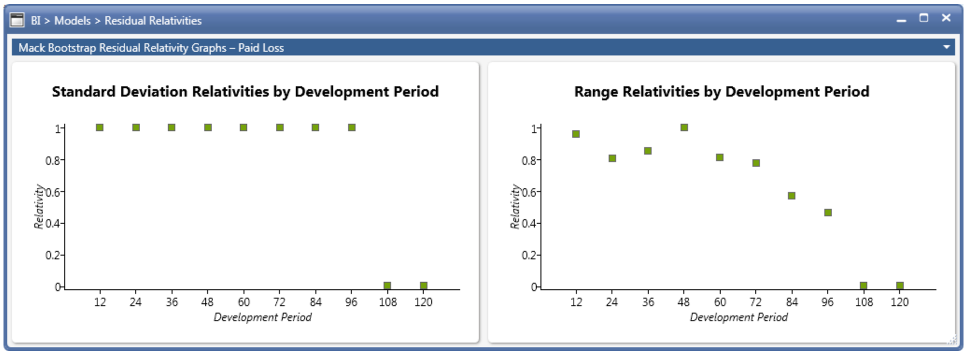


Graph 6-1:
Plots of Residuals

For illustration purposes, we are using the BI data in the ODP_Mack_Hayne.apj file that is included with the system files in the *C:\Users\username\Documents\Milliman\Arius\DemoFiles* directory, where the *username* is your Windows user name.

In the Mack bootstrap model, residuals are resampled with replacement – that is, they are taken from any location in the residual triangle, and placed in another random location to form a sample triangle.

Therefore, the residuals should all be independent, identically distributed random numbers. Unlike the ODP bootstrap model, the standardized residuals are calculated using the standard deviation of each development period, so heteroscedasticity (i.e., different variances) does not occur. Thus, there are no heteroscedasticity adjustment factors for the Mack bootstrap model, but from the **Plot of Residuals against Development Period** (in Graph 6-1) and the **Residual Relativities** table (illustrated in Graph 6-2) you can that the standard deviation relativities are all consistent.



Graph 6-2:
Plots of Residual Relativities

The remaining DIAGNOSTIC collection windows include the **Adjusted Triangle**, **Residuals** and **Age-to-Age Factors** tables. If you have checked the **Enable Exposure Adjustment** option, then the incremental values will be divided by the exposures in each period. If you have stub period data (see Section 8) then the last diagonal will be grossed up to a full period. Otherwise, this will simply be the difference in the cumulative values you entered (as illustrated in Image 6-1).

The next diagnostic output is the **Age-to-Age Factors** table (as illustrated in Image 6-8). These factors are essentially calculated as if from a deterministic analysis, except that exposure adjustments and/or stub period adjustments for the last diagonal are also included. In other words, the factors are calculated after cumulating the adjusted incremental values (as illustrated in Image 6-7).

BI > Models > Age-To-Age Factors

Mack Bootstrap Age-To-Age Factors – Paid Loss

Accident Year	12-24	24-36	36-48	48-60	60-72	72-84	84-96	96-108	108-120	120-ult.
12-2004	3.118	1.448	1.182	1.080	1.035	1.014	1.006	1.001	1.001	
12-2005	3.347	1.480	1.177	1.093	1.044	1.011	1.006	1.003		
12-2006	3.479	1.451	1.220	1.095	1.038	1.015	1.008			
12-2007	3.192	1.549	1.210	1.100	1.047	1.014				
12-2008	3.602	1.550	1.219	1.097	1.044					
12-2009	3.869	1.567	1.234	1.108						
12-2010	3.680	1.555	1.224							
12-2011	3.750	1.506								
12-2012	3.568									
12-2013										
Averages - Paid Loss										
Volume Wtd										
All Periods	3.515	1.514	1.210	1.096	1.042	1.014	1.006	1.002	1.001	

Image 6-5:
Age-to-Age Factor Triangle

In addition to the age-to-age factors, this table also includes the volume weighted **Averages** for the age-to-age factors (including exposure and/or last diagonal adjustments if appropriate).

The next diagnostic output is the standardized residuals shown in the **Residuals** table. These residuals are the basis for the model's simulations. The calculations for the residuals are described in Appendix

A, although the residuals will be based on the data adjusted for exposures and/or stub periods (as illustrated in Image 6-6).

BI > Models > Residuals

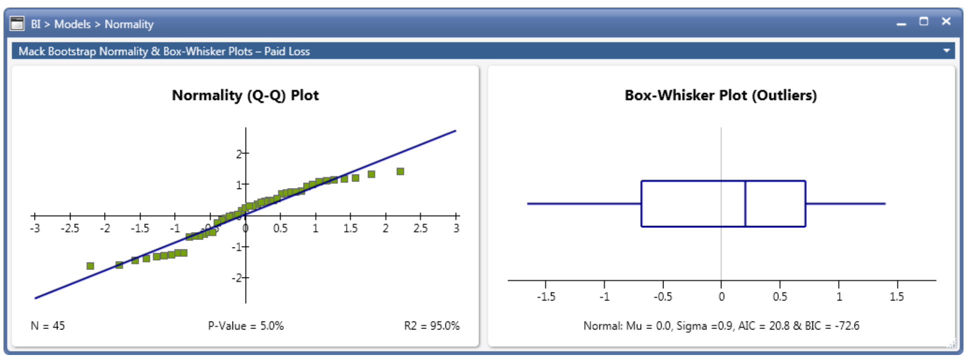
Mack Bootstrap Standardized Residuals – Paid Loss

Accident Year	12-24	24-36	36-48	48-60	60-72	72-84	84-96	96-108	108-120	120-ult.
12-2004	(1.620)	(1.364)	(1.287)	(1.657)	(1.331)	0.203	(0.567)	(0.689)	0.000	
12-2005	(0.629)	(0.680)	(1.481)	(0.277)	0.382	(1.404)	(0.567)	0.725		
12-2006	(0.124)	(1.239)	0.438	(0.054)	(0.707)	0.960	1.165			
12-2007	(1.241)	0.698	(0.021)	0.454	1.151	0.258				
12-2008	0.346	0.767	0.435	0.110	0.506					
12-2009	1.298	1.096	1.124	1.398						
12-2010	0.672	0.912	0.710							
12-2011	1.047	(0.177)								
12-2012	0.265									
12-2013										

Image 6-6:
Standardized Residuals

STEP 3A: IDENTIFY AND EXCLUDE OUTLIERS

The next DIAGNOSTICS window, **Normality**, will help you judge the general improvement in the model as you change the model options. For example, look at Graph 6-3 below which corresponds to the plots shown above in Graph 6-1.



Graph 6-3:
Normality & Box-Whisker Plots

As noted in Appendix B, the changes in the P-Value, R^2, AIC and BIC values under the Normality (Q-Q) Plot and Box-Whisker Plot are a useful guide. You can also review these graphs before and after other changes to the model options.

STEP 3B: DEFINE HOW TO HANDLE TAIL FACTORS

After you a better sense from the diagnostics about how your data fits the requirements of the model, and before you run your model simulations, the last consideration that is common to all deterministic methods is the potential inclusion of a tail factor. In the **Navigation Pane**, select the STOCHASTIC | ODP BOOTSTRAP | PAID LOSS | DIAGNOSTICS collection. You can then open the **Tail Factor** window as illustrated in Image 6-7.

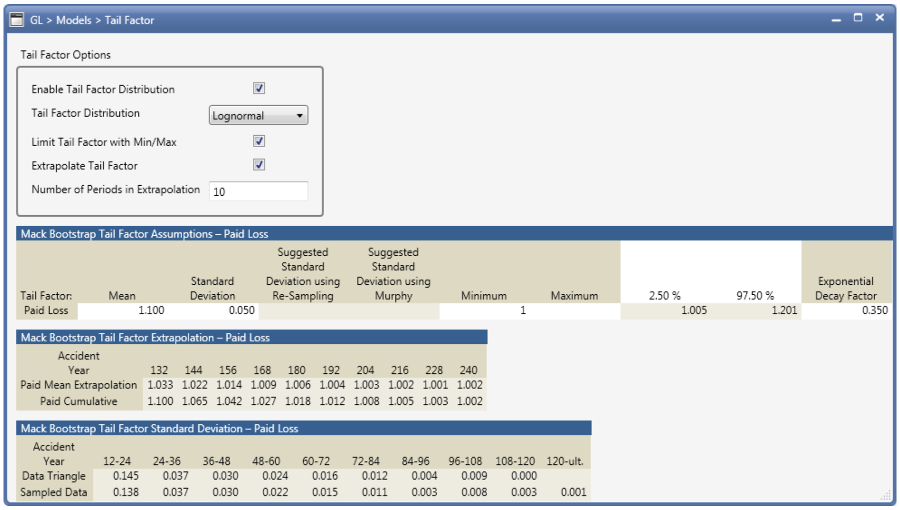


Image 6-7:
Tail Factor assumptions

12. **Enable Tail Factor Distribution** – The system provides two options:
    - Check this option and the system will select a tail for each iteration based on your supplied mean, standard deviation and distribution type; or
    - Uncheck this option and the system will use your **Mean** tail factor amount in each iteration.
13. **Tail Factor Distribution** – If you check **Enable Tail Factor Distribution** (above), you can select from a lognormal or normal distribution from which to simulate the tail factor.
14. **Limit Tail Factor with Min/Max** – If you choose to randomly select a tail factor, you can also provide specific minimums and/or maximum amounts for the model to use. If you provide min/max levels, and check this option, any random amounts outside these levels will be limited to these levels. An example might be to limit factors to a minimum of 1.00.
15. **Extrapolate Tail Factor** – One of the outputs of the simulation is an estimate of the cash flows resulting from the estimated unpaid amounts. These can be presented two ways:
    - If you check this option, the future payments related to the tail will be extended out beyond the development of the triangle itself; or
    - If you uncheck this option, the future payments related to the tail are all accumulated into one final period in the Estimated Cash Flow exhibit.

    The future cash flows related to the tail are extrapolated into the future based on the **Number of Periods in Extrapolation** field and using the value entered into the **Exponential Decay Factor** field. This is important if you want a meaningful **Estimated Cash Flow** table and will also affect the discounted results.
16. **Number of Periods in Extrapolation** – This is an estimate of how many future periods are assumed to be in the tail factor, used as noted above to extrapolate the Cash Flows to future periods.

    The incurred selection will effectively be converted to the number of periods for the paid selection. For example, if the paid model extrapolates 5 years and the incurred model is set to not extrapolate, the simulated paid values will be adjusted to sum to the same ultimate values as the incurred values and the extrapolation for an additional 5 years will be included. Alternatively, if the incurred model extrapolates 5 years and the paid model is set to not extrapolate, the adjustment of the paid simulations will include zeroes beyond the end of the triangle since no payment pattern is simulated beyond the end of the triangle.

17. **Tail Factor** – The parameterization of the tail factor has several related parts (as illustrated in Image 6-7):
    - **Mean** – Enter your best estimate of a tail factor.
    - **Standard Deviation** – If you have checked the **Enable Tail Factor Distribution** option, enter an estimate of the standard deviation of the tail factor.
    - **Suggested Std Deviation: Tail Factor** – After you RUN DIAGNOSTICS**,** suggested parameters for the standard deviation of the tail factor will be shown here based on two different methods:
        - **Resampling** – this method extends the residual resampling that is used for the body of the triangles into the tail of the triangle. 10,000 resampling iterations are done and the implied standard deviation is shown here.
        - **Murphy** – this method uses your a priori loss ratio input (from Model Assumptions | Bornhuetter-Ferguson) and Ultimate Premiums (or Exposures) and calculates the selected ultimate loss for each accident year. The implied tail factor is derived from the difference between the chain ladder ultimate, excluding the tail factor, and the user's selected ultimate for each year. The standard deviation from this set of implied tail factors is shown here.
    - **Min / Max** – If you have checked the **Limit Tail Factor with Min/Max** option, select a minimum and/or maximum for your tail factor.
    - **Percentile** – The 95% confidence interval for the tail factor distribution is shown. Similar to the confidence interval for the Bornhuetter-Ferguson a prior assumption, you can change the percentiles in the heading to see a different interval.
    - **Exponential Decay Factor** – When extrapolation is turned on, one minus this factor is multiplied times the tail factor for each period in the extrapolation. Since the tail factor is a factor to ultimate, each successive factor is a new factor **to ultimate one period** later and dividing each factor by the next factor results in incremental age-to-age factors for the tail (as illustrated in Image 6-8, with a Tail Factor = 1.1, Decay Factor = 35.0% and Number of Years = 10).

Tail Factor Extrapolation					
Period	TF	TF + 1	TF + 2	…	TF + n-1
Extrapolation	TF / TF + 1	TF + 1 / TF + 2	TF + 2 / TF + 3	…	TF + n-1
Cumulative	TF(Mean)	1 + [("TF" - 1) x (1 - Decay)]	1 + [("TF + 1" - 1) x (1 - Decay)]	…	1 + [("TF + n-2" - 1) x (1 - Decay)]
Period	TF	TF + 1	TF + 2	…	TF + 10-1
Extrapolation	1.1000 / 1.0650	1.0650 / 1.0423	1.0423 / 1.0275	…	1.0013
Cumulative	1.1000	1 + [(1.1000 - 1) x (1 - 0.35)]	1 + [(1.0650 - 1) x (1 - 0.35)]	…	1 + [(1.0021 - 1) x (1 - 0.35)]
Period	TF	TF + 1	TF + 2	…	TF + 9
Extrapolation	1.0329	1.0218	1.0144	…	1.0013
Cumulative	1.1000	1.0650	1.0423	…	1.0013

Image 6-8:
Tail Factor Extrapolation Example with Formulas

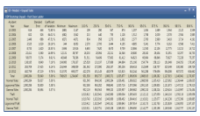
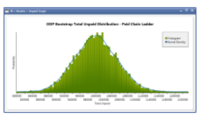
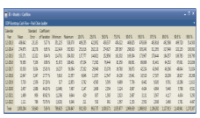
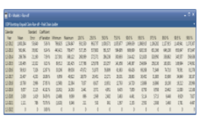
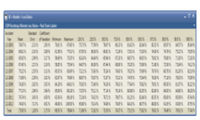
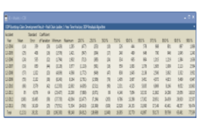
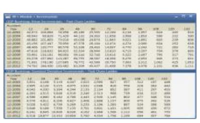
18. **Tail Factor Standard Deviations** – In addition to the average age-to-age factors (illustrated in Image 5-11 below), the standard deviations of the Age-to-Age factors are shown in the Data Triangle row (illustrated in Image 6-7 above). After you run the simulations, the **Sampled Data**

row of this table is also shown, which is based on all of the simulated data. Both of the standard deviation rows can be used to help you select a standard deviation for the tail factors.

In addition to the model diagnostics described above, the results output also has diagnostic features. Thus, running the model using RUN SIMULATIONS, reviewing the model output and adjusting model parameters and assumptions is part of the diagnostic process. Reviewing the model output is discussed in more detail in the remainder of this Section.

SUMMARY OF OUTPUT

The results for each model are shown in their own collection. For example, in the **Navigation Pane**, select the STOCHASTIC | MACK BOOTSTRAP | PAID LOSS | ULTIMATE collection to view all of the simulation results for the Mack Bootstrap model. For the TIME HORIZON results, there is an additional table.

	Estimated Unpaid	Mean, Standard Error, Coefficient of Variation, Min, Max and Percentiles. Total Distributions and TVaRs.
	Total Unpaid Distribution	Histogram and kernel density of total unpaid.
	Estimated Cash Flow	Future calendar period payments.
	Estimated Run-off	Total unpaid as future calendar periods are removed.
	Estimated Loss Ratios	Time zero to ultimate loss ratios.
	Estimated CDR	Claim Development Results (for Time Horizon options only).
	Incremental Values	Mean and standard deviation values for each incremental cell, historical and future.

STEP 4: EVALUATE THE OUTPUT FOR EACH MODEL

After the model diagnostics have been set up and reviewed, the next step in the evaluation of each model is to use RUN SIMULATIONS to run the simulations for the segment you are analyzing. To illustrate the diagnostic elements of the simulation output we will review the results for the Mack bootstrap model.

Estimated Unpaid Results

The first diagnostic element of the **Unpaid Table** (illustrated in Table 6-1) can be seen by reviewing the Standard Error and Coefficient of Variation columns. As general rules, the standard error should go up as you move from the oldest years to the most recent years and the standard error for the total of all years should be larger than any individual year. In Table 6-1, the standard errors follow these general rules. For the coefficients of variation, they should go down when moving from the oldest years to the more recent years and the coefficient of variation for all years combined should be less than for any individual year.[40] Except for the 2013 year, the coefficients of variation in Table 6-1 also follow the general rules.

BI > Models > Unpaid Table

Mack Bootstrap Unpaid – Paid Chain Ladder

Accident Year	Mean	Standard Error	Coefficient of Variation	Minimum	Maximum	10.0 %	25.0 %	50.0 %	75.0 %	90.0 %	95.0 %	97.5 %	99.0 %	99.5 %	99.9 %
12-2004	617	350	56.75 %	(645)	1,963	172	379	617	853	1,060	1,194	1,299	1,426	1,510	1,711
12-2005	824	403	48.97 %	(865)	2,400	313	550	823	1,096	1,347	1,487	1,613	1,748	1,876	2,100
12-2006	1,442	517	35.84 %	(699)	3,374	781	1,097	1,440	1,788	2,105	2,296	2,454	2,644	2,757	2,955
12-2007	3,522	713	20.23 %	938	6,309	2,608	3,038	3,530	4,009	4,421	4,682	4,900	5,151	5,332	5,740
12-2008	8,754	1,030	11.77 %	4,965	12,593	7,432	8,059	8,759	9,445	10,063	10,442	10,777	11,119	11,374	12,085
12-2009	22,542	1,987	8.81 %	15,614	30,678	19,989	21,203	22,539	23,887	25,064	25,778	26,382	27,179	27,700	28,667
12-2010	58,408	3,903	6.68 %	44,305	73,681	53,477	55,750	58,332	61,027	63,395	64,889	66,231	67,833	68,768	70,718
12-2011	139,300	8,962	6.43 %	105,369	170,081	127,763	133,243	139,298	145,315	150,776	154,251	157,033	160,195	162,428	167,254
12-2012	285,484	17,572	6.16 %	214,755	364,249	262,802	273,471	285,423	297,164	307,872	314,839	320,331	327,068	331,925	341,526
12-2013	487,652	38,540	7.90 %	332,438	636,439	438,535	461,586	487,120	512,864	538,262	552,371	564,840	580,351	590,143	611,330
Total	1,008,544	45,983	4.56 %	825,201	1,183,538	949,414	977,268	1,007,963	1,038,578	1,068,444	1,085,737	1,099,580	1,118,220	1,131,745	1,154,388
Normal %iles	1,008,544	45,981	4.56 %			949,617	977,530	1,008,544	1,039,557	1,067,470	1,084,175	1,098,664	1,115,511	1,126,982	1,150,635
Lognormal %iles	1,008,544	46,031	4.56 %			950,285	976,969	1,007,496	1,038,976	1,068,150	1,086,000	1,101,723	1,120,292	1,133,114	1,160,018
Gamma %iles	1,008,544	45,982	4.56 %			950,082	977,160	1,007,845	1,039,166	1,067,903	1,085,351	1,100,636	1,118,587	1,130,921	1,156,634
TVaR						1,017,324	1,027,868	1,045,215	1,067,644	1,090,803	1,105,282	1,118,651	1,135,225	1,146,368	1,164,885
Normal TVaR						1,017,510	1,028,026	1,045,231	1,066,990	1,089,239	1,103,389	1,116,038	1,131,092	1,141,518	1,163,365
Lognormal TVaR						1,017,255	1,027,734	1,045,240	1,067,912	1,091,663	1,107,060	1,121,014	1,137,859	1,149,676	1,174,844
Gamma TVaR						1,017,334	1,027,822	1,045,222	1,067,576	1,090,798	1,105,750	1,119,236	1,135,431	1,146,739	1,170,680

Table 6-1: Estimated Unpaid Model Output

Note: For policy period data or incomplete accident period data, the unpaid data in the last row(s) will be reduced to only include earned exposures.

The reason the standard errors (value scale) tend to go up is that they tend to follow the magnitude of the mean or expected value estimates. The reason the coefficients of variation (percent scale) tend to go down has more to do with the independence in the incremental claim payment stream. For the oldest accident year, there is typically only one (or a few) incremental payment(s) left so the variability of that payment(s) is (almost) fully reflected in the coefficient. For the most current accident year, the "up and down" variations in the future incremental payment stream can offset each other thus causing the total variation to be a function of the correlation between each incremental payment for that accident year (i.e., the incremental payments are assumed independent).

Note: Caution should be exercised in the interpretation and adjustments for increases in the coefficient of variation in recent years. While keeping the theory in mind is appropriate, this must be balanced with the need to keep from underestimating the uncertainty of the more recent years.

The coefficient of variation rules noted above are a reflection of the step 7's described in Section 3 (and Appendix A), in the sense that they describe the process variance in the model. While the coefficients of variation should go down, if they do start going back up in the most recent year(s), as illustrated in Table 6-1 for 2013, then this could be the result of the following issues:

1. The parameter uncertainty tends to increase when moving from the oldest years to the more recent years as more and more parameters are used in the model. In the most recent year(s), the parameter uncertainty could be "overpowering" the process uncertainty, causing the coefficient of variation to start going back up. At the very least, the increasing parameter uncertainty will cause the rate of decrease in the coefficient of variation to slow down.

40 These standard error and coefficient of variation rules are based on the independence of the incremental process risk and assume that the underlying exposures are relatively stable from year to year – i.e., no radical changes. In practice, random changes do occur from one year to the next which could cause the actual standard errors to deviate from these rules somewhat. In other words, these rules should generally hold true, but are not considered hard and fast rules in every case. Strictly speaking, the total all years rules assume that the individual years are not positively correlated.

2. If the increase in the most recent year(s) is significant, then this could indicate that the model is overestimating the uncertainty in those years. If this is the case, then an adjustment to the model parameters may be needed (e.g., limit incrementals to zero, etc.).

While we mentioned the rules for the standard error and coefficient of variation for the total of all years, it is also worth noting that in addition to the correlation (independence) within each accident year the total of all years also includes the impact of the correlation (independence) between accident years. In essence, when one or more accident years are "bad" we do not expect all accident years to be "bad." To see this impact, you can add the accident year standard errors and note that they will not sum to the standard error for all years combined.[41]

The next diagnostic element in the **Unpaid Table** is the **Minimum** and **Maximum** columns. In these columns, the smallest and largest values, respectively, from among all iterations of the simulation are displayed. These values can be reviewed judgmentally to make sure that they are not outside the "realm of possibility." If they do seem a bit unrealistic then they could indicate the need to review the model options. For example, the presence of negative numbers might lead to changing one or both of the options which limit incremental values to zero. Sometimes "extreme" outliers in the results will show up in these columns and may also distort the histogram (discussed later in this section).

Risk Measures

Also included in Table 6-1, notice that there are three rows of "Percentile" numbers and then four rows of TVaR numbers at the bottom of these tables under each of the percentile columns. For the three "Percentile" rows, the normal, lognormal and gamma distributions, respectively, have been fit to the Total unpaid claim distribution. The fitted mean, standard deviation and selected percentiles are shown under the Mean, Standard Error and Percentile columns, respectively, so that the smoothed results can be used to judge the quality of fit for each distribution or other purposes such as parameterizing a DFA model or using smoothed results in the tail of the distribution.

The Tail Value at Risk (TVaR)[42] is the average of all of the simulated values equal to or greater than the percentile value. For example, in Table 6-1 the 75th percentile value for the total unpaid for all accident years combined is 1,038,578 and the average of all simulated values that are greater than or equal to 1,038,578 is 1,067,644. The "Normal TVAR," "Lognormal TVaR" and "Gamma TVaR" rows are calculated the same way, except that instead of using the actual simulated values from the model the respective fitted distributions are used in the calculations.

To interpret the TVaR numbers, the question we are trying to answer with a TVaR number is "if the actual outcome does exceed the X percentile value, on average how much might it exceed that value by?" This is an important question related to risk based capital calculations and other technical aspects of enterprise risk management, although a more complete discussion is beyond the scope of this manual. It is worth noting, however, that the purpose of the normal, lognormal and gamma TVaR numbers is to provide "smoothed" values in the sense that some of the random noise is kept from distorting the calculations.

Estimated Cash Flow Results

In addition to the results by accident year, we can also review the model output by calendar year (or by future diagonal) in the **Cash Flow** table as illustrated in Table 6-2. Comparing Table 6-2 to 6-1, notice that the Total row is identical since the total is the same whether you add the parts horizontally or

41 Likewise, the minimum, maximum and each of the percentile columns will not sum to the total for all years combined. In contrast, adding the mean values for each accident year will sum to the total for all years combined.

42 The Tail Value at Risk is sometimes referred to as the Conditional Tail Expectation.

diagonally. Similar diagnostic issues can be reviewed in this table, except that the relative values of the standard errors and coefficients of variation move in the opposite direction for calendar years compared to accident years. This should make intuitive sense as the “final” payments projected the farthest out into the future should be the smallest yet relatively most uncertain.

BI > Models > Cashflow

Mack Bootstrap Cash Flow – Paid Chain Ladder

Calendar Year	Mean	Standard Error	Coefficient of Variation	Minimum	Maximum	10.0 %	25.0 %	50.0 %	75.0 %	90.0 %	95.0 %	97.5 %	99.0 %	99.5 %	99.9 %
12-2014	441,139	20,589	4.67 %	366,055	515,031	414,547	427,121	440,949	454,946	467,518	475,262	482,063	489,613	494,784	502,910
12-2015	277,401	16,168	5.83 %	225,035	346,447	256,712	266,396	277,240	288,276	298,059	304,247	309,462	316,261	320,200	332,067
12-2016	155,088	10,369	6.69 %	118,564	198,981	142,023	148,007	154,929	162,006	168,356	172,324	176,203	179,882	182,635	187,702
12-2017	77,529	5,565	7.18 %	54,144	100,951	70,460	73,828	77,463	81,236	84,695	86,760	88,675	90,858	92,498	95,359
12-2018	33,835	2,983	8.82 %	23,352	46,388	30,044	31,783	33,795	35,835	37,666	38,776	39,718	40,843	41,870	43,339
12-2019	12,751	1,386	10.87 %	7,742	17,774	10,990	11,814	12,748	13,675	14,543	15,068	15,486	16,034	16,519	17,089
12-2020	5,744	1,007	17.53 %	2,000	9,520	4,444	5,067	5,749	6,411	7,035	7,415	7,729	8,096	8,313	8,839
12-2021	2,462	772	31.36 %	(444)	5,261	1,477	1,939	2,461	2,994	3,450	3,732	3,967	4,241	4,404	4,880
12-2022	1,489	679	45.59 %	(1,039)	4,179	613	1,037	1,488	1,938	2,371	2,613	2,804	3,040	3,221	3,601
12-2023	1,107	606	54.68 %	(1,305)	3,545	333	698	1,100	1,513	1,890	2,123	2,319	2,521	2,696	3,111
Total	1,008,544	45,983	4.56 %	825,201	1,183,538	949,414	977,268	1,007,963	1,038,578	1,068,444	1,085,737	1,099,580	1,118,220	1,131,745	1,154,388

Table 6-2: Estimated Cash Flow Model Output

Estimated Unpaid Claim Runoff Results

Another report similar to the **Cash Flow** table is the **Run-off** table. Rather than looking at individual diagonal results, the **Run-off** table starts with the total unpaid results and then looks at how the total unpaid will decrease over time as successive diagonals are removed, as illustrated in Table 6-3. Comparing Table 6-3 to 6-1 & 6-2, notice that the first row of Table 6-3 is identical to the Total rows in Tables 6-1 and 6-2. Each successive row in Table 6-3 is then the total of the remaining diagonals.

BI > Models > Run-off

Mack Bootstrap Unpaid Claim Run-off – Paid Chain Ladder

Calendar Year	Mean	Standard Error	Coefficient of Variation	Minimum	Maximum	10.0 %	25.0 %	50.0 %	75.0 %	90.0 %	95.0 %	97.5 %	99.0 %	99.5 %	99.9 %
12-2013	1,008,544	45,983	4.56 %	825,201	1,183,538	949,414	977,268	1,007,963	1,038,578	1,068,444	1,085,737	1,099,580	1,118,220	1,131,745	1,154,388
12-2014	567,405	29,830	5.26 %	452,168	686,420	529,569	547,361	567,076	587,212	605,803	616,696	626,624	639,863	648,631	666,753
12-2015	290,004	17,037	5.87 %	220,463	356,955	268,304	278,308	290,010	301,233	311,962	318,252	324,328	330,972	335,588	343,344
12-2016	134,916	9,061	6.72 %	94,450	167,812	123,325	128,688	134,847	140,907	146,489	149,941	153,158	156,764	159,140	165,175
12-2017	57,387	5,113	8.91 %	38,845	76,184	50,873	53,925	57,355	60,764	63,816	65,737	67,650	69,881	71,041	73,789
12-2018	23,552	3,248	13.79 %	9,510	34,882	19,425	21,336	23,515	25,761	27,662	28,906	29,977	31,209	32,011	33,607
12-2019	10,802	2,433	22.52 %	522	19,884	7,691	9,159	10,803	12,455	13,916	14,811	15,621	16,441	17,230	18,409
12-2020	5,058	1,785	35.29 %	(2,788)	11,189	2,774	3,852	5,060	6,242	7,346	7,996	8,598	9,305	9,739	10,607
12-2021	2,596	1,193	45.95 %	(2,344)	7,338	1,067	1,804	2,589	3,387	4,140	4,596	4,956	5,388	5,702	6,316
12-2022	1,107	606	54.68 %	(1,305)	3,545	333	698	1,100	1,513	1,890	2,123	2,319	2,521	2,696	3,111
12-2023	0	0	0.00 %	0	0	0	0	0	0	0	0	0	0	0	0

Table 6-3: Estimated Unpaid Claim Run-Off Model Output

Estimated Ultimate Loss Ratio Results

The next collection table shows the ultimate **Loss Ratios** by accident year as illustrated in Table 6-4. If there are no earned premiums or ultimate premiums input into the model, then this table will not be filled in since the model cannot calculate a loss ratio without the premium information.[43]

43 Earned premiums are used as the denominator of the loss ratios. However, if earned premiums are not input then earned premiums are estimated from the ultimate premiums.

BI > Models > Loss Ratios

Mack Bootstrap Ultimate Loss Ratios – Paid Chain Ladder

Accident Year	Mean	Standard Error	Coefficient of Variation	Minimum	Maximum	10.0 %	25.0 %	50.0 %	75.0 %	90.0 %	95.0 %	97.5 %	99.0 %	99.5 %	99.9 %
12-2004	78.60 %	0.09 %	0.11 %	78.29 %	78.94 %	78.49 %	78.54 %	78.60 %	78.66 %	78.72 %	78.75 %	78.78 %	78.81 %	78.83 %	78.88 %
12-2005	69.78 %	0.09 %	0.14 %	69.39 %	70.15 %	69.66 %	69.72 %	69.78 %	69.85 %	69.90 %	69.94 %	69.97 %	70.00 %	70.03 %	70.08 %
12-2006	65.90 %	0.11 %	0.17 %	65.44 %	66.32 %	65.76 %	65.83 %	65.90 %	65.98 %	66.05 %	66.09 %	66.12 %	66.16 %	66.19 %	66.23 %
12-2007	67.46 %	0.15 %	0.22 %	66.92 %	68.04 %	67.27 %	67.36 %	67.46 %	67.56 %	67.65 %	67.70 %	67.75 %	67.80 %	67.84 %	67.92 %
12-2008	75.22 %	0.21 %	0.28 %	74.43 %	76.02 %	74.94 %	75.07 %	75.22 %	75.36 %	75.49 %	75.57 %	75.64 %	75.71 %	75.76 %	75.91 %
12-2009	71.73 %	0.40 %	0.56 %	70.33 %	73.37 %	71.22 %	71.46 %	71.73 %	72.00 %	72.24 %	72.39 %	72.51 %	72.67 %	72.77 %	72.97 %
12-2010	74.30 %	0.72 %	0.97 %	71.69 %	77.13 %	73.39 %	73.81 %	74.29 %	74.79 %	75.22 %	75.50 %	75.75 %	76.05 %	76.22 %	76.58 %
12-2011	77.33 %	1.46 %	1.89 %	71.79 %	82.35 %	75.45 %	76.34 %	77.33 %	78.31 %	79.20 %	79.77 %	80.22 %	80.74 %	81.10 %	81.89 %
12-2012	76.96 %	2.53 %	3.28 %	66.79 %	88.29 %	73.70 %	75.23 %	76.95 %	78.64 %	80.18 %	81.18 %	81.97 %	82.94 %	83.64 %	85.02 %
12-2013	75.58 %	5.18 %	6.85 %	54.72 %	95.58 %	68.98 %	72.08 %	75.51 %	78.97 %	82.39 %	84.28 %	85.96 %	88.04 %	89.36 %	92.21 %
Total	73.64 %	0.86 %	1.17 %	70.20 %	76.92 %	72.53 %	73.05 %	73.63 %	74.20 %	74.76 %	75.09 %	75.35 %	75.69 %	75.95 %	76.37 %

Table 6-4: Estimated Loss Ratio Model Output

Note: For policy period data or incomplete accident period data, the unpaid data in the last row(s) will be reduced to only include earned exposures. However, since the earned exposures are divided by the Earned Premium to calculate the loss ratios we have a match of losses to premium.

Unlike the **Loss Ratios** table for the ODP bootstrap model, the loss ratios for the Mack bootstrap model are calculated from only the simulated values beyond the end of the triangles, since the data in the triangle is not simulated. Because we are using only future simulated values, the standard errors and coefficients of variation in Table 6-4 increase when comparing the latest to the oldest years for the reasons cited earlier for the **Unpaid Table**.

Estimated Incremental Results

The next collection table is designed to help you take a deeper look at the simulations and to understand the reasons for increases in the coefficients of variation (illustrated in Tables 6-1 and 6-4). They show the mean and standard deviations, respectively, by accident year by incremental period. As illustrated in Table 6-5, both the Mean and Standard Deviation **Incrementals** can be reviewed down each column or across each row to look for any irregularities in the expected patterns.

As you can see by looking down the 24 and 36 month columns in Table 6-5, it appears as though there might be "too much" variability in future incremental values for years 2012 and 2013 – i.e., those "future" values do not appear consistent with the values in the prior years. This does not imply that the "historical" values are correct and that the "future" values are overstated, just that they are not always consistent. These inconsistencies appear to be impacting both the unpaid and loss ratio results for 2012 and 2013.

As noted above, the Mack bootstrap model only uses simulation for future incremental values and as such the historical triangle is unchanged for each iteration. Thus, the standard deviations for each historical incremental is zero as illustrated in Table 6-5.

BI > Models > Incrementals

Mack Bootstrap Mean Incrementals – Paid Chain Ladder

Accident Year	12	24	36	48	60	72	84	96	108	120	132
12-2004	51,215	108,476	71,537	42,035	21,987	10,427	4,260	1,724	438	243	617
12-2005	43,948	103,143	70,532	38,412	23,874	12,222	3,182	1,638	777	233	591
12-2006	42,294	104,841	66,430	47,021	24,860	10,927	4,556	2,370	606	234	602
12-2007	45,501	99,722	79,700	47,157	27,252	14,159	4,399	2,006	636	247	634
12-2008	45,236	117,700	89,643	55,393	29,841	14,885	4,785	2,258	721	281	709
12-2009	40,081	115,007	87,955	56,942	32,509	13,916	4,722	2,220	705	276	703
12-2010	48,990	131,279	100,046	62,889	32,926	15,740	5,342	2,511	790	306	794
12-2011	59,239	162,883	112,499	70,369	38,882	18,553	6,293	2,959	944	367	934
12-2012	69,972	179,688	128,218	79,420	43,922	20,962	7,086	3,340	1,066	415	1,054
12-2013	74,691	187,464	134,918	83,481	46,143	22,015	7,458	3,517	1,113	435	1,107

Mack Bootstrap Standard Deviation Incrementals – Paid Chain Ladder

Accident Year	12	24	36	48	60	72	84	96	108	120	132
12-2004	0	0	0	0	0	0	0	0	0	0	350
12-2005	0	0	0	0	0	0	0	0	0	222	339
12-2006	0	0	0	0	0	0	0	0	309	226	343
12-2007	0	0	0	0	0	0	0	470	318	232	363
12-2008	0	0	0	0	0	0	681	509	343	253	403
12-2009	0	0	0	0	0	1,653	674	501	340	251	397
12-2010	0	0	0	0	3,090	1,770	725	537	375	276	443
12-2011	0	0	0	6,657	3,470	2,001	818	606	408	314	515
12-2012	0	0	10,339	7,429	3,924	2,200	895	668	446	338	575
12-2013	0	15,906	13,387	9,168	4,907	2,676	1,028	721	479	358	606

Table 6-5:
Estimated Increments by Accident Year by Development Period

Estimated Claim Development Result

If the Mack bootstrap Time Horizon model is selected, either from the Default Model Selection tab of the **Model Options** dialog (as illustrated in 4-14) or from the **Choose Models** dialog (as illustrated in 4-16), from the **Navigation Pane** you can select the STOCHASTIC | MACK BOOTSTRAP | PAID LOSS | TIME HORIZON collection. In addition to all of the windows discussed above for the ULTIMATE collection, the TIME HORIZON collection includes the **CDR** or Claim Development Result table (as illustrated in Table 6-6).

BI > Models > CDR

Mack Bootstrap Claim Development Result – Paid Chain Ladder, 1-Year Time Horizon Algorithm

Accident Year	Mean	Standard Error	Coefficient of Variation	Minimum	Maximum	10.0 %	25.0 %	50.0 %	75.0 %	90.0 %	95.0 %
12-2004	0	350	(607,166,576,969,246)	(1,261)	1,347	(444)	(237)	0	237	444	577
12-2005	(3)	398	(130)	(1,444)	1,625	(511)	(269)	(3)	271	507	651
12-2006	2	472	275	(1,824)	1,568	(604)	(312)	2	324	597	777
12-2007	6	614	106	(2,554)	2,074	(783)	(411)	17	423	787	986
12-2008	3	824	290	(3,519)	2,850	(1,052)	(555)	14	559	1,044	1,351
12-2009	(22)	1,770	(79)	(7,011)	6,463	(2,296)	(1,212)	(14)	1,171	2,268	2,866
12-2010	(21)	3,351	(158)	(13,442)	14,820	(4,321)	(2,330)	(48)	2,230	4,299	5,587
12-2011	(7)	7,861	(1,085)	(29,913)	30,066	(10,088)	(5,380)	49	5,314	10,077	12,914
12-2012	49	14,754	303	(56,655)	56,131	(18,919)	(9,959)	97	9,919	18,862	24,327
12-2013	(207)	34,298	(165)	(129,095)	130,244	(44,308)	(23,397)	(952)	22,634	44,499	56,414
Total	(202)	40,782	(202)	(150,847)	150,432	(53,173)	(27,576)	(621)	27,397	52,782	67,519

Table 6-6:
Estimated Claim Development Result Output

The output for this table is calculated by subtracting the mean of the **Unpaid Table** from the ULTIMATE collection from each of the iterations used to calculate the **Unpaid Table** from the TIME HORIZON collection. For example, the **Unpaid Table** for the TIME HORIZON collection is shown in Table 6-7. Subtracting the Total **Mean** from Table 6-1 of 1,008,544 from the Total **Mean** from Table 6-7 of 1,008,341, results in the Total **Mean** for Table 6-6 of (202). The CDR is used to calculate the required capital for Solvency II regulations in Europe.

BI > Models > Unpaid Table

Mack Bootstrap Unpaid – Paid Chain Ladder, 1-Year Time Horizon Algorithm

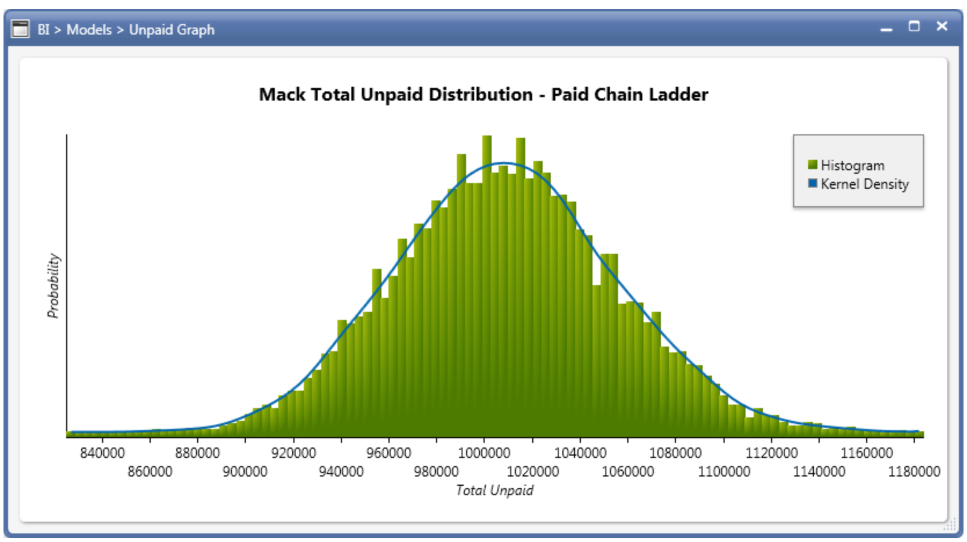
Accident Year	Mean	Standard Error	Coefficient of Variation	Minimum	Maximum	10.0 %	25.0 %	50.0 %	75.0 %	90.0 %	95.0 %	97.5 %	99.0 %	99.5 %	99.9 %
12-2004	617	350	56.75 %	(645)	1,963	172	379	617	853	1,060	1,194	1,299	1,426	1,510	1,711
12-2005	821	398	48.42 %	(620)	2,449	313	555	821	1,095	1,331	1,475	1,600	1,741	1,800	2,040
12-2006	1,443	472	32.68 %	(382)	3,010	838	1,129	1,444	1,766	2,039	2,219	2,383	2,539	2,637	2,836
12-2007	3,528	614	17.40 %	968	5,596	2,739	3,111	3,539	3,945	4,309	4,508	4,713	4,947	5,123	5,388
12-2008	8,757	824	9.41 %	5,235	11,604	7,702	8,199	8,767	9,312	9,798	10,105	10,356	10,697	10,849	11,244
12-2009	22,519	1,770	7.86 %	15,531	29,005	20,246	21,330	22,528	23,712	24,810	25,408	25,902	26,478	26,990	28,008
12-2010	58,387	3,351	5.74 %	44,966	73,228	54,087	56,078	58,361	60,638	62,707	63,996	65,000	66,123	66,989	68,786
12-2011	139,293	7,861	5.64 %	109,387	169,365	129,212	133,920	139,349	144,614	149,377	152,214	154,560	157,550	159,183	163,745
12-2012	285,532	14,754	5.17 %	228,828	341,614	266,565	275,525	285,581	295,403	304,346	309,811	314,620	320,874	324,671	331,505
12-2013	487,445	34,298	7.04 %	358,557	617,895	443,344	464,255	486,700	510,286	532,151	544,066	555,324	567,992	578,478	598,435
Total	1,008,341	40,782	4.04 %	857,697	1,158,976	955,371	980,967	1,007,923	1,035,940	1,061,326	1,076,062	1,090,034	1,105,179	1,115,686	1,132,321
Normal %iles	1,008,341	40,780	4.04 %			956,079	980,836	1,008,341	1,035,847	1,060,603	1,075,419	1,088,269	1,103,210	1,113,384	1,134,362
Lognormal %iles	1,008,342	40,829	4.05 %			956,588	980,384	1,007,516	1,035,400	1,061,156	1,076,875	1,090,697	1,106,992	1,118,227	1,141,753
Gamma %iles	1,008,341	40,791	4.05 %			956,432	980,537	1,007,791	1,035,546	1,060,957	1,076,361	1,089,841	1,105,654	1,116,508	1,139,110
TVaR						1,016,181	1,025,535	1,040,907	1,060,690	1,080,751	1,093,720	1,105,132	1,118,069	1,126,000	1,143,116
Normal TVaR						1,016,293	1,025,620	1,040,879	1,060,177	1,079,910	1,092,459	1,103,677	1,117,029	1,126,276	1,145,652
Lognormal TVaR						1,016,095	1,025,396	1,040,897	1,060,919	1,081,837	1,095,367	1,107,612	1,122,369	1,132,706	1,154,683
Gamma TVaR						1,016,157	1,025,464	1,040,880	1,060,651	1,081,154	1,094,338	1,106,217	1,120,469	1,130,411	1,151,437

Table 6-7:
Estimated Unpaid Model Output for **Time Horizon** option

Total Unpaid Distribution Graph

The final model output from the simulations is a histogram of the estimated unpaid amounts for the total of all accident years combined, as illustrated in Graph 6-6. The **Unpaid Graph**, or histogram, is created by dividing the range of all values from the simulation (using the maximum and minimum values) into one hundred "buckets" of equal size and counting the number of simulations that fall within each "bucket." Dividing by the total number of simulations (10,000 in this case) results in the frequency or probability for each "bucket" in the graph.

Since the simulation results often look "jagged" (as they do in Graph 6-6) a kernel density function is also used to calculate a "smoothed" line fit to the histogram values. The kernel density distribution is represented by the blue line in Graph 6-6.[44]



Graph 6-4:
Total Unpaid Distribution

44 In simple terms, a kernel density function can be thought of as a weighted average of values "close" to each point in the "jagged" distribution with progressively less weight being given to values the further they are from the point being evaluated. For a more detailed discussion of Kernel density functions, see Wand & Jones, "Kernel Smoothing," Chapman & Hall. 1995.

When you initially parameterize and run the model, you may find the resulting graph to be extremely narrow – almost a straight line. This is normally caused by a handful of extreme iterations. Many of the percentile results in the **Unpaid Table** may still appear reasonable, but it is still important to remove these extreme iterations since they will unduly affect your mean result. One of the most common causes of the extreme iteration is negative incremental values which can sometimes also result in an unrealistically high age-to-age factor. Thus, checking the **Limit Incremental to Zero** constraints may help remove these extreme iterations.

[THIS PAGE INTENTIONALLY LEFT BLANK]

7. Using the Hayne MLE Models

In addition to two modeling families based on commonly understood volume weighted age-to-age factors and cumulative data, Arius contains a third family based on fitting model parameters to incremental data and using Maximum Likelihood. Four different model frameworks (i.e., Berquist-Sherman, Cape Cod, Chain Ladder and Hoerl Curve) are included which allow you to fit parameters based on each framework and then simulate possible outcomes directly from the model parameters.

In addition, you can start with claim count data to estimate ultimate claim counts using the Frequency models and then use your estimated ultimate counts and the loss data to estimate ultimate unpaid claims using the Severity models. Or, you can simply enter ultimate claim count estimates and just use the Severity models.

Even though the Arius system has options to help you obtain the best model possible for your data, you can obtain valuable diagnostic information and even initial distribution estimates for a line of business with only a few steps, which can be summarized as:

- enter the data to be modeled,
- run the model diagnostics to calculate the model parameters and populate the necessary statistics and fields, and
- run the simulation to estimate future results (i.e., use the default model settings).

Of course, the diagnostics and model results can be used to evaluate and improve how your model fits your data. Understanding the purpose and use of the diagnostic tools requires some prior statistical knowledge so we direct the interested reader to Appendix B, which provides a general overview of the diagnostic process. Therefore, this section assumes prior knowledge of statistics and starts with the basics of running a model and builds on that foundation by exploring all of the different models, model options, diagnostics and model output. Note however, that Appendix B is based on using the diagnostic tools for the ODP Bootstrap model, so the differences when using the Hayne MLE models will be discussed here.

REQUIRED DATA: FREQUENCY MODELS

Inputs for the frequency models are relatively simple. You can start with nothing more than a triangle of reported claim count data and a vector of ultimate exposures, but if:

IN ADDITION TO CLAIM COUNT DATA, IF YOU PROVIDE:	THE SYSTEM CAN:
▪ a vector of earned premium data	▪ provide loss ratios by accident period at various percentiles
▪ A triangle of incurred or reported loss data	▪ reconcile Ultimate Losses using Paid, Case Reserves, and IBNR
▪ a vector of ultimate exposure data	▪ simulate based on exposure-adjusted losses rather than only the raw data

There are certain limitations that are imposed on the data by the mathematics involved in the model. Specifically:

- the triangle shape must be symmetrical in terms of row and column periods – i.e., it must be annual x annual or quarter x quarter;

Note:

If you have a partial last exposure period, then you should enter the earned premium in the appropriate column, but the ultimate premium and ultimate exposure are for the **full period**. For example, if you have an annual triangle but a 6 month last diagonal, then you should enter the premiums earned for the first 6 months in the earned premium column and the fully annualized premium and/or exposure in the ultimate premium and ultimate exposure columns, respectively. For more details see Section 9.

    - The system *will* work with triangles that contain a stub period (e.g., annual x annual with most recent diagonal evaluated at 6 months)
    - The system *will* work with triangles where the first development period is different from the rest (e.g., development columns of 6/18/30/42… or 3/15/27/39…)
    - The system *will not* work with truly asymmetrical triangles, such as annual accident periods x quarterly development.
- there must be at least 3 diagonals of data; and
- blank cells are acceptable anywhere in the triangle except on the most recent two diagonals, unless a whole row is blank (i.e., a triangle in run-off is OK)
- Individual negative age-to-age factors are acceptable, and the average for a column can be negative.
- Do not enter “0” values where the values are unknown. The model will treat cells with “0” values as information (that is, no losses occurred in this period), and blank cells as unknown.

REQUIRED DATA: SEVERITY MODELS

Inputs for the severity models are also relatively simple. You can start with nothing more than a triangle of paid loss data and a vector of ultimate claim counts, but if:

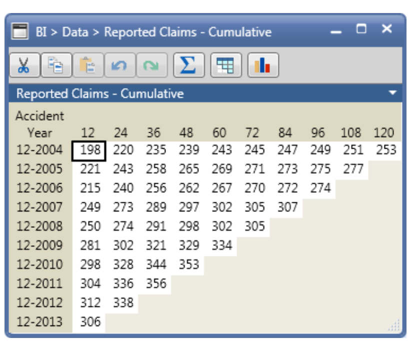
IN ADDITION TO DATA TRIANGLES, IF YOU PROVIDE:	THE SYSTEM CAN:
▪ a vector of earned premium data	▪ provide loss ratios by accident period at various percentiles
▪ a vector of ultimate exposure data	▪ simulate based on exposure-adjusted losses rather than only the raw data

All of the limitations that are imposed on the data by the mathematics involved in the model for the claim count data also apply to the paid loss data.

STEP 1: ENTER BASIC MODEL DATA

To get started, select one of your segments using the **Segment** drop down box below the **HOME** ribbon. In the **Navigation Pane**, select the DATA | INPUTS | ALL INPUTS collection. Notice that the first three tables in the collection, **Paid Loss**, **Case Loss Reserves** and **Incurred Loss**, are white; these are the data entry tables. You can fill in any two of these tables and the third with change to tan, which means it will be filled automatically and that you cannot enter data here any longer. Similarly for the claim counts, you can fill in any two of the **Closed Claims**, **Open Claims** or **Reported Claims** tables and the third will be filled in automatically.

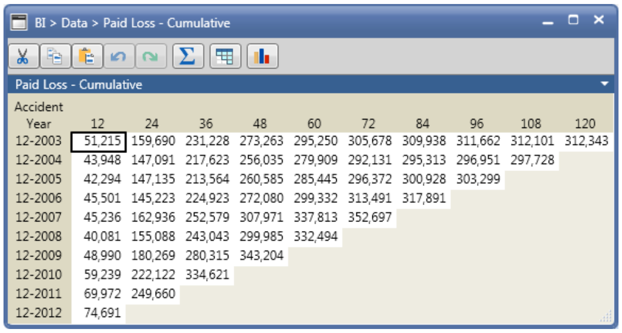
Note:
You can use the icon to switch between cumulative and incremental or the icon to switch between accident and calendar views, or both, prior to bringing in the data.



Accident Year	12	24	36	48	60	72	84	96	108	120
12-2004	198	220	235	239	243	245	247	249	251	253
12-2005	221	243	258	265	269	271	273	275	277	
12-2006	215	240	256	262	267	270	272	274		
12-2007	249	273	289	297	302	305	307			
12-2008	250	274	291	298	302	305				
12-2009	281	302	321	329	334					
12-2010	298	328	344	353						
12-2011	304	336	356							
12-2012	312	338								
12-2013	306									

Image 7-1:
Reported Claim Count Triangle

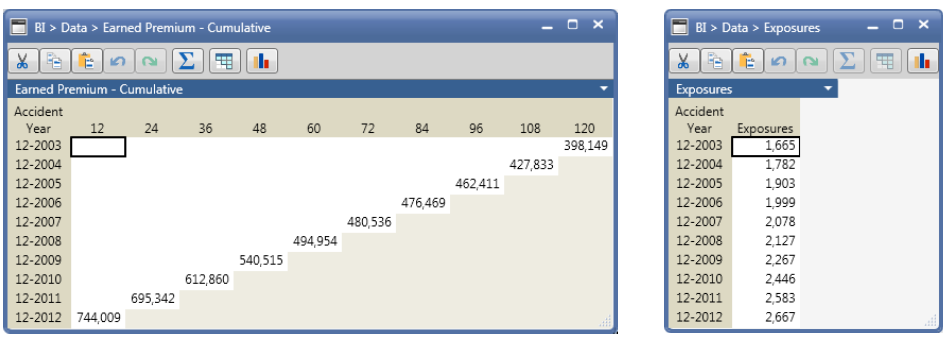
1. Enter data for the **Reported Claims** triangle (as illustrated in Image 7-1). You can either type in data or paste it in from another source.
2. Enter data for the **Paid Loss** triangle (as illustrated in Image 7-2). You can either type in data or paste it in from another source.



Accident Year	12	24	36	48	60	72	84	96	108	120
12-2003	51,215	159,690	231,228	273,263	295,250	305,678	309,938	311,662	312,101	312,343
12-2004	43,948	147,091	217,623	256,035	279,909	292,131	295,313	296,951	297,728	
12-2005	42,294	147,135	213,564	260,585	285,445	296,372	300,928	303,299		
12-2006	45,501	145,223	224,923	272,080	299,332	313,491	317,891			
12-2007	45,236	162,936	252,579	307,971	337,813	352,697				
12-2008	40,081	155,088	243,043	299,985	332,494					
12-2009	48,990	180,269	280,315	343,204						
12-2010	59,239	222,122	334,621							
12-2011	69,972	249,660								
12-2012	74,691									

Image 7-2:
Paid Loss Data Triangle

3. Also from the ALL INPUTS collection, you can enter **Earned Premium** and **Exposure** data, if you have that available (as illustrated in Image 7-3). Having this additional data allows the model to provide more information; this is especially true of Premium data, which allows the projection of ultimate loss ratios.



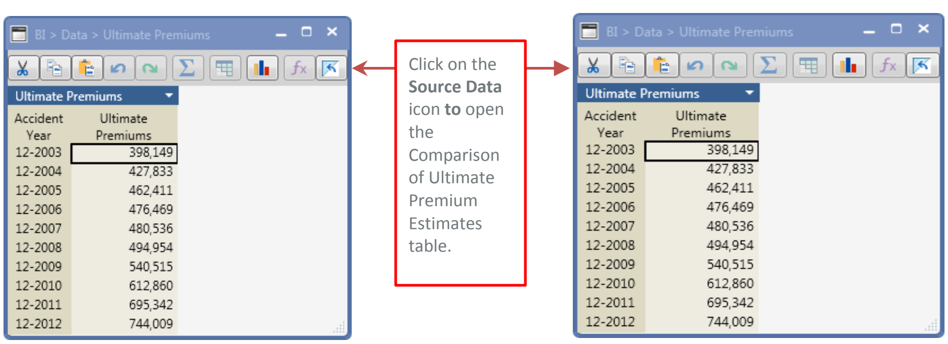
Accident Year	12	24	36	48	60	72	84	96	108	120
12-2003										398,149
12-2004									427,833	
12-2005								462,411		
12-2006							476,469			
12-2007						480,536				
12-2008					494,954					
12-2009				540,515						
12-2010			612,860							
12-2011		695,342								
12-2012	744,009									

Accident Year	Exposures
12-2003	1,665
12-2004	1,782
12-2005	1,903
12-2006	1,999
12-2007	2,078
12-2008	2,127
12-2009	2,267
12-2010	2,446
12-2011	2,583
12-2012	2,667

Image 7-3:
Earned Premium and Exposure tables

Note:
The earned premiums are entered in a triangle so that they can be developed in the Deterministic portion of the system.

4. In order to enter the **Ultimate Premium** data (again from the ALL INPUTS collection) you must open the table and click on the Source Data icon in order to get to the Deterministic table used to estimate **Ultimate Premium**. This is illustrated in Image 7-4.



Accident Year	Ultimate Premiums
12-2003	398,149
12-2004	427,833
12-2005	462,411
12-2006	476,469
12-2007	480,536
12-2008	494,954
12-2009	540,515
12-2010	612,860
12-2011	695,342
12-2012	744,009

Accident Year	Ultimate Premiums
12-2003	398,149
12-2004	427,833
12-2005	462,411
12-2006	476,469
12-2007	480,536
12-2008	494,954
12-2009	540,515
12-2010	612,860
12-2011	695,342
12-2012	744,009

Image 7-4:
Ultimate Premiums and Comparison of Ultimate Premiums Estimates tables

5. In order to enter the **Ultimate Claims** data, you must first open the **Object Library** from the **Home** ribbon, then in **Navigation Pane** within the **Object Library** open the Data | Results collection and select the **Ultimate Claims** table. Once you have the **Ultimate Claims** table open click on the Source Data icon to get to the Deterministic table used to estimate **Ultimate Claims**. This is illustrated in Image 7-5.

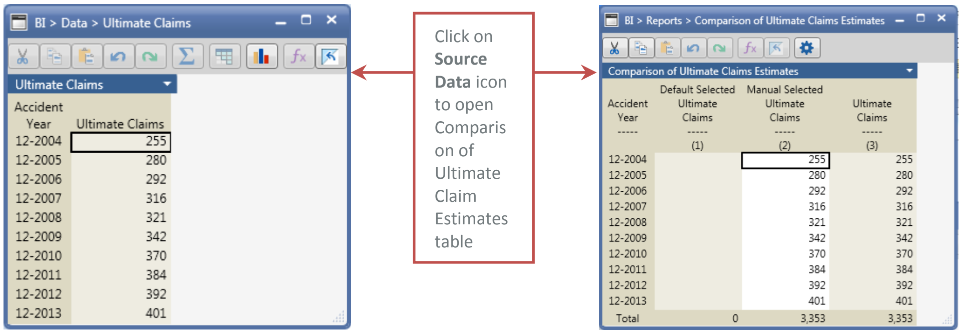


Image 7-5:
Ultimate Claims and Comparison of Ultimate Claims Estimates tables

STEP 2: REVIEW THE MODEL ASSUMPTIONS

If you have not already done so, from the **Home** ribbon select the **Chose Models** icon to select one or more of the Hayne MLE models, as illustrated in Image 7-6.

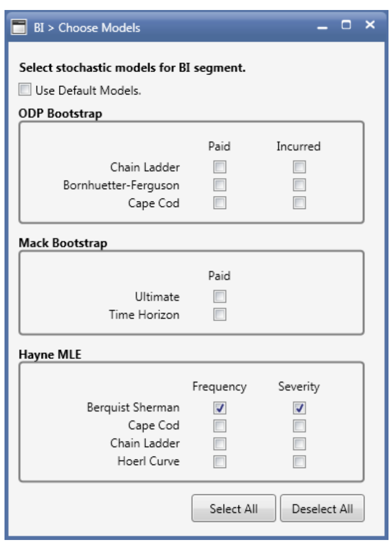


Image 7-6:
Choose Models dialog, with Hayne MLE models selected

Once a model has been selected, you can **Run Diagnostics** which will find the parameters and statistics for each model selected. Then from the **Navigation Pane**, select the Stochastic | Hayne MLE |

INCREMENTAL SEVERITY | BERQUIST-SHERMAN collection[45] and open the **Fit Details** window to review the assumptions as illustrated in Image 7-7.

BI > Models > Fit Details

Hayne MLE Cumulative Claim Severity – Berquist Sherman, Incremental Severity

Accident Year	12	24	36	48	60	72	84	96	108	120
12-2004	200.841	626.237	906.775	1,071.619	1,157.844	1,198.735	1,215.442	1,222.205	1,223.924	1,224.876
12-2005	156.957	525.324	777.225	914.410	999.675	1,043.326	1,054.690	1,060.539	1,063.314	
12-2006	144.842	503.886	731.385	892.415	977.552	1,014.974	1,030.576	1,038.694		
12-2007	143.991	459.566	711.783	861.014	947.254	992.061	1,005.983			
12-2008	140.921	507.588	786.849	959.412	1,052.375	1,098.745				
12-2009	117.196	453.474	710.652	877.150	972.206					
12-2010	132.407	487.214	757.609	927.579						
12-2011	154.269	578.442	871.408							
12-2012	178.501	636.889								
12-2013	186.262									

Hayne MLE Incremental Claim Severity – Berquist Sherman, Incremental Severity

Accident Year	12	24	36	48	60	72	84	96	108	120
12-2004	200.841	425.395	280.538	164.844	86.225	40.892	16.707	6.763	1.719	0.952
12-2005	156.957	368.367	251.900	137.186	85.264	43.651	11.365	5.849	2.774	
12-2006	144.842	359.044	227.499	161.031	85.137	37.422	15.602	8.117		
12-2007	143.991	315.576	252.217	149.231	86.240	44.807	13.922			
12-2008	140.921	366.667	279.261	172.563	92.963	46.369				
12-2009	117.196	336.277	257.179	166.497	95.056					
12-2010	132.407	354.808	270.394	169.971						
12-2011	154.269	424.174	292.965							
12-2012	178.501	458.388								
12-2013	186.262									

Image 7-7:
Hayne MLE **Fit Details** tables

Hayne MLE Parameters – Berquist Sherman, Incremental Severity

	12	24	36	48	60	72	84	96	108	120
Mean	145.815	355.258	247.437	151.040	83.694	40.691	13.926	6.735	2.282	0.918
Std Dev	7.323	16.951	11.727	7.304	4.273	2.284	0.911	0.530	0.236	0.144
	Acc. Period Trend	Kappa	Power	Model AIC	Model BIC					
Mean	0.013	1.867	0.939	434.695	460.790					
Std Dev	0.007	0.630	0.066							

Hayne MLE Fitted Incremental Severities – Berquist Sherman, Incremental Severity

Accident Year	12	24	36	48	60	72	84	96	108	120
12-2004	147.726	359.914	250.680	153.019	84.790	41.224	14.109	6.823	2.312	0.930
12-2005	149.662	364.631	253.965	155.025	85.902	41.764	14.294	6.913	2.343	0.942
12-2006	151.624	369.410	257.293	157.056	87.027	42.311	14.481	7.003	2.373	0.954
12-2007	153.611	374.251	260.665	159.115	88.168	42.866	14.671	7.095	2.404	0.967
12-2008	155.624	379.156	264.081	161.200	89.323	43.428	14.863	7.188	2.436	0.980
12-2009	157.664	384.125	267.542	163.312	90.494	43.997	15.058	7.282	2.468	0.992
12-2010	159.730	389.159	271.049	165.453	91.680	44.574	15.255	7.378	2.500	1.005
12-2011	161.823	394.259	274.601	167.621	92.882	45.158	15.455	7.474	2.533	1.019
12-2012	163.944	399.426	278.200	169.818	94.099	45.749	15.658	7.572	2.566	1.032
12-2013	166.093	404.661	281.846	172.043	95.332	46.349	15.863	7.672	2.600	1.045

Image 7-7 (cont.):
Hayne MLE **Fit Details** tables

In the **Fit Details** window, there are four specific sections – i.e., the *Cumulative Claim*, *Incremental Claim*, *Parameters* and *Fitted Incremental* sections. The *Cumulative Claim* and *Incremental Claim* triangles are based on either the input **Paid Loss** triangle divided by the **Ultimate Claims** (for Incremental Severity models) or the **Reported Claim** triangle divided by the **Ultimate Exposures** (for Incremental Frequency models). The *Parameters* are calculated when you RUN DIAGNOSTICS based on the specific Hayne model and the *Incremental Claim* section data. The *Fitted Incremental* data is derived using the *Parameters* of the specific model (See Appendix A for examples).

If you have not done so, save your file at this point.

STEP 3: EVALUATE YOUR DATA WITH THE MODEL'S DIAGNOSTICS

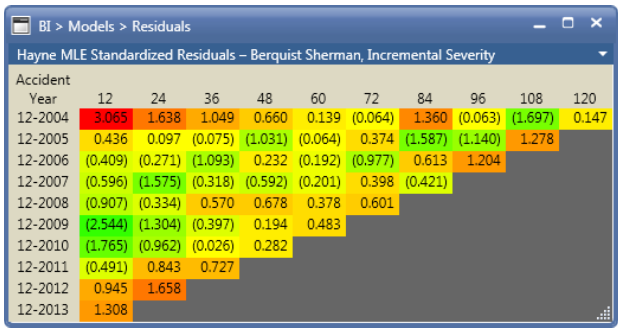
The Hayne MLE models are based on fitting parameters to assumptions about the incremental values in a triangle. In order to increase the model's predictive power, the data must be consistent with the assumptions that are inherent in the deterministic form of the model (or the model should be adjusted

[45] Within this manual we will be illustrating the Hayne MLE Berquist-Sherman model using Incremental Severity data, but the results for all other Hayne MLE models are consistent. The only difference between these models is the parameters which depend on the individual model assumptions as noted in Section 3 and Appendix A.

to be consistent with the data). The diagnostic output includes a variety of tables and graphs to help test these assumptions and then to adjust the model options to improve the statistical fit of the model to the data.

First, from the **Home** ribbon, click on the Run Diagnostics icon to populate the tables and graphs. In an iterative process, you will now want to analyze the diagnostic output, make adjustments to the model options, and then Run Diagnostics again to update the diagnostics results. An additional part of this iterative process is to click on the Run Simulations icon from the **Home** ribbon to run the simulations for the segment you are analyzing. This will allow you to review the model output for the segment, make adjustments to the model options and then either run diagnostics or simulations again until you have optimized the model.

For illustration purposes, we are using the BI data in the ODP_Mack_Hayne.apj file that is included with the system files in the *C:\Users\username\Documents\Milliman\Arius\DemoFiles* directory, where the *username* is your Windows user name.

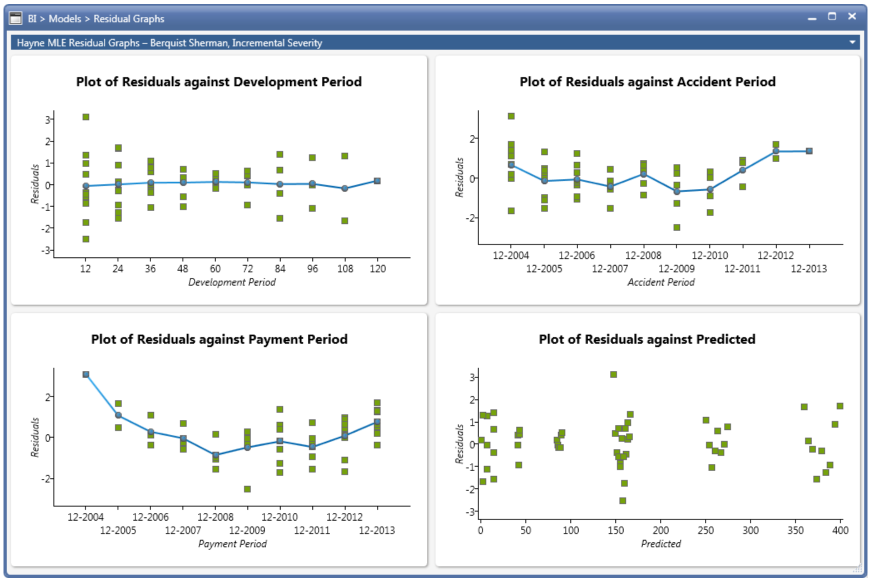


Accident Year	12	24	36	48	60	72	84	96	108	120
12-2004	3.065	1.638	1.049	0.660	0.139	(0.064)	1.360	(0.063)	(1.697)	0.147
12-2005	0.436	0.097	(0.075)	(1.031)	(0.064)	0.374	(1.587)	(1.140)	1.278	
12-2006	(0.409)	(0.271)	(1.093)	0.232	(0.192)	(0.977)	0.613	1.204		
12-2007	(0.596)	(1.575)	(0.318)	(0.592)	(0.201)	0.398	(0.421)			
12-2008	(0.907)	(0.334)	0.570	0.678	0.378	0.601				
12-2009	(2.544)	(1.304)	(0.397)	0.194	0.483					
12-2010	(1.765)	(0.962)	(0.026)	0.282						
12-2011	(0.491)	0.843	0.727							
12-2012	0.945	1.658								
12-2013	1.308									

Image 7-8:
Standardized Residuals

In the **Navigation Pane**, select the Stochastic | Hayne MLE | Incremental Severity | Berquist-Sherman collection. The first diagnostic output is the standardized residuals shown in the **Residuals** table. The calculations for the residuals are described in Appendix A, although the residuals will be based on the data adjusted for exposures and/or stub periods (as illustrated in Image 7-8). As you can see in Image 7-8, the standardized residuals for the Hayne MLE models are colored as a heat map similar to the Deterministic Age-to-Age factors. The difference is that the largest value overall is red and the smallest overall value is green, whereas for the development factors the colors are only based on the values in each development column.

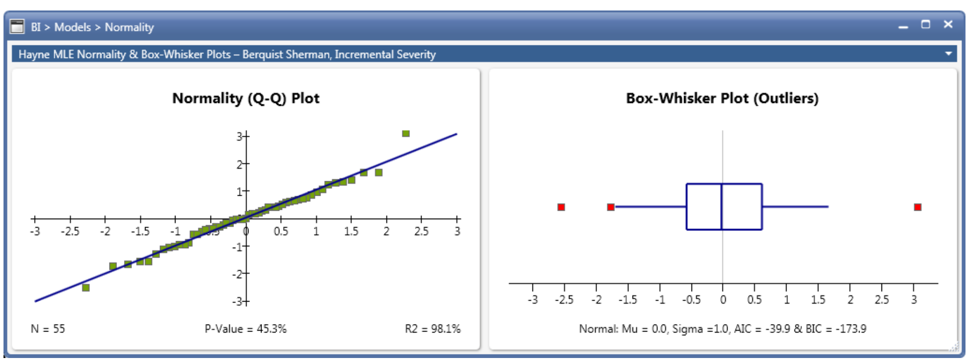
As a tool to help evaluate the residuals, each model collection includes a **Residual Graphs** window (as illustrated in Graph 7-1). These graphics show plots of the residuals (from Image 7-8) against the development, accident, and payment periods, as well as a plot of the residuals vs. the fitted (i.e., predicted) values. These will help you identify trends or other features in your data that may not be completely modeled, thus indicating that the Hayne MLE predictions from the data may be less than optimal.



Graph 7-1:
Plots of Residuals

In the Hayne MLE models, new parameters for each iteration are sampled based on the mean, standard deviation (illustrated in Image 7-7) and variance-covariance matrix (not shown) parameters for that model. Unlike the ODP bootstrap model, the standardized residuals are calculated using the standard deviation of each development period, so heteroscedasticity (i.e., different variances) does not occur. Thus, there are no heteroscedasticity adjustment factors for the Hayne MLE models.

STEP 3.A: IDENTIFY AND EXCLUDE OUTLIERS



Graph 7-2:
Normality & Box-Whisker Plots

The next diagnostics window, **Normality**, will help you judge the overall quality of the model and general improvement in the model if you exclude outliers. For example, look at Graph 7-2 below which corresponds to the plots shown above in Graph 7-1. As noted in Appendix B, the P-Value, R^2, AIC and BIC values under the Normality (Q-Q) Plot and Box-Whisker Plot are a useful guide. You can also review these graphs before and after excluding outliers.

Note:
Removing outliers should be done with caution as this will usually reduce the "extremes" of the resulting model distribution.

If shown on the Box-Whisker plot, it might be reasonable to remove outliers from the model. When you do want to "remove" an outlier from the data, the procedure for doing so is to determine the correct cell(s) and identify it (them) with a one ("1") in the corresponding cell(s) in the **Outliers** triangle (as illustrated in Image 7-9 for the largest and smallest residuals in Image 7-8).

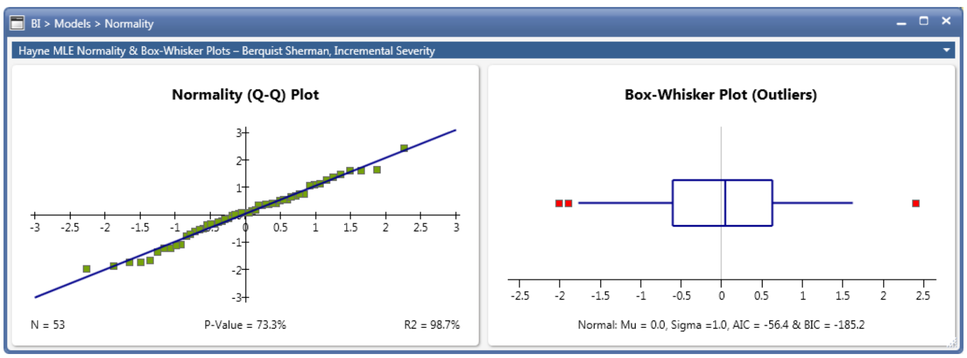
BI > Models > Outliers

Hayne MLE Outliers – Berquist Sherman, Incremental Severity

Accident Year	12	24	36	48	60	72	84	96	108	120
12-2004	1	0	0	0	0	0	0	0	0	0
12-2005	0	0	0	0	0	0	0	0	0	
12-2006	0	0	0	0	0	0	0	0		
12-2007	0	0	0	0	0	0	0			
12-2008	0	0	0	0	0	0				
12-2009	1	0	0	0	0					
12-2010	0	0	0	0						
12-2011	0	0	0							
12-2012	0	0								
12-2013	0									

Image 7-9:
Outliers triangle with two outliers selected

After the outlier(s) have been identified in this manner, use RUN DIAGNOSTICS again to update the tables and graphs. After the tables are graphs have been updated, the selected outlier(s) will no longer be visible in any of the graphs, but you can review the statistics in the Normality graphs (as illustrated in Graph 7-3) to see if they have improved (compare statistics in Graph 7-3 to Graph 7-2).



Graph 7-3:
Normality & Box-Whisker Plots, *after* excluding outliers

You can still see which cell(s) have been given zero weight in the model by opening the **Outliers** table (Image 7-9) or by opening the **Residuals** (Image 7-10) or **Fit Details** (Image 7-11) windows. To restore an outlier (give it weight again), you must change the one(s) ("1") in the **Outliers** table to a zero ("0") and use RUN DIAGNOSTICS again.

BI > Models > Residuals

Hayne MLE Standardized Residuals – Berquist Sherman, Incremental Severity

Accident Year	12	24	36	48	60	72	84	96	108	120
12-2004		2.406	1.572	1.017	0.321	0.035	1.592	(0.025)	(1.764)	0.139
12-2005	1.068	0.369	0.092	(1.126)	0.011	0.482	(1.774)	(1.254)	1.325	
12-2006	(0.064)	(0.175)	(1.249)	0.343	(0.213)	(1.152)	0.622	1.237		
12-2007	(0.359)	(1.891)	(0.366)	(0.734)	(0.299)	0.365	(0.592)			
12-2008	(0.820)	(0.420)	0.640	0.716	0.306	0.521				
12-2009		(1.711)	(0.633)	0.045	0.342					
12-2010	(2.014)	(1.379)	(0.278)	0.054						
12-2011	(0.537)	0.737	0.531							
12-2012	1.114	1.623								
12-2013	1.452									

Image 7-10:
Standardized Residuals, *after* excluding outliers

BI > Models > Fit Details

Hayne MLE Cumulative Claim Severity – Berquist Sherman, Incremental Severity

Accident Year	12	24	36	48	60	72	84	96	108	120
12-2004		626.237	906.775	1,071.619	1,157.844	1,198.735	1,215.442	1,222.205	1,223.924	1,224.876
12-2005	156.957	525.324	777.225	914.410	999.675	1,043.326	1,054.690	1,060.539	1,063.314	
12-2006	144.842	503.886	731.385	892.415	977.552	1,014.974	1,030.576	1,038.694		
12-2007	143.991	459.566	711.783	861.014	947.254	992.061	1,005.983			
12-2008	140.921	507.588	786.849	959.412	1,052.375	1,098.745				
12-2009		453.474	710.652	877.150	972.206					
12-2010	132.407	487.214	757.609	927.579						
12-2011	154.269	578.442	871.408							
12-2012	178.501	636.889								
12-2013	186.262									

Hayne MLE Incremental Claim Severity – Berquist Sherman, Incremental Severity

Accident Year	12	24	36	48	60	72	84	96	108	120
12-2004		425.395	280.538	164.844	86.225	40.892	16.707	6.763	1.719	0.952
12-2005	156.957	368.367	251.900	137.186	85.264	43.651	11.365	5.849	2.774	
12-2006	144.842	359.044	227.499	161.031	85.137	37.422	15.602	8.117		
12-2007	143.991	315.576	252.217	149.231	86.240	44.807	13.922			
12-2008	140.921	366.667	279.261	172.563	92.963	46.369				
12-2009		336.277	257.179	166.497	95.056					
12-2010	132.407	354.808	270.394	169.971						
12-2011	154.269	424.174	292.965							
12-2012	178.501	458.388								
12-2013	186.262									

Hayne MLE Parameters – Berquist Sherman, Incremental Severity

	12	24	36	48	60	72	84	96	108	120
Mean	137.176	343.078	240.038	147.070	81.824	39.927	13.728	6.650	2.256	0.913
Std Dev	6.345	13.676	9.498	5.964	3.542	1.936	0.802	0.480	0.222	0.141
	Acc. Period Trend	Kappa	Power	Model AIC	Model BIC					
Mean	0.020	1.817	0.906	399.328	424.942					
Std Dev	0.006	0.628	0.066							

Hayne MLE Fitted Incremental Severities – Berquist Sherman, Incremental Severity

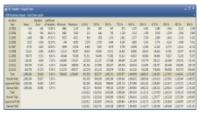
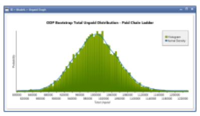
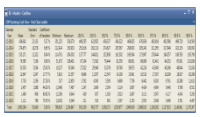
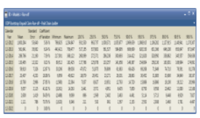
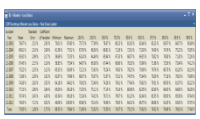
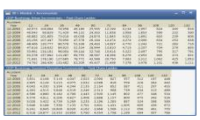
Accident Year	12	24	36	48	60	72	84	96	108	120
12-2004	139.956	350.030	244.902	150.051	83.482	40.736	14.006	6.784	2.302	0.931
12-2005	142.792	357.123	249.865	153.092	85.174	41.561	14.290	6.922	2.349	0.950
12-2006	145.686	364.360	254.928	156.194	86.900	42.403	14.580	7.062	2.396	0.970
12-2007	148.638	371.744	260.094	159.359	88.661	43.263	14.875	7.205	2.445	0.989
12-2008	151.650	379.277	265.365	162.588	90.458	44.139	15.176	7.351	2.494	1.009
12-2009	154.723	386.963	270.743	165.883	92.291	45.034	15.484	7.500	2.545	1.030
12-2010	157.859	394.805	276.229	169.245	94.161	45.947	15.798	7.652	2.597	1.051
12-2011	161.058	402.805	281.827	172.675	96.069	46.878	16.118	7.807	2.649	1.072
12-2012	164.321	410.968	287.538	176.174	98.016	47.828	16.445	7.965	2.703	1.094
12-2013	167.651	419.296	293.365	179.744	100.002	48.797	16.778	8.127	2.758	1.116

Image 7-11:
Hayne MLE Fit Details tables, *after* excluding

In addition to the model diagnostics described above, the results output also has diagnostic features. Thus, running the model using Run Simulations, reviewing the model output and adjusting model parameters and assumptions is part of the diagnostic process. Reviewing the model output is discussed in more detail in the remainder of this Section.

SUMMARY OF OUTPUT

The results for each model are shown in their own collection. For example, in the **Navigation Pane**, select the STOCHASTIC | HAYNE MLE | INCREMENTAL SEVERITY | BERQUIST-SHERMAN collection to view all of the simulation results for the Hayne MLE Berquist-Sherman Incremental Severity model.

	Estimated Unpaid	Mean, Standard Error, Coefficient of Variation, Min, Max and Percentiles. Total Distributions and TVaRs.
	Total Unpaid Distribution	Histogram and kernel density of total unpaid.
	Estimated Cash Flow	Future calendar period payments.
	Estimated Run-off	Total unpaid as future calendar periods are removed.
	Estimated Ultimate	Time zero to ultimate values
	Estimated Loss Ratios	Time zero to ultimate loss ratios.
	Incremental Values	Mean and standard deviation values for each incremental cell, historical and future.

STEP 4: EVALUATE THE OUTPUT FOR EACH MODEL

After the model diagnostics have been set up and reviewed, the next step in the evaluation of each model is to use RUN SIMULATIONS to run the simulations for the segment you are analyzing. To illustrate the diagnostic elements of the simulation output we will review the results for the Hayne MLE Berquist-Sherman model.

Estimated Unpaid Results

The first diagnostic element of the **Unpaid Table** (illustrated in Table 7-1) can be seen by reviewing the Standard Error and Coefficient of Variation columns. As general rules, the standard error should go up as you move from the oldest years to the most recent years and the standard error for the total of all years should be larger than any individual year. In Table 7-1, the standard errors follow these general rules. For the coefficients of variation, they should go down when moving from the oldest years to the more recent years and the coefficient of variation for all years combined should be less than for any individual year.[46] The coefficients of variation in Table 7-1 also follow the general rules.

BI > Models > Unpaid Table

Hayne MLE Unpaid – Berquist Sherman, Incremental Severity

Accident Year	Mean	Standard Error	Coefficient of Variation	Minimum	Maximum	10.0 %	25.0 %	50.0 %	75.0 %	90.0 %	95.0 %	97.5 %	99.0 %	99.5 %	99.9 %
12-2004	0	0	0.00 %	0	0	0	0	0	0	0	0	0	0	0	0
12-2005	264	60	22.87 %	(70)	567	189	224	262	301	342	367	388	419	440	484
12-2006	973	140	14.41 %	392	1,682	801	881	971	1,062	1,151	1,204	1,256	1,315	1,353	1,480
12-2007	3,306	374	11.32 %	1,758	4,923	2,841	3,057	3,299	3,549	3,785	3,920	4,052	4,224	4,333	4,633
12-2008	8,198	779	9.50 %	4,909	11,435	7,210	7,684	8,194	8,710	9,186	9,487	9,760	10,073	10,397	10,875
12-2009	23,885	2,060	8.63 %	16,296	30,863	21,284	22,494	23,842	25,267	26,565	27,324	28,003	28,817	29,248	30,179
12-2010	60,129	4,590	7.63 %	43,500	79,321	54,299	57,022	60,045	63,155	66,005	67,731	69,300	71,188	72,252	75,616
12-2011	127,541	8,670	6.80 %	96,008	159,497	116,539	121,614	127,424	133,281	138,653	142,132	144,993	148,400	150,736	155,224
12-2012	241,028	15,497	6.43 %	176,109	315,899	221,323	230,303	240,653	251,320	261,158	266,986	272,086	278,108	281,585	290,988
12-2013	411,717	25,383	6.17 %	318,442	517,556	379,650	394,424	411,118	428,694	444,955	453,946	461,985	472,437	479,533	493,234
Total	877,040	39,867	4.55 %	718,876	1,028,426	825,926	849,515	876,016	903,655	928,904	943,868	957,414	972,724	982,786	1,004,920
Normal %iles	877,040	39,865	4.55 %			825,950	850,151	877,040	903,929	928,130	942,613	955,175	969,781	979,726	1,000,233
Lognormal %iles	877,040	39,892	4.55 %			826,548	849,677	876,134	903,416	928,696	944,162	957,785	973,871	984,980	1,008,286
Gamma %iles	877,040	39,856	4.54 %			826,366	849,838	876,436	903,583	928,490	943,611	956,857	972,414	983,101	1,005,383
TVaR						884,606	893,749	908,876	928,470	948,513	961,344	972,734	986,468	995,957	1,014,970
Normal TVaR						884,814	893,931	908,848	927,713	947,003	959,271	970,237	983,290	992,329	1,011,270
Lognormal TVaR						884,591	893,672	908,842	928,489	949,068	962,407	974,497	989,090	999,327	1,021,128
Gamma TVaR						884,659	893,750	908,831	928,206	948,331	961,289	972,975	987,010	996,809	1,017,554

Table 7-1: Estimated Unpaid Model Output

Note: For policy period data or incomplete accident period data, the unpaid data in the last row(s) will be reduced to only include earned exposures.

The reason the standard errors (value scale) tend to go up is that they tend to follow the magnitude of the mean or expected value estimates. The reason the coefficients of variation (percent scale) tend to go down has more to do with the independence in the incremental claim payment stream. For the oldest accident year, there is typically only one (or a few) incremental payment(s) left so the variability of that payment(s) is (almost) fully reflected in the coefficient. For the most current accident year, the "up and down" variations in the future incremental payment stream can offset each other thus causing the total variation to be a function of the correlation between each incremental payment for that accident year (i.e., the incremental payments are assumed independent).

Note: Caution should be exercised in the interpretation and adjustments for increases in the coefficient of variation in recent years. While keeping the theory in mind is appropriate, this must be balanced with the need to keep from underestimating the uncertainty of the more recent years.

The coefficient of variation rules noted above are a reflection of the step 6's described in Section 3 (and Appendix A), in the sense that they describe the process variance in the model. While the coefficients of variation should go down, if they do start going back up in the most recent year(s), then this could be the result of the following issues:

1. The parameter uncertainty tends to increase when moving from the oldest years to the more recent years as more and more parameters are used in the model. In the most recent year(s), the parameter uncertainty could be "overpowering" the process uncertainty causing the coefficient of variation to start going back up. At the very least, the increasing parameter uncertainty will cause the rate of decrease in the coefficient of variation to slow down.
2. If the increase in the most recent year(s) is significant, then this could indicate that the model is overestimating the uncertainty in those years. If this is the case, then an adjustment to the model parameters may be needed (e.g., limit incrementals to zero, etc.).

While we mentioned the rules for the standard error and coefficient of variation for the total of all years, it is also worth noting that in addition to the correlation (independence) within each accident year the total of all years also includes the impact of the correlation (independence) between accident years. In essence, when one or more accident years are "bad" we do not expect all accident years to be

46 These standard error and coefficient of variation rules are based on the independence of the incremental process risk and assume that the underlying exposures are relatively stable from year to year – i.e., no radical changes. In practice, random changes do occur from one year to the next which could cause the actual standard errors to deviate from these rules somewhat. In other words, these rules should generally hold true, but are not considered hard and fast rules in every case. Strictly speaking, the total all years rules assume that the individual years are not positively correlated.

"bad." To see this impact, you can add the accident year standard errors and note that they will not sum to the standard error for all years combined.[47]

The next diagnostic element in the **Unpaid Table** is the **Minimum** and **Maximum** columns. In these columns, the smallest and largest values, respectively, from among all iterations of the simulation are displayed. These values can be reviewed judgmentally to make sure that they are not outside the "realm of possibility." If they do seem a bit unrealistic then they could indicate the need to review the model options. For example, the presence of negative numbers might lead to changing one or both of the options which limit incremental values to zero. Sometimes "extreme" outliers in the results will show up in these columns and may also distort the histogram (discussed later in this Section).

Risk Measures

Also included in Table 7-1, notice that there are three rows of "Percentile" numbers and then four rows of TVaR numbers at the bottom of these tables under each of the percentile columns. For the three "Percentile" rows, the normal, lognormal and gamma distributions, respectively, have been fit to the Total unpaid claim distribution. The fitted mean, standard deviation and selected percentiles are shown under the Mean, Standard Error and Percentile columns, respectively, so that the smoothed results can be used to judge the quality of fit for each distribution or other purposes such as parameterizing a DFA model or using smoothed results in the tail of the distribution.

The Tail Value at Risk (TVaR)[48] is the average of all of the simulated values equal to or greater than the percentile value. For example, in Table 7-1 the 75th percentile value for the total unpaid for all accident years combined is 903,655 and the average of all simulated values that are greater than or equal to 903,655 is 928,470. The "Normal TVAR," "Lognormal TVaR" and "Gamma TVaR" rows are calculated the same way, except that instead of using the actual simulated values from the model the respective fitted distributions are used in the calculations.

To interpret the TVaR numbers, the question we are trying to answer with a TVaR number is "if the actual outcome does exceed the X percentile value, on average how much might it exceed that value by?" This is an important question related to risk based capital calculations and other technical aspects of enterprise risk management, although a more complete discussion is beyond the scope of this manual. It is worth noting, however, that the purpose of the normal, lognormal and gamma TVaR numbers is to provide "smoothed" values in the sense that some of the random noise is kept from distorting the calculations.

Estimated Cash Flow Results

In addition to the results by accident year, we can also review the model output by calendar year (or by future diagonal) in the **Cash Flow** table as illustrated in Table 7-2. Comparing Table 7-2 to 7-1, notice that the Total row is identical since the total is the same whether you add the parts horizontally or diagonally. Similar diagnostic issues can be reviewed in this table, except that the relative values of the standard errors and coefficients of variation move in the opposite direction for calendar years compared to accident years. This should make intuitive sense as the "final" payments projected the farthest out into the future should be the smallest yet relatively most uncertain.

47 Likewise, the minimum, maximum and each of the percentile columns will not sum to the total for all years combined. In contrast, adding the mean values for each accident year will sum to the total for all years combined.

48 The Tail Value at Risk is sometimes referred to as the Conditional Tail Expectation.

BI > Models > Cash Flow

Hayne MLE Cash Flow – Berquist Sherman, Incremental Severity

Calendar Year	Mean	Standard Error	Coefficient of Variation	Minimum	Maximum	10.0 %	25.0 %	50.0 %	75.0 %	90.0 %	95.0 %	97.5 %	99.0 %	99.5 %	99.9 %
12-2014	392,309	23,153	5.90 %	296,723	475,367	362,834	376,357	392,025	407,983	421,924	430,649	438,853	448,928	454,564	465,211
12-2015	240,411	15,551	6.47 %	178,477	305,070	220,936	229,914	240,150	250,396	260,609	266,479	271,847	278,459	282,835	289,617
12-2016	132,464	9,688	7.31 %	99,200	174,541	120,284	125,853	132,221	138,855	144,924	148,691	152,378	156,071	158,751	165,109
12-2017	65,963	5,297	8.03 %	43,741	88,896	59,301	62,331	65,859	69,556	72,701	74,714	76,574	78,901	80,485	83,726
12-2018	28,855	2,622	9.09 %	19,919	39,772	25,588	27,052	28,750	30,600	32,227	33,284	34,220	35,333	36,145	37,641
12-2019	10,688	1,056	9.88 %	6,786	15,187	9,363	9,971	10,659	11,372	12,057	12,477	12,836	13,294	13,602	14,291
12-2020	4,482	521	11.61 %	2,597	6,616	3,827	4,137	4,472	4,815	5,142	5,361	5,560	5,807	5,945	6,307
12-2021	1,448	208	14.34 %	554	2,458	1,191	1,305	1,441	1,579	1,716	1,800	1,882	1,974	2,058	2,203
12-2022	421	92	21.78 %	59	847	309	358	416	479	541	577	613	656	684	751
Total	877,040	39,867	4.55 %	718,876	1,028,426	825,926	849,515	876,016	903,655	928,904	943,868	957,414	972,724	982,786	1,004,920

Table 7-2:
Estimated Cash Flow Model Output

Estimated Unpaid Claim Runoff Results

Another report similar to the **Cash Flow** table is the **Run-off** table. Rather than looking at individual diagonal results, the **Run-off** table starts with the total unpaid results and then looks at how the total unpaid will decrease over time as successive diagonals are removed, as illustrated in Table 7-3. Comparing Table 7-3 to 7-1 & 7-2, notice that the first row of Table 7-3 is identical to the Total rows in Tables 7-1 and 7-2. Each successive row in Table 7-3 is then the total of the remaining diagonals.

BI > Models > Run-off

Hayne MLE Unpaid Claim Run-off – Berquist Sherman, Incremental Severity

Calendar Year	Mean	Standard Error	Coefficient of Variation	Minimum	Maximum	10.0 %	25.0 %	50.0 %	75.0 %	90.0 %	95.0 %	97.5 %	99.0 %	99.5 %	99.9 %
12-2013	877,040	39,867	4.55 %	718,876	1,028,426	825,926	849,515	876,016	903,655	928,904	943,868	957,414	972,724	982,786	1,004,920
12-2014	484,731	24,823	5.12 %	379,236	577,612	453,069	467,813	484,055	500,980	517,083	526,395	535,443	545,794	551,981	565,547
12-2015	244,321	14,239	5.83 %	196,728	300,823	226,243	234,636	243,927	253,676	263,106	268,521	273,378	278,714	282,543	291,177
12-2016	111,856	7,292	6.52 %	81,332	148,130	102,619	106,921	111,738	116,623	121,253	124,151	126,581	129,340	131,601	136,662
12-2017	45,893	3,393	7.39 %	34,137	59,455	41,621	43,561	45,810	48,141	50,242	51,539	52,729	54,448	55,289	56,964
12-2018	17,039	1,428	8.38 %	12,296	23,182	15,240	16,067	17,010	17,966	18,901	19,457	19,960	20,500	20,914	21,838
12-2019	6,350	647	10.20 %	4,154	9,255	5,525	5,915	6,336	6,767	7,173	7,438	7,691	7,984	8,174	8,474
12-2020	1,868	252	13.51 %	902	3,051	1,553	1,697	1,859	2,031	2,195	2,302	2,394	2,497	2,575	2,758
12-2021	421	92	21.78 %	59	847	309	358	416	479	541	577	613	656	684	751
12-2022	0	0	0.00 %	0	0	0	0	0	0	0	0	0	0	0	0

Table 7-3:
Estimated Unpaid Claim Run-Off Model Output

Estimated Ultimate Results

The next collection table is the **Ultimate Table** as illustrated in Table 7-4. Unlike the **Unpaid**, **Cash Flow** and **Run-off** tables, the values in the **Ultimate Table** are calculated from all simulated values, not just the values beyond the end of the triangles. In other words, since the model parameters are used to simulate the entire rectangle, we have enough information to estimate the complete variability in the ultimate values from day one in each accident year until all claims are completely paid and settled.

BI > Models > Ultimate Table

Hayne MLE Ultimate – Berquist Sherman, Incremental Severity

Accident Year	Mean	Standard Error	Coefficient of Variation	Minimum	Maximum	10.0 %	25.0 %	50.0 %	75.0 %	90.0 %	95.0 %	97.5 %	99.0 %	99.5 %	99.9 %
12-2004	270,487	16,601	6.14 %	211,516	331,023	249,474	259,125	270,356	281,438	291,929	298,275	303,574	309,846	313,763	322,850
12-2005	300,816	17,210	5.72 %	223,677	373,860	279,268	289,172	300,513	312,470	322,936	329,230	334,843	341,588	345,876	358,681
12-2006	318,344	17,257	5.42 %	250,675	396,135	296,486	307,016	318,332	329,700	340,398	346,667	352,240	359,730	364,344	372,951
12-2007	348,437	17,550	5.04 %	272,930	413,057	326,128	336,913	348,366	359,916	370,967	377,359	382,299	389,303	394,725	405,752
12-2008	358,978	18,051	5.03 %	290,864	433,152	335,920	346,997	358,874	370,908	381,703	388,598	394,669	401,569	406,281	419,381
12-2009	387,145	19,249	4.97 %	317,604	470,638	362,876	374,358	386,851	399,894	412,052	419,182	425,442	432,618	438,302	451,091
12-2010	424,621	20,832	4.91 %	337,402	512,681	398,299	411,005	424,543	437,907	451,425	459,485	466,294	475,366	480,419	491,520
12-2011	446,379	22,636	5.07 %	345,428	530,079	417,797	431,231	446,559	461,338	475,169	483,755	491,226	499,833	506,066	518,884
12-2012	461,715	24,801	5.37 %	378,359	567,721	429,945	444,864	461,647	478,304	493,407	503,187	510,765	518,986	527,230	543,210
12-2013	478,322	27,369	5.72 %	374,529	598,909	443,525	459,724	477,498	496,615	513,671	523,453	532,754	544,580	552,128	565,745
Total	3,795,245	87,896	2.32 %	3,463,482	4,171,013	3,682,108	3,736,087	3,795,918	3,854,450	3,907,456	3,940,852	3,967,258	4,001,378	4,018,463	4,063,547

Table 7-4:
Estimated Ultimate Model Output

Because we are using all simulated values, the standard errors in Table 7-4 should be proportionate to the means while the coefficients of variation should be relatively constant by accident year. Diagnostically, any increases in standard error and coefficient of variation for the latest few years will be consistent with the reasons cited earlier for the **Unpaid Table**.

Estimated Ultimate Loss Ratio Results

The next collection table shows the ultimate **Loss Ratios** by accident year as illustrated in Table 7-5. If there are no earned premiums or ultimate premiums input into the model, then this table will not be filled in since the model cannot calculate a loss ratio without the premium information.[49]

BI > Models > Loss Ratios

Mack Bootstrap Ultimate Loss Ratios – Paid Chain Ladder

Accident Year	Mean	Standard Error	Coefficient of Variation	Minimum	Maximum	10.0 %	25.0 %	50.0 %	75.0 %	90.0 %	95.0 %	97.5 %	99.0 %	99.5 %	99.9 %
12-2004	78.60 %	0.09 %	0.11 %	78.29 %	78.94 %	78.49 %	78.54 %	78.60 %	78.66 %	78.72 %	78.75 %	78.78 %	78.81 %	78.83 %	78.88 %
12-2005	69.78 %	0.09 %	0.14 %	69.39 %	70.15 %	69.66 %	69.72 %	69.78 %	69.85 %	69.90 %	69.94 %	69.97 %	70.00 %	70.03 %	70.08 %
12-2006	65.90 %	0.11 %	0.17 %	65.44 %	66.32 %	65.76 %	65.83 %	65.90 %	65.98 %	66.05 %	66.09 %	66.12 %	66.16 %	66.19 %	66.23 %
12-2007	67.46 %	0.15 %	0.22 %	66.92 %	68.04 %	67.27 %	67.36 %	67.46 %	67.56 %	67.65 %	67.70 %	67.75 %	67.80 %	67.84 %	67.92 %
12-2008	75.22 %	0.21 %	0.28 %	74.43 %	76.02 %	74.94 %	75.07 %	75.22 %	75.36 %	75.49 %	75.57 %	75.64 %	75.71 %	75.76 %	75.91 %
12-2009	71.73 %	0.40 %	0.56 %	70.33 %	73.37 %	71.22 %	71.46 %	71.73 %	72.00 %	72.24 %	72.39 %	72.51 %	72.67 %	72.77 %	72.97 %
12-2010	74.30 %	0.72 %	0.97 %	71.69 %	77.13 %	73.39 %	73.81 %	74.29 %	74.79 %	75.22 %	75.50 %	75.75 %	76.05 %	76.22 %	76.58 %
12-2011	77.33 %	1.46 %	1.89 %	71.79 %	82.35 %	75.45 %	76.34 %	77.33 %	78.31 %	79.20 %	79.77 %	80.22 %	80.74 %	81.10 %	81.89 %
12-2012	76.96 %	2.53 %	3.28 %	66.79 %	88.29 %	73.70 %	75.23 %	76.95 %	78.64 %	80.18 %	81.18 %	81.97 %	82.94 %	83.64 %	85.02 %
12-2013	75.58 %	5.18 %	6.85 %	54.72 %	95.58 %	68.98 %	72.08 %	75.51 %	78.97 %	82.39 %	84.28 %	85.96 %	88.04 %	89.36 %	92.21 %
Total	73.64 %	0.86 %	1.17 %	70.20 %	76.92 %	72.53 %	73.05 %	73.63 %	74.20 %	74.76 %	75.09 %	75.35 %	75.69 %	75.95 %	76.37 %

Table 7-5:

Estimated Loss Ratio Model Output

Note:
For policy period data or incomplete accident period data, the unpaid data in the last row(s) will be reduced to only include earned exposures. However, since the earned exposures are divided by the Earned Premium to calculate the loss ratios we have a match of losses to premium.

Unlike the **Unpaid**, **Cash Flow** and **Run-off** tables, the values in the **Loss Ratios** table are calculated from all simulated values, not just the values beyond the end of the triangles. In other words, since the model parameters are used to simulate the entire rectangle, we have enough information to estimate the complete variability in the loss ratio from day one in each accident year until all claims are completely paid and settled.[50]

Because we are using all simulated values, the standard errors in Table 7-5 should be proportionate to the means while the coefficients of variation should be relatively constant by accident year. Diagnostically, any increases in standard error and coefficient of variation for the latest few years will be consistent with the reasons cited earlier for the **Unpaid Table**.

Estimated Incremental Results

The next collection table is designed to help you take a deeper look at the simulations and to understand the reasons for increases in the coefficients of variation (in Tables 7-1 and 7-4). They show the mean and standard deviations, respectively, by accident year by development period. As illustrated in Table 7-6, both the Mean and Standard Deviation **Incrementals** can be reviewed down each column or across each row to look for any irregularities in the expected patterns.

[*Note that in some versions of Arius the **Incrementals** table may not be included in the standard Hayne MLE collections, and if not, you can add it to (drag it into) your project from the Object Library.*]

The Hayne MLE models use the same parameters for each incremental cell and as such the resulting mean and standard deviation in each cell is consistent with those model parameters, as illustrated in Table 7-6.

49 Earned premiums are used as the denominator of the loss ratios. However, if earned premiums are not input then earned premiums are estimated from the ultimate premiums.

50 If we are only interested in the "remaining" volatility in the loss ratio, then the values in the estimated **Unpaid Table** can be added to the cumulative values in the data input table and divided by the premiums.

BI > Models > Incrementals

Hayne MLE Mean Incrementals – Berquist Sherman, Incremental Severity

Accident Year	12	24	36	48	60	72	84	96	108	120
12-2004	37,612	91,714	63,883	38,976	21,638	10,500	3,597	1,742	590	236
12-2005	41,821	101,859	71,118	43,436	24,021	11,698	4,003	1,939	656	264
12-2006	44,242	108,041	75,201	45,884	25,387	12,343	4,230	2,044	694	279
12-2007	48,422	118,241	82,131	50,345	27,821	13,526	4,646	2,240	761	306
12-2008	49,922	121,858	84,524	51,885	28,661	13,930	4,789	2,311	783	314
12-2009	53,857	131,221	91,421	55,898	30,864	15,057	5,151	2,493	845	340
12-2010	59,104	143,788	100,257	61,344	33,941	16,510	5,647	2,731	927	373
12-2011	62,088	151,395	105,355	64,312	35,680	17,363	5,952	2,867	976	392
12-2012	64,187	156,500	109,015	66,655	36,890	17,927	6,154	2,974	1,007	405
12-2013	66,606	161,997	113,064	68,983	38,194	18,567	6,366	3,083	1,042	421

Hayne MLE Standard Deviation Incrementals – Berquist Sherman, Incremental Severity

Accident Year	12	24	36	48	60	72	84	96	108	120
12-2004	4,754	11,100	7,882	4,853	2,830	1,452	553	290	113	56
12-2005	5,075	11,783	8,340	5,256	3,047	1,571	588	306	122	60
12-2006	5,175	12,115	8,539	5,430	3,133	1,606	609	324	126	63
12-2007	5,393	12,534	8,898	5,582	3,306	1,717	651	343	135	67
12-2008	5,474	12,952	9,198	5,789	3,383	1,756	664	357	139	70
12-2009	5,736	13,438	9,699	6,073	3,583	1,862	704	375	149	74
12-2010	6,100	14,289	10,313	6,526	3,875	1,985	758	404	160	81
12-2011	6,356	14,967	10,748	6,747	4,032	2,090	797	421	167	84
12-2012	6,548	15,619	11,217	7,107	4,189	2,176	821	438	172	88
12-2013	6,877	16,301	11,652	7,582	4,374	2,287	864	462	182	92

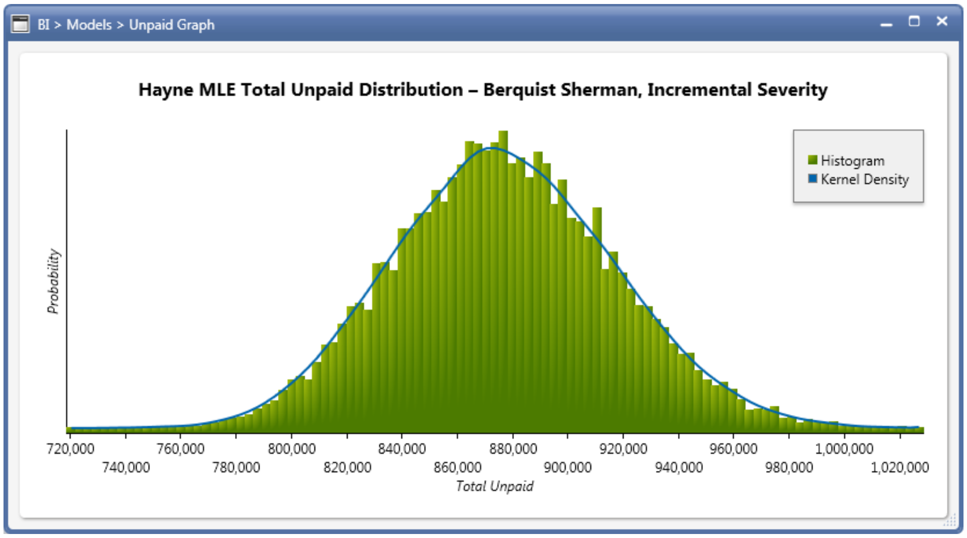
Table 7-6:
Estimated Incrementals by Accident Year by Development Period

Total Unpaid Distribution Graph

The final model output from the simulations is a histogram of the estimated unpaid amounts for the total of all accident years combined, as illustrated in Graph 7-4. The **Unpaid Graph**, or histogram, is created by dividing the range of all values from the simulation (using the maximum and minimum values) into one hundred "buckets" of equal size and counting the number of simulations that fall within each "bucket." Dividing by the total number of simulations (10,000 in this case) results in the frequency or probability for each "bucket" in the graph.

Since the simulation results often look "jagged" (as they do in Graph 7-4) a kernel density function is also used to calculate a "smoothed" line fit to the histogram values. The kernel density distribution is represented by the blue line in Graph 7-4.[51]

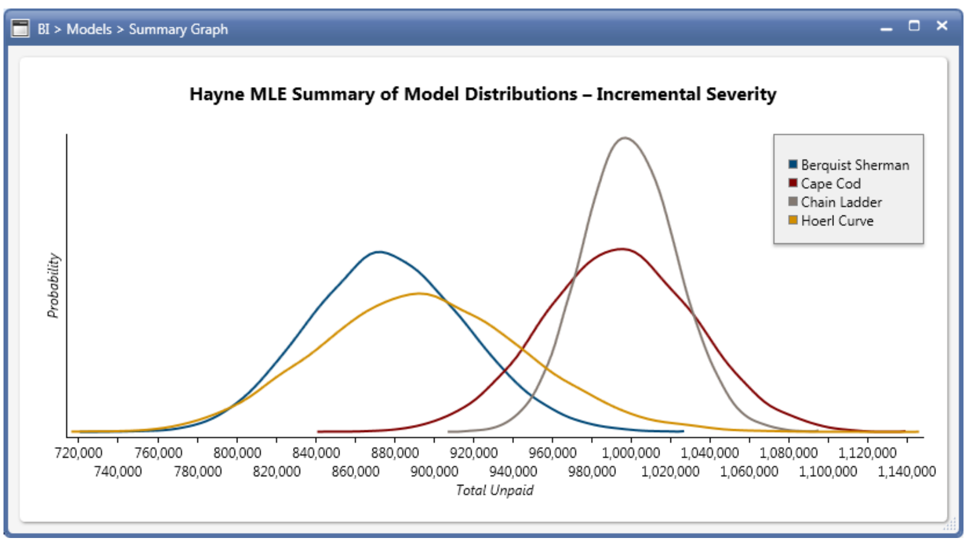
51 In simple terms, a kernel density function can be thought of as a weighted average of values "close" to each point in the "jagged" distribution with progressively less weight being given to values the further they are from the point being evaluated. For a more detailed discussion of Kernel density functions, see Wand & Jones, "Kernel Smoothing," Chapman & Hall. 1995.



Graph 7-4:
Total Unpaid Distribution

STEP 5: COMPARE GRAPHS FOR EACH MODEL

If you are using more than one of the Hayne models, after all of them have been reviewed, open the **Object Library** from the **HOME** ribbon, then in **Navigation Pane** within the **Object Library** select MODELS | HAYNE MLE | INCREMENTAL SEVERITY and select the **Summary Graph**, as illustrated in Graph 7-5. The **Summary Graph** is created by combining the kernel density graphs from each of the Incremental Severity models.



Graph 7-5:
Summary Graph

8. Evaluating Multiple Lines Together

OVERVIEW: ODP BOOTSTRAP AGGREGATION

In the **Navigation Pane**, select the ODP BOOTSTRAP AGGREGATION to aggregate the final results of all the LOBs (or segments) together, taking into account the estimated correlation among them.[52] This section of the **Navigation Pane** has two collections.

Assumptions | Correlation

- **User Selected** Rank Correlation Matrix

  This table is the correlation matrix and degrees of freedom used in calculating the correlated aggregate results. This is where you select, edit, and/or enter the matrix to use when you are aggregating distributions.

  Also shown at the bottom of this table are the degrees of freedom for the T-distribution used in the correlation, which affects the strength of the correlation in the tail of the distributions. The degrees of freedom range between 1 and 99. 99 will give a correlation based on a normal distribution. As the degrees of freedom move closer to 1, the model will give a correlation based on a fatter tailed T-distribution. Having a fatter tail means that you expect stronger correlations in extreme outcomes, for a given level of correlation.

- Rank Correlation of **Simulated** Results

  This table shows the correlation values from the simulation results; they provide some confidence that the simulated results match your intended correlation.

- **Calculated** Rank Correlation of Residuals

  These are tables of the rank correlation based on the input data, both before and after adjustment for heteroscedasticity. These matrices are provided to help you select the correlation to use in calculating aggregate distributions. Each matrix of residual correlation factors also has a corresponding matrix of p-values; these provide a measure of the statistical significance of each correlation coefficient.

  The correlation statistics will be different depending on the **Estimate Correlation Using** option you select on the **Model Options** dialog. The default is **MLE Copula**, which uses a maximum likelihood estimation copula to solve for all correlations at once, including the degrees of freedom for the T-distribution. The **Pairwise** option calculates the correlation between each pair of LOBs, but does not include a calculation of the degrees of freedom for the T-distribution, so a default value of 99 is entered as part of the output.

Results | Aggregate

- This collection has all of the same tables and graphs of results as the individual model collections. The amounts here, however, are the totals of the weighted (and "shifted" if turned on) values for all segments, taking into account the effect of the **User Selected** Rank Correlation Matrix.

52 Aggregation is used to combine only the weighted and/or shifted results for each segment. Thus, since weighting and shifting is currently only applicable for the ODP Bootstrap models the Mack Bootstrap and Hayne MLE models are not available for aggregation. When the Mack Bootstrap and Hayne MLE models are available for use in weighting and shifting (i.e., in a future release of Arius), they will also be included in the aggregation.

After you have analyzed each segment separately, it's time to look at the combined total of all lines in the project. But as was noted earlier, it is not enough to simply add all the results together. In most cases, this will result in an unrealistic total distribution of unpaid claims, depending on the level of correlation among the various segments. In general, the aggregate distribution of unpaid claims can be materially narrower than the sum of the individual distributions, when the aggregate accounts for the effect of correlation between the segments. This difference between the correlated aggregate and the sum of the segments will not be as material in cases where the segments are all strongly positively correlated, where there is little variability in the individual distributions, or where one segment is far larger than the rest.

In general, the process is as follows:

- Complete the analysis of every segment, including model weights and selected ("shifted") unpaid
- Run Diagnostics for all Segments & Correlation
- Select a correlation matrix to use in deriving the correlated aggregates
- Run Simulations for all Segments & Aggregation

STEP 0: PREPARE EACH SEGMENT FOR CORRELATION CALCULATIONS

The model calculates a series of correlation matrices for you to analyze or choose from. However, for the model to calculate these most effectively, each line of business should first be analyzed to make sure you have the best models for each line, as well as the weighted (and, if used, "shifted") "best distribution."

STEP 1: RUN ALL DIAGNOSTICS

From the **HOME** ribbon, click on the **RUN DIAGNOSTICS** icon and select the RUN DIAGNOSTICS FOR ALL SEGMENTS & CORRELATION option. This will do two things:

- it runs the diagnostics for all the segments in your project (to insure they are all up to date); and
- it calculates four different correlation matrices that will then be available for review in the **Correlation** collection.

The four correlation matrices will be different depending on the **Estimate Correlation Using** option you select on the MODEL OPTIONS dialog. The default is the **MLE Copula** option, which uses a maximum likelihood estimation copula to solve for all correlations at once, including the degrees of freedom for the T-distribution. In general, this option will tend to give a more robust solution since it is analyzing all of the data at once. However, it can be less than ideal when data is not used or missing for one or more segments. For example, if you are only using two year average age-to-age ratios for one segment, then only the data for the last three diagonals can be used in the estimation process. The maximum likelihood copula must only use data points that are common for every segment, so it is possible to have a situation where there is no data points common to every segment and the option will not work at all (e.g., one segment is in run-off for past 6 years [with no data for those years] and another only started up 5 years ago [with no prior data]).

Note:
Whenever you select anything less than the all-year average for the link ratios or exclude outliers, some of the residuals (that would otherwise be included) will be excluded from both the calculations for that LOB and the correlation matrix calculations.

The **Pairwise** option calculates the correlation between each pair of LOBs, but does not include a calculation of the degrees of freedom for the T-distribution, so a default value of 99 is entered as part of the output. The advantage of this option is that the common data is only a requirement for each pair of segments, so even if a smaller amount of data is present for one segment it will only impact the correlation calculations for that one segment and not the rest of the pairs.

STEP 2: SET UP YOUR CORRELATION MATRIX

The system needs a correlation matrix so it can factor in the effects of correlation among all your lines of business in the aggregate distribution. You must provide this information, based on your knowledge of the businesses involved and on the information you glean from the diagnostic matrices.

In the **Navigation Pane**, select the ODP Bootstrap Aggregation | Assumptions | Correlation collection. This collection contains three different correlation matrix windows:

- **User Selected** – where you enter or select your correlation factors
- **Simulated** – a check on your selections
- **Calculated** – tools to help you arrive at your selections

Specifically:

1. The **User Selected** Rank Correlation Matrix, illustrated in Image 8-1, contains the factors used to induce correlation in the simulated aggregate distribution.

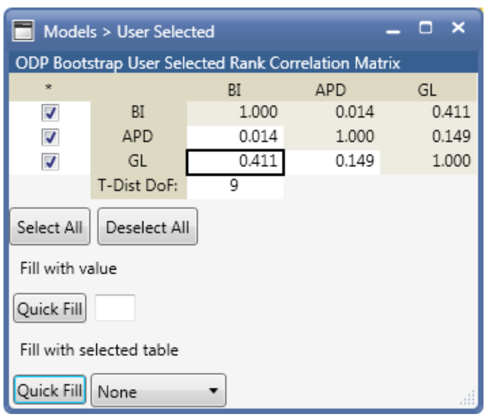


Image 8-1:
User Selected Rank Correlation Matrix

   - Part of this matrix is a data entry area, signified by the white background.
   - The remainder of the matrix is a mirror of that data entry area, with 1's down the center diagonal; this area is not available for data entry.
   - Each row or column represents a segment in the project file, and wherever rows and columns intersect the factor representing the expected correlation between those two lines is shown (thus the diagonal of 1's down the center, as each line is perfectly correlated with itself).

2. Enter the factors to represent the expected correlation between the lines of business. These are decimal numbers from 1 to -1, completely positively correlated to completely negatively correlated.

   The easiest way to do this is typically with one of the **Quick Fill** buttons. This allows you to automatically fill the matrix with a single value or with results from the system's various diagnostic calculations as illustrated in Image 8-2. You can choose to fill the matrix with:

   - **Fill with value** – i.e., the same value for every cell

Note:
Some correlation matrices are theoretically not possible. For example, it is impossible for more than two lines of business to all be 100% negatively correlated with each other – i.e., a matrix of -1's. When an impossible matrix is entered, the system will automatically adjust the matrix so that it will work.

- **Fill with selected table** – you can select:
  - values calculated from paid residuals before or after any heteroscedasticity adjustments,
  - values calculated from incurred residuals before or after heteroscedasticity adjustments,

For example, a common approach might be to start with correlation factors based on paid data after heteroscedasticity adjustments; these are listed in the QUICK FILL drop down box as *CorrAfterHeteroPaid*.

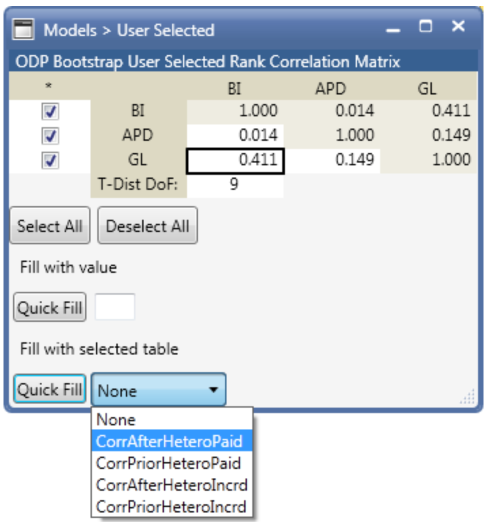


Image 8-2:
Quick Fill Drop Down Box

3. As with any similar decisions, you should review these factors to make sure they make sense based on your understanding of the data. For example, does it make sense for specific lines to be positively or negatively correlated, and if so, do the correlation factors in your matrix reflect that? You may not always want to change the factors, but you should at least review them.

   To help you assess the reasonableness of your chosen factors, with each of the four diagnostic correlation matrices the system also provides a corresponding matrix of p-values.

   - A large *p*-value (p>.05) indicates that the correlation is not significantly different than zero. Therefore, as a general rule, you could replace correlations with *p*-values >.05 with zeroes in your **User Selected** Rank Correlation Matrix unless you have a better way to estimate correlation between those lines.

4. In the **User Selected** Rank Correlation Matrix enter the **T-Dist DoF** to be used in the correlation process. The degrees of freedom effectively allows you to select between two different distributions for the correlation process:

   - The normal distribution is used if you set the degrees of freedom to 99, or
   - The T-distribution is used if you set the degrees of freedom to between 98 and 1, inclusive.
   - The normal distribution is symmetrical so it gives symmetrical weight to the entire distribution during the correlation process, whereas the T-distribution will progressively strengthen the weight given to the “tail” of the distribution as the degrees of freedom moves from 98 toward 1.

5. In the **User Selected** Rank Correlation Matrix check the segments that you want to include in the Aggregate results. By checking and unchecking different groups of segments, this allows you the ability to create sub-aggregations for different divisions, legal entities or strategic business units within a group.

Note:
If you selected one of the options (other than **No**) for **Save Results to File** in the **Model Options** dialog, then the model will take a little longer as it will also need to compile the data you requested.

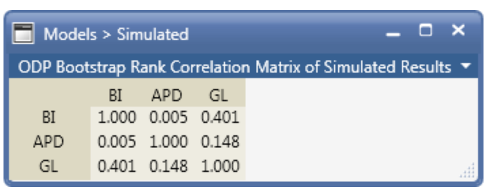
STEP 3: RUN THE CORRELATED SIMULATION AND REVIEW YOUR RESULTS

1. From the **Home** ribbon, click on the **Run Simulations** icon and select the Run Simulations for All Segments & Aggregation option. In a few minutes it will return with results for each segment and the aggregate results.

2. Review your results for reasonableness. One of the first checks is to review the Rank Correlation of **Simulated** Results matrix.

   The Rank Correlation of **Simulated** Results table (illustrated in Image 8-3) should generally appear very similar to the correlation factors you selected when running the model. During the simulation process, the system sorts the results based on the correlation factors you selected. It then measures the correlation again using the sorted results. The resulting simulation correlation matrix appears as the second window in the Correlation collection. It helps provide a reasonableness check of some of the model's calculations. If the factors in this matrix appear to be materially different from the factors you entered, then most likely either

   - you didn't run enough iterations for the correlation to have its full effect (generally at least 5,000), so change this and re-run, or
   - your selected correlation matrix doesn't work and the model corrected it to the closest matrix that will work.[53]



	BI	APD	GL
BI	1.000	0.005	0.401
APD	0.005	1.000	0.148
GL	0.401	0.148	1.000

Image 8-3:
Rank Correlation of **Simulated** Results

3. In the **Navigation Pane**, select the ODP Bootstrap Aggregation | Results | Aggregate collection. This collection contains an **Unpaid Table**, **Cash Flow**, **Run-off**, **Loss Ratios**, **CDR** and **Incrementals** tables as well as an **Unpaid Graph** for the aggregate of all lines you chose to include in the **User Selected** Rank Correlation Matrix. These are similar to the corresponding tables and graph for each segment. The mean of this distribution should be the same as the sum of the means of all the individual segments. However, as you look at the different percentiles, you will find divergence between the aggregate and the raw sum of the segments, with this difference being greater at the higher percentiles. These differences are illustrated in Graph C-1 in Appendix C.

Like with the individual lines, remember that the model can be run iteratively with different correlation factors, allowing you to fine-tune your results to best match your understanding of the underlying business and to test the impact of different correlation assumptions.

53 See the note on the prior page. In statistical terms a correlation matrix that "will work" is referred to as a positive semi-definite matrix, so if you entered a correlation matrix that is not positive semi-definite then the system will use an algorithm to find the positive semi-definite matrix that is closest to your matrix.

[THIS PAGE INTENTIONALLY LEFT BLANK]

9. Working with Unusual Data

Up to this point in the manual, the examples and illustrations have assumed that the data triangle and the exposures are completely symmetrical (i.e., the number of development periods and accident periods in the data triangle are identical and the lengths of all exposure periods are the same). Unfortunately, in many common analyses the exposures are not always symmetrical. Fortunately, however, the Milliman model can accommodate many common situations with non-symmetrical exposures.

STUB PERIOD TRIANGLES

One of the most common situations is the analysis of data with a stub period – i.e., where the last diagonal is not the same as the rest of the triangle. For example, with an annual x annual triangle the development intervals are 12/24/36/... for a year end analysis, but when an interim analysis is performed as of June 30 the last diagonal has development intervals of 6/18/30/… As noted in Section 4, from the **HOME** ribbon select the **PROJECT SETTINGS** icon which opens the PROJECT SETTINGS dialog box to the **Data Structure** tab (as illustrated in Image 9-1).

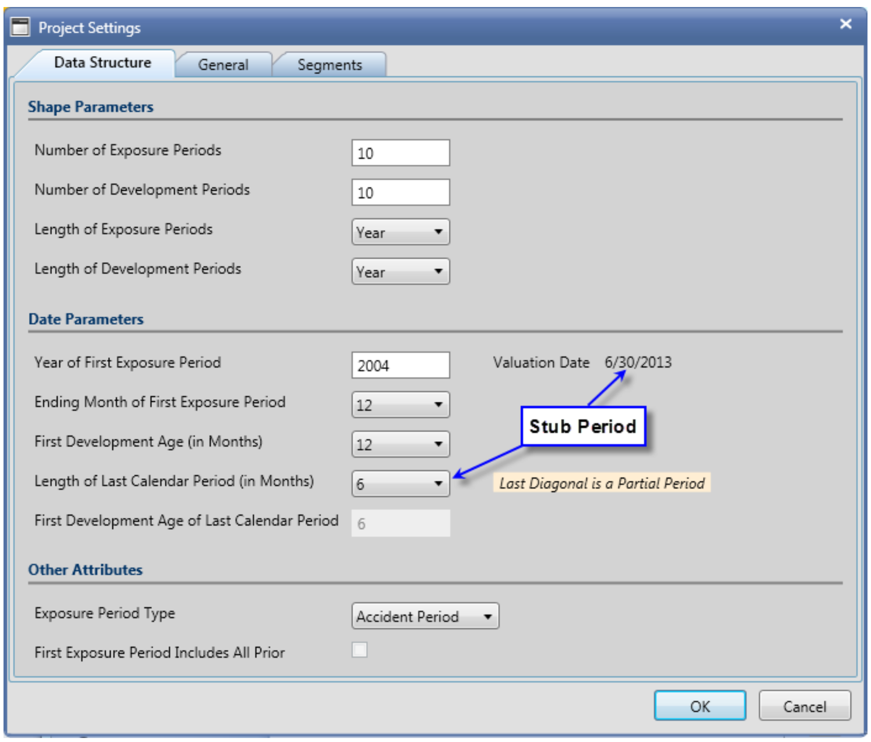


Image 9-1:
Project Settings Dialog Box (with Stub Period)

In order to specify the stub period, the Length of Last Calendar Period (in Months) will be different than the Length of Development Periods setting, but the First Development Age (in Months) will still be consistent with the Length of Development Periods setting. For example, in Image 9-1 the Length of Last Calendar Period (in Months) is set to 6 in order to specify a June 30 analysis while the overall data structure is based on annual x annual data.

With a stub period specified in the PROJECT SETTINGS dialog box, the model will automatically make several adjustments to the normal model simulation steps. In order to illustrate the adjustments, we will discuss the changes to the Paid Chain Ladder model simulation as described in Section 3 (and Appendix A). The changes to the other models are essentially the same so will not be illustrated in detail, although a few comments are included for completeness.

1. Use a triangle of cumulative paid losses as input. Calculate a triangle of age-to-age development factors.
    - When the last incremental diagonal is a stub period, actuaries ordinarily exclude this from the age-to-age calculations.
    - However, in Arius incremental values in the last diagonal are grossed-up to estimate a full period's worth of data and included in the age-to-age factor calculation.
    - The gross-up factors are calculated using assumptions of linear earnings and are therefore different for accident period and policy period data.
    - The gross-up factors are also adjusted for actual earnings by comparing Earned Premium to Ultimate Premium by year if included.
2. Calculate a new triangle of "fitted values" – i.e., use the age-to-age factors to "undevelop" each value in the latest diagonal to form a new triangle of values predicted by the model assumptions.
    - The fitted values use the grossed-up last diagonal and the age-to-age factor calculation from step 1 that includes the grossed-up diagonal.

3-4. Working from the incremental versions of these triangles, calculate a triangle of residuals using the fitted triangle and the original data. These are called "unscaled Pearson residuals" in the model. These are then standardized using the hat matrix adjustment factors resulting in "standardized Pearson residuals".

- The model uses the grossed-up last diagonal values so that all residuals represent the same amount of exposure.

5. Create a new incremental sample triangle by selecting randomly with replacement from among the triangle of standardized Pearson residuals.
    - The sample triangles include the grossed-up last diagonal values since they are based on the fitted values from step 2 and the residuals in step 4.
6. Develop and square that sample triangle, adding tail factors, and estimating ultimate losses.
    - The incremental values in the last diagonal for the sample triangle are "reduced" to match the exposures of the original stub period triangle.
    - The complements of the "reduced" incremental values in the last diagonal are added to the incremental values in the next future diagonal. Similarly, each future diagonal has a portion "reduced" to adjust for the exposure and added to the next future diagonal, so that the future incremental development periods match the exposures along the last historical diagonal. For example, if the last diagonal has development dates of 6/18/30/…, then after this "reallocation" process the future incremental values will be converted from 12/24/36/… development to 6/18/30/… development.
    - For the most recent accident period (or for several of the recent policy periods), the ultimate value represents a full period of exposure, so the future incremental values are "reduced" to exclude the exposure beyond the evaluation date using assumptions of linear earnings or actual earnings if Earned and Ultimate Premiums are included.

- For the Bornhuetter-Ferguson and Cape Cod models, the ultimate values are calculated using the Ultimate Premiums to match the grossed-up exposures of the last diagonal, then the exposure reduction portion of the algorithm is applied.

7. Add process variance to the future incremental values from Step 6 (which will change the estimated ultimate to a possible outcome).
   - The future incremental values will have the same development period length as the original data, but the starting point for the development will be "shifted" to match the exposures of the last diagonal. For example, for the analysis in Image 9-1, the future periods will be 18/30/42....
8. Calculate the total future payments (estimated unpaid amounts) for each year and in total for this iteration of the model.
   - The unpaid estimates are consistent with the "shifted" periods from step 7.
9. Repeat the random selection, new triangle creation, and resulting unpaid calculations in Steps 5 through 8, *X* times.

The result from the X simulations is an estimate of the distribution of possible outcomes. From this we can calculate the mean, standard deviation, percentiles, etc.

- The distributions are consistent with the "shifted" periods from step 7 and the "reduced" exposures for step 6.
- The loss ratio distributions use the Earned Premiums in order to match exposures of losses with revenue.

SHORT FIRST DEVELOPMENT PERIOD TRIANGLES

Another common situation is the analysis of data with a shortened first development period – i.e., where the development period in the first column is not the same as the rest of the triangle. For example, with an annual x annual triangle the development intervals are 12/24/36/... for a year end analysis, but when an interim analysis is performed as of June 30 the column development intervals of 6/18/30/... are used for all data when the first development column is shortened. As noted in Section 4, from the **HOME** ribbon select the **PROJECT SETTINGS** icon which opens the PROJECT SETTINGS dialog box to the **Data Structure** tab (as illustrated in Image 9-2).

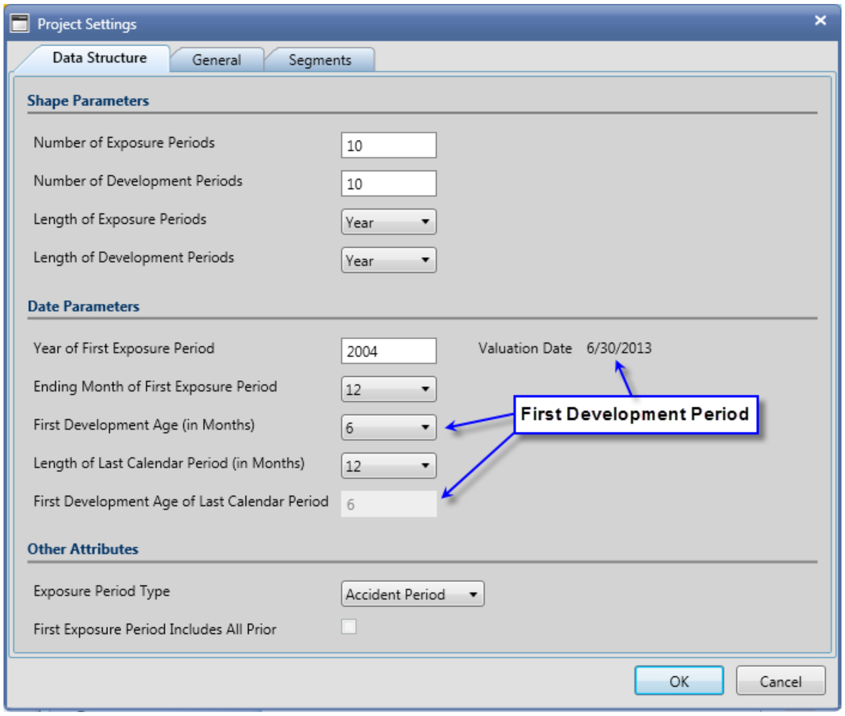


Image 9-2:
Project Settings Dialog Box (with short first development period)

Note:
The column headings in the development tables will reflect the data in the analysis (e.g., 6/18/30/...) since they are the same for all columns and rows.

In order to specify the short initial development period, the **First Development Age (in Months)** will be different than the **Length of Development Periods** setting, but the **Length of Last Calendar Period (in Months)** will still be consistent with the **Length of Development Periods** setting.[54] For example, in Image 9-2 the **First Development Period (in Months)** is set to 6 in order to specify a June 30 analysis based on annual x annual data with a short initial development period.

With a short initial development period specified in the PROJECT SETTINGS dialog box, the model will automatically make several adjustments to the normal model simulation steps. In order to illustrate the adjustments, we will discuss the changes to the Paid Chain Ladder model simulation as described in Section 3 (and Appendix A). The changes to the other models are essentially the same so will not be illustrated in detail, although a few comments are included for completeness.

1. Use a triangle of cumulative paid losses as input. Calculate a triangle of age-to-age development factors.
   - No adjustments are required for this step.
2. Calculate a new triangle of "fitted values" -- i.e., use the age-to-age factors to "undevelop" each value in the latest diagonal to form a new triangle of values predicted by the model assumptions.
   - No adjustments are required for this step.

3-4. Working from the incremental versions of these triangles, calculate a triangle of residuals using the fitted triangle and the original data. These are called "unscaled Pearson residuals" in the

54 As a technical note, it is possible to specify **both** a short initial development period **and** a stub period along the last diagonal by specifying a **Length of Last Calendar Period (in Months)** that is less than the **First Development Age (in Months)**. The Milliman Arius model will work in this situation and will adjust for both of these as described in the two subsections in Section 9, but we will not describe it as a separate unusual situation since it is far less common.

model. These are then standardized using the hat matrix adjustment factors resulting in "standardized Pearson residuals".

- No adjustments are required for these steps.

5. Create a new incremental sample triangle by selecting randomly with replacement from among the triangle of standardized Pearson residuals.

   - No adjustments are required for this step.

6. Develop and square that sample triangle, adding tail factors, and estimating ultimate losses.

   - For the most recent accident period (or for several of the recent policy periods), the ultimate value represents a full period of exposure, so the future incremental values are "reduced" to exclude the exposure beyond the evaluation date using assumptions of linear earnings or actual earnings if Earned and Ultimate Premiums are included.

   - For the Bornhuetter-Ferguson and Cape Cod models, the ultimate values are calculated using the Ultimate Premiums to match the grossed-up exposures of the last diagonal, then the exposure reduction portion of the algorithm is applied.

7. Add process variance to the future incremental values from Step 6 (which will change the estimated ultimate to a possible outcome).

   - The future incremental values will have the same development period length as the original data, with the starting point for the development consistent with the original exposures. For example, for the analysis in Image 9-2, the future periods will be 18/30/42....

8. Calculate the total future payments (estimated unpaid amounts) for each year and in total for this iteration of the model.

   - The unpaid estimates are consistent with the original exposure periods from step 7.

9. Repeat the random selection, new triangle creation, and resulting unpaid calculations in Steps 5 through 8, *X* times.

The result from the *X* simulations is an estimate of the distribution of possible outcomes. From this we can calculate the mean, standard deviation, percentiles, etc.

- The distributions are consistent with the original exposure periods from step 7 and the "reduced" exposures from step 6.

- The loss ratio distributions use the Earned Premiums in order to match exposures of losses with revenue.

MISSING VALUES

Another common data issue is missing values. For example, while the PROJECT SETTINGS dialog specifies the "dimensions" for every triangle in the analysis for every segment, you could have started writing business for some segments after the rest or some might be discontinued and in runoff. Other common issues resulting in missing values are calendar periods that start after the initial accident (or policy) period, which results in a "missing triangle" in the upper left corner, or simply having a few missing values somewhere in the middle of the triangle.

If the dataset has missing values, then these should be left blank. **Do not** fill them in with zeroes, unless they are truly a zero value. As noted in Section 5, blank cells are acceptable anywhere in the triangle except on the most recent two diagonals. The Arius system understands that a blank cell is different than a zero and functions accordingly.

MISSING WHOLE COLUMNS

The model must be able to calculate development factors at each point in the triangle. Therefore, whole missing columns are not allowed for the ODP Bootstrap and Mack Bootstrap models. However, the Hayne MLE models will work with a missing column.

WHOLE COLUMNS OF ZEROES

The model can handle columns of zeroes only if they are sequential and start from the first development column. For example, there cannot be a column of zeroes in the middle of your cumulative triangle, with non-zero columns on either side. An incremental column of zeroes is acceptable in the middle of a triangle.

10. Frequently Asked Questions

WHAT CAN CAUSE A SIMULATION TO RUN UNSUCCESSFULLY?

My simulation appeared to run. Why did my Unpaid Table, Cash Flow, etc. exhibits remain empty?

I thought I filled in all the necessary data for the model. So why did I receive a "Job Failed" message when I simulated a line of business?

A number of issues or data interrelationships can prevent the simulation from successfully completing. When a model fails to run successfully, you should receive descriptive information back from the system.

Note:
Running the simulations will NOT update your diagnostics.

In addition, here are some of the more common causes of models failing to run:

1. What you MUST have for ALL models to run:
   - All cells in the Heteroscedasticity Adjustment Groups table must contain values.
     - For your first test simulation you can run with all zeroes, just to calculate an initial set of diagnostics.
   - You must enter a Mean Tail Factor for whichever model you are simulating with either the ODP or Mack bootstrap models (Hayne does not currently include tail extrapolation).
   - You need both Paid and Incurred data and options filled in if you are simulating based on Incurred data.
2. If Extrapolate Tail Factor is checked, then
   - Exponential Decay Factor must be provided
3. If Enable Exposure Adjustment is checked, then
   - Ultimate Exposure data must be provided, and
   - if you have a blank cell in your Ultimate Exposure data, you cannot have data on that same row in your Paid or Incurred triangles.
4. It is impossible for the model to calculate Degrees of Freedom (DoF) with very small triangles (e.g., 4 x 4 or smaller). It also may not be able to calculate the DoFs if you have larger triangles (e.g., 6 x 6) but have excluded several residuals as part of your diagnostic analysis.
5. Hetero groups must have at least 2 residuals that are not selected as outliers and are not zero. For instance, the last development period will always have one residual of zero. This cannot be selected as a group by itself.

IS THERE A DIFFERENCE BETWEEN RUNNING DIAGNOSTICS AND RUNNING SIMULATIONS?

Yes. The analysis process contains two distinct phases, and these two system operations reflect these different phases. You must use both diagnostics and simulation functions to get to a final distribution estimate.

Note:
The simulations should run using the default settings, but it will NOT update your diagnostics.

1. **Run Diagnostics for *Segment*** and **Run Diagnostics for All Segments & Correlation** are essential steps in preparing your model for simulation; *they do not perform the simulation*. They provide you the necessary information to allow you to evaluate the quality of your model, to understand how well your data fits the models, and to adjust the models for the optimal possible fit. They fill

the DIAGNOSTICS collection tables and graphs, providing incremental data, age-to-age factors and averages, residuals, and the graphics that allow you to identify potential adjustments before you run a simulation.

Note that **Run Diagnostics for All Segments & Correlation** also calculates the correlation between the various lines of business, and returns various correlation matrices to the ODP Bootstrap Aggregation | Assumptions | Correlation collection.

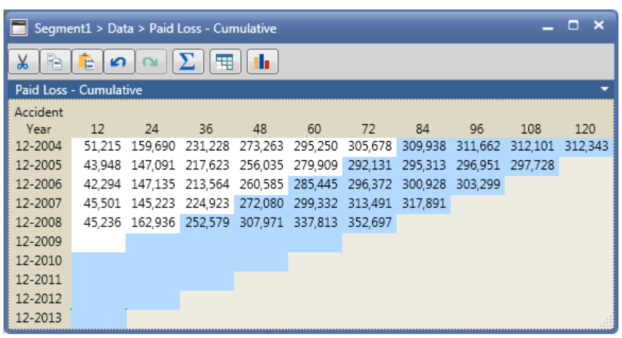
2. The simulation functions – **Run Simulations for *Segment*** and **Run Simulations for All Segments & Aggregation** – perform the actual stochastic projections to ultimate (or time horizon, if selected), and return results to the Arius system. *You will use the Simulation function(s) after you have reviewed the results of the diagnostic calculations and understand the quality of the model fit with your data.* The Simulation functions only return results to the model specific collections by segment (for the selected models) and Aggregate Results collection. They do not return values to any of the DIAGNOSTICS collection tables or graphs.

Note: Running the simulatios will NOT update your diagnostics.

CAN I MODEL TRIANGLES OF DIFFERENT SIZES & SHAPES TOGETHER IN THE SAME PROJECT?

Yes, as long as the triangles have the same 'as of' dates and the same stub periods or short first period, if any. They must also all have the same initial development period and the same development length (e.g., if one LOB is 9-21-33 development, all LOBs should contain this type of data). For instance, you can have a 10x10 triangle and a 5x5 triangle in the same workbook. The top 5 rows of the segment with the 5x5 triangle will be blank. When specifying the inputs for a new project you should always define the shape based on the size of the largest triangle.

If you have a 10x10 triangle and a 5x10 triangle (e.g., in run-off) then you would specify inputs as 10x10 triangles. If you do that, you have to remain cognizant of the implications on the model. For example, if you specify a weighted average of the last 3 years, then the shaded cells in blue will be factored into the calculations, as illustrated in Image 10-1.

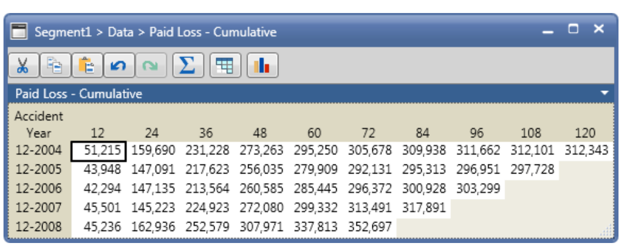

Segment1 > Data > Paid Loss - Cumulative

Paid Loss - Cumulative

Accident Year	12	24	36	48	60	72	84	96	108	120
12-2004	51,215	159,690	231,228	273,263	295,250	305,678	309,938	311,662	312,101	312,343
12-2005	43,948	147,091	217,623	256,035	279,909	292,131	295,313	296,951	297,728	
12-2006	42,294	147,135	213,564	260,585	285,445	296,372	300,928	303,299		
12-2007	45,501	145,223	224,923	272,080	299,332	313,491	317,891			
12-2008	45,236	162,936	252,579	307,971	337,813	352,697				
12-2009										
12-2010										
12-2011										
12-2012										
12-2013										

Image 10-1: Triangle in run-off. Shaded cells used in 3-year average age-to-age ratios.

To get around this, select an all year weighted average, or specify a 5x10 triangle in the inputs. However, this prevents the model from also containing the 10x10 triangle.


Segment1 > Data > Paid Loss - Cumulative

Paid Loss - Cumulative

Accident Year	12	24	36	48	60	72	84	96	108	120
12-2004	51,215	159,690	231,228	273,263	295,250	305,678	309,938	311,662	312,101	312,343
12-2005	43,948	147,091	217,623	256,035	279,909	292,131	295,313	296,951	297,728	
12-2006	42,294	147,135	213,564	260,585	285,445	296,372	300,928	303,299		
12-2007	45,501	145,223	224,923	272,080	299,332	313,491	317,891			
12-2008	45,236	162,936	252,579	307,971	337,813	352,697				

Image 10-2: Triangle size 5x10.

HOW ARE TAIL FACTORS APPLIED IN MY MODEL?

For each iteration, the first step is to sample a random tail factor using the distribution, mean and standard deviation specified in the **Tail Factor** window in the DIAGNOSTICS collection, assuming random tail factors are enabled (i.e., the **Enable Tail Factor Distribution** option is checked). Otherwise, the mean tail factor is used in each iteration. Then, if both a minimum and/or maximum value are entered and the limits are enabled (i.e., the **Limit Tail Factor with Min/Max** option is checked), the sampled tail factor is compared to the minimum and/or maximum and limited if needed.

Next, if extrapolation is enabled (i.e., the **Extrapolate Tail Factor** option is checked) the decay rate is applied to extrapolate the limited sampled tail factor up to the **Number of Periods in Extrapolation** and calculate the incremental age-to-age factors as illustrated in Section 5 and Image 5-8.[55] These [random], [limited], [extrapolated] age-to-age factors are applied along with the calculated average age-to-age factors in the appropriate step(s) for each model to square the triangle and project an ultimate point estimate.

Finally, process variance is added using the selected option to convert all future incremental values to possible outcomes. Note that the hetero group selected for the last development period in the triangle is applied to all the future periods that use the [random], [limited], [extrapolated] tail factor.

For the **ODP Process** and **ODP Residual** algorithms, the resulting possible outcomes (described just above) are used up to the Nth future diagonal when calculating updated average age-to-age factors, and any remaining [random], [limited], [extrapolated] age-to-age factors beyond the Nth future diagonal, if any, are used to calculate the remaining point estimates.

For the **Mack Bootstrap Time Horizon** algorithm, the [random], [limited], [extrapolated] age-to-age factors are used within the diagonal by diagonal iterative process up to the Nth future diagonal and any remaining [random], [limited], [extrapolated] age-to-age factors beyond the Nth future diagonal, if any, are used to calculate the remaining point estimates.

HOW DO I EXCLUDE AN OUTLIER FROM MY MODEL?

You can potentially improve the fit of your model by eliminating certain outliers from the simulation process in any of the ODP bootstrap models.

Outliers in the residuals can be identified in either the **Residual Graphs**, in the **Normality** Plot or in the Box-Whiskers plot. Though they are most clearly identified in the Box-Whisker Plot (that’s the main reason for this plot) if a point is a significant outlier, it will show up clearly in all three diagnostic exhibits.

If a point is both materially outside the range of the other residuals, and you want to give it no weight in the model’s calculations, you should identify that point to the system.

- It is generally advisable to identify hetero groups first and rerun the diagnostics. This step will often eliminate potential outliers, or for example may eliminate two of your three outliers, leaving you with only one outlier to evaluate for potential elimination from the model.
- Find the potential outlier(s) in the Box-Whiskers plot **After Hetero**scedasticity adjustments. You can find that same cell(s) in the **Residual Graphs** and simply click on the dot in one of the four plots, which will change the color from green to red. Clicking on a dot will place a “1” in the

55 The illustration in Section 5 and Image 5-8 is for the mean, but the calculation for each iteration will be based on the [random], [limited] tail factor.

corresponding cell in the **Outliers** table. Conversely, you can directly change the value in the appropriate cell in the **Outliers** table and it will "mark" the dots in the **Residual Graphs** by changing their color to red (or green).

- Eliminate the effects of this point on the model by running diagnostics to recalculate all of the diagnostics statistics after giving zero weight to the identified outliers.

Note that outliers should be reviewed carefully, and should generally only be removed from the model when they represent an occurrence that would not be assumed to happen again with this data set. Certain outlier-type events are inherent in the nature of many insurance lines and the model is often more realistic when it includes the additional skewness that their presence provides.

BLANK CELLS

Are blank cells treated as zeroes in the model?

What do I need to understand about blank cells?

Blanks are treated differently in different places within the model, depending on the most appropriate and expected definition in each specific context.

- The model's residual calculations are based on incremental data. When there is a blank cell in the cumulative triangle, the model cannot determine what the related incremental value is for that cell nor can it determine the incremental value in the next cell to the right. Thus both cells are left blank in both the incremental triangle and the residual triangles.
- Blanks are treated the same as zeroes in the **Outliers** triangle.
- A blank cell is not allowed in either the most recent or next-to-most recent diagonal if there is any other data elsewhere on that same row in the triangle.
- Blanks are not treated as zeroes in LDF calculations. If you are using a 5-year average, and there is a blank cell among the most recent five values, the system will use only four values; it will not reach back and pick up another value to make up five, and it will not count the blank as a zero and keep the same number (5) as the denominator.
- Also, extending the above item, if you have data for a run-off triangle (say, 24 columns x 20 rows, with the bottom four rows missing) and you place that data into a fully dimensioned triangle space (say, 24 x 24), the bottom four cells in the first four columns will be blanks. However, the LDFs for this triangle will be calculated as if it were a 24x24 triangle instead of a triangle in run-off. This is not an issue if you are selecting an all year weighted average. To get the correct development factor calculations, if at all possible you should define the size and shape of the data triangle to match the data you will be using. If you are entering 24x20 runoff data, then you should specify a 24x20 triangle in the PROJECT SETTINGS dialog.

WHEN I PASTE A TRIANGLE INTO ARIUS I LOSE FORMATTING FROM THE ORIGINAL TEMPLATE. HOW DO I PREVENT THIS?

Whenever pasting data into the Arius tables from other sources, none of the original formatting is retained to make sure that Arius' formatting stays intact.

CAN INPUT DATA BE LINKED TO ARIUS FROM ANOTHER FILE, RATHER THAN ENTERED DIRECTLY INTO A TRIANGLES?

No, the data entry areas can't be linked to cells in other programs such as an Excel. However, the Application Programming Interface (API) can be used to pull data into Arius from another source or push data or results back out to another source. In addition, Arius can import data and other model settings from either ReservePro or RVM.

HOW DO I REDUCE PROCESSING TIME?

Model Weights and set Save Results to File both require significant computer resources to run. If you can remove those selections from at least some runs, those runs will be noticeably faster than if the options are turned on.

HOW DO I GETTING A DISTRIBUTION OF IBNR INSTEAD OF UNPAID LOSS?

The ODP bootstrap's Incurred Model is designed to output a distribution of total unpaid, including case reserves, so that it is comparable to the paid model. If you want to simulate a distribution of IBNR (i.e., unpaid excluding case reserves) you can enter the incurred data in the paid data input array and run the paid model with incurred data.

[THIS PAGE INTENTIONALLY LEFT BLANK]

Appendices

A. Examples of the Basic Calculations

A WALKTHROUGH OF THE BASIC CALCULATION BASED ON PAID LOSS DATA

STEP 1.
BUILD A BASIC DEVELOPMENT MODEL.

Use the standard chain-ladder method and the all-period volume-weighted average (VWA) to calculate age-to-age factors.

CUMULATIVE PAID LOSS DATA

	12	24	36	48	60
2009	352	783	1,045	1,183	1,295
2010	255	572	710	750	
2011	279	638	767		
2012	311	717			
2013	308				

	12-24	24-36	36-48	48-60	60+
VWAs	2.264	1.265	1.101	1.095	1.000

STEP 2.
CREATE A "FITTED" TRIANGLE.

Start with the most recent cumulative diagonal and "un-develop" backwards using the appropriate VWAs.

TRIANGLE FITTED BACKWARDS FROM LATEST DIAGONAL

	12	24	36	48	60
2009	375	849	1,074	1,183	1,295
2010	238	538	681	750	
2011	268	606	767		
2012	317	717			
2013	308				

STEP 3.
CALCULATE PEARSON RESIDUALS.

Working from the incremental forms of both triangles, subtract the fitted from the actual amount for each cell. Divide each result by the square root of the fitted amount.[56] This results in "unscaled Pearson residuals."

$$r_{UP} = \frac{C - \hat{m}}{\sqrt{abs(\hat{m})}}$$

C = actual incremental amount

- $\hat{m}$ = fitted incremental for matching location in each triangle

UNSCALED PEARSON RESIDUALS

	12	24	36	48	60
2009	-1.18	-1.97	2.45	2.78	0.00
2010	1.12	0.96	-0.40	-3.50	
2011	0.69	1.12	-2.51		
2012	-0.32	0.28			
2013	0.00				

56 While the theory uses the square root of the fitted amount, by using the square root of the absolute value of the fitted amount we can include negative incremental amounts without any material impact.

STEP 4a. STANDARDIZE THE RESIDUALS.

Multiply each unscaled residual by the hat matrix adjustment factor.[57]

$$f_{ij} = \frac{1}{\sqrt{1-h_{ij}}}$$

h_{ij} = the diagonal of the hat matrix H

$H = X(X^TWX)^{-1}X^TW$

X = the design matrix from the Generalized Linear Model (GLM)

W = the weight matrix from the GLM

HAT MATRIX ADJUSTMENT FACTORS

	12	24	36	48	60
2009	1.4922	1.6088	1.4725	1.6849	1.0000
2010	1.3675	1.4675	1.3033	1.3415	
2011	1.4172	1.5350	1.3485		
2012	1.5606	1.7546			
2013	1.0000				

STANDARDIZED PEARSON RESIDUALS

	12	24	36	48	60
2009	-1.76	-3.17	3.60	4.69	0.00
2010	1.54	1.40	-0.53	-4.69	
2011	0.98	1.72	-3.39		
2012	-0.50	0.50			
2013	0.00				

STEP 4b. CALCULATE THE SCALE PARAMETER.

N = number of observations (less outliers)

p = number of parameters in the model (typically the number of columns in the data triangle + the number of rows in the data triangle + the number of additional hetero groups – 1)

ADDITIONAL STATISTICS CALCULATED

Pearson chi-squared statistic	=	sum of squares of unscaled Pearson residuals
	=	41.82
Degrees of freedom	=	# of observations in the model minus # of parameters (# columns + # rows – 1)
	=	15 – 9 = 6
Scale parameter	=	chi-squared statistic ÷ degrees of freedom
	=	6.97

57 See Pinheiro, Paulo J. R., João Manuel Andrade e Silva and Maria de Lourdes Centeno. 2001. Bootstrap Methodology in Claim Reserving. ASTIN Colloquium. As a technical issue, our model uses matrix decomposition rather than matrix multiplication, which is numerically equivalent yet more stable.

SIMULATION STEPS:

STEP 5. RANDOMLY CREATE A NEW TRIANGLE OF "SAMPLE" DATA.

5a. Build a new triangle by randomly selecting (with replacement) from among the non-zero standardized Pearson residuals in Step 4a.[58]

5b. Create a triangle of sample data based on the randomly selected residuals. For each cell:

$$C' = r_{SP}\sqrt{abs(\hat{m})} + \hat{m}$$

RANDOMLY SELECTED RESIDUALS

	12	24	36	48	60
2009	1.54	-0.50	-3.39	-0.50	1.54
2010	-3.17	-3.17	1.54	-4.69	
2011	-0.53	4.69	-0.50		
2012	1.40	1.54			
2013	4.69				

SAMPLE INCREMENTAL TRIANGLE CALCULATED BASED ON THE RANDOM RESIDUALS

	12	24	36	48	60
2009	405	463	174	104	128
2010	189	246	161	30	
2011	259	425	155		
2012	342	431			
2013	390				

STEP 6. COMPLETE THE NEW RANDOMLY-GENERATED TRIANGLE.

Calculate new VWAs and use them to complete the bottom right of the triangle.

NOTE: A randomly generated tail factor could also be applied here to extrapolate future development periods – i.e., beyond 60 months in this example.

SAMPLE CUMULATIVE TRIANGLE

	12	24	36	48	60
2009	405	868	1,042	1,146	1,274
2010	189	434	596	626	
2011	259	684	838		
2012	342	773			
2013	390				

	12-24	24-36	36-48	48-60	60+
VWAs	2.310	1.247	1.082	1.112	1.000

COMPLETED CUMULATIVE TRIANGLE WITH FUTURE EXPECTED PAYMENTS

	12	24	36	48	60
2009	405	868	1,042	1,146	1,274
2010	189	434	596	626	696
2011	259	684	838	907	1,008
2012	342	773	963	1,042	1,159
2013	390	902	1,124	1,216	1,352

58 As noted in Section 3, Step 5, this is the default option for randomly simulating a new triangle with the same statistical properties as the original data. Other options for simulating a new triangle are described in Section 5, but are not illustrated here.

STEP 7.
INTRODUCE PROCESS VARIANCE.

Calculate the incremental payments in the future payment stream from the cumulative completed triangle.

To add *process variance* in the simulation, replace every future incremental paid amount with a randomly selected point from a gamma distribution[59] where:

Mean = the incremental paid loss amount

Variance = Mean x Scale

Parameter (see Step 4b)

COMPLETED INCREMENTAL TRIANGLE

	12	24	36	48	60
2009	405	463	174	104	128
2010	189	246	161	30	70
2011	259	425	155	68	102
2012	342	431	191	79	117
2013	390	511	223	92	136

RANDOMLY GENERATED FUTURE INCREMENTAL PAYMENTS BASED ON ABOVE

	12	24	36	48	60
2009					
2010					68
2011				58	80
2012			194	84	127
2013		469	235	79	160

STEP 8.
CALCULATE TOTAL UNPAID AMOUNTS.

Sum the future incremental values to estimate the total unpaid loss by period and in total (in this example, 1,556).

This provides *one* estimated possible outcome.

TOTAL ESTIMATED FUTURE INCREMENTAL PAYMENTS

	12	24	36	48	60	TOTAL
2009						
2010					68	68
2011				58	80	138
2012			194	84	127	406
2013		469	235	79	160	944
						1,556

STEP 9.
REPEAT AND SUMMARIZE.

Repeat Steps 5 through 8 the specified number of times (e.g., 10,000 or more), capturing the resulting cash flows and the unpaid amounts by period and in total for each iteration. The resulting distribution of unpaid amounts from all the iterations can be used to calculate means, percentiles, etc.

TOTAL SIMULATED RESULTS

	MEAN UNPAID	STANDARD ERROR	COEFFICIENT OF VARIATION
2009	0	0	0.0%
2010	71	30	42.3%
2011	159	47	29.4%
2012	379	78	20.6%
2013	763	159	20.9%
	1,372	213	15.6%

59 As a technical note, the gamma distribution is used as an approximation to the over-dispersed Poisson. The use of volume weighted average age-to-age factors is derived from the GLM assumption of an over-dispersed Poisson distribution, but the gamma is a very close approximation and runs much faster in simulation software. When the mean is negative, the absolute value is used and then twice the absolute value of the mean is subtracted from the random sample in order to produce a negative result while keeping the shape of the distribution skewed to the right.

A WALKTHROUGH OF THE BASIC CALCULATION BASED ON INCURRED LOSS DATA

STEP 1.
BUILD A BASIC DEVELOPMENT MODEL.

Use the standard chain-ladder method and the all-period volume-weighted average (VWA) to calculate age-to-age factors.

CUMULATIVE INCURRED LOSS DATA

	12	24	36	48	60
2009	715	1,077	1,184	1,285	1,295
2010	654	794	804	835	
2011	655	886	910		
2012	837	937			
2013	747				

	12-24	24-36	36-48	48-60	60+
VWAs	1.291	1.051	1.066	1.008	1.000

STEP 2.
CREATE A "FITTED" TRIANGLE.

Start with the most recent cumulative diagonal and "un-develop" backwards using the appropriate VWAs.

TRIANGLE FITTED BACKWARDS FROM LATEST DIAGONAL

	12	24	36	48	60
2009	888	1,146	1,205	1,285	1,295
2010	577	745	783	835	
2011	671	866	910		
2012	726	937			
2013	747				

STEP 3.
CALCULATE PEARSON RESIDUALS.

Working from the incremental forms of both triangles, subtract the fitted from the actual amount for each cell. Divide each result by the square root of the fitted amount. This results in "unscaled Pearson residuals."

$$r_{UP} = \frac{C - \hat{m}}{\sqrt{abs(\hat{m})}}$$

C = actual incremental amount

$\hat{m}$ = fitted incremental for matching location in each triangle

UNSCALED PEARSON RESIDUALS

	12	24	36	48	60
2009	-5.80	6.44	6.32	2.35	0.00
2010	3.21	-2.16	-4.55	-2.91	
2011	-0.60	2.56	-3.05		
2012	4.13	-7.66			
2013	0.00				

STEP 4a.
STANDARDIZE THE RESIDUALS.

Multiply each unscaled residual by the hat matrix adjustment factor.

$$f_{ij} = \frac{1}{\sqrt{1-h_{ij}}}$$

h_{ij} = the diagonal of the hat matrix H

$H = X(X^TWX)^{-1}X^TW$

X = the design matrix from the GLM

W = the weight matrix from the GLM

HAT MATRIX ADJUSTMENT FACTORS

	12	24	36	48	60
2009	2.2611	1.3536	1.3399	1.6454	1.0000
2010	2.0614	1.2555	1.1985	1.3264	
2011	2.2445	1.2904	1.2379		
2012	2.4376	1.3153			
2013	1.0000				

STANDARDIZED PEARSON RESIDUALS

	12	24	36	48	60
2009	-13.12	8.71	8.47	3.86	0.00
2010	6.61	-2.71	-5.46	-3.86	
2011	-1.34	3.30	-3.77		
2012	10.07	-10.07			
2013	0.00				

STEP 4b.
CALCULATE THE SCALE PARAMETER.

N = number of residuals (less outliers)

p = number of parameters in the model (typically the number of columns in the residual triangle + the number of rows in the data triangle + number of additional hetero groups – 1)

ADDITIONAL STATISTICS CALCULATED

Pearson chi-squared statistic	=	sum of squares of unscaled Pearson residuals
	=	256.55
Degrees of freedom	=	# of residuals in the model *minus* # of parameters (# columns + # rows – 1)
	=	15 – 9 = 6
Scale parameter	=	Chi-squared statistic ÷ degrees of freedom
	=	42.76

SIMULATION STEPS:

STEP 5. RANDOMLY CREATE A NEW TRIANGLEOF "SAMPLE" DATA.

5a. Build a new triangle by randomly selecting (with replacement) from among the non-zero standardized Pearson residuals in Step 4a.

RANDOMLY SELECTED RESIDUALS

	12	24	36	48	60
2009	-1.34	6.61	-3.86	-2.71	8.71
2010	3.30	-2.71	8.47	-3.77	
2011	-13.12	-2.71	-13.12		
2012	3.30	6.61			
2013	10.07				

5b. Create a triangle of sample data based on the randomly selected residuals. For each cell:

$$C' = r_{SP}\sqrt{abs(\hat{m})} + \hat{m}$$

SAMPLE INCREMENTAL TRIANGLE CALCULATED BASED ON THE RANDOM RESIDUALS

	12	24	36	48	60
2009	848	365	29	56	38
2010	656	133	90	25	
2011	331	157	-43		
2012	815	307			
2013	1,022				

STEP 6. COMPLETE THE NEW RANDOMLY GENERATED TRIANGLE.

Calculate new VWAs and use them to complete the bottom right of the triangle.

NOTE: A randomly generated tail factor could also be applied here to extrapolate future development periods – i.e., beyond 60 months in this example.

SAMPLE CUMULATIVE TRIANGLE

	12	24	36	48	60
2009	848	1,213	1,242	1,297	1,335
2010	656	789	880	904	
2011	331	488	445		
2012	815	1,122			
2013	1,022				

	12-24	24-36	36-48	48-60	60+
VWAs	1.363	1.031	1.038	1.029	1.000

COMPLETED CUMULATIVE TRIANGLE WITH FUTURE EXPECTED INCURRED AMOUNTS

	12	24	36	48	60
2009	848	1,213	1,242	1,297	1,335
2010	656	789	880	904	930
2011	331	488	445	462	475
2012	815	1,122	1,157	1,201	1,235
2013	1,022	1,394	1,436	1,491	1,534

STEP 7.
INTRODUCE PROCESS VARIANCE AND CALCULATE TOTALS.

Calculate the incremental incurred amounts from the cumulative completed triangle.

To add *process variance* in the simulation, replace every future incremental incurred amount with a randomly selected point from a gamma distribution where:

Mean = the incremental incurred loss amount

Variance = Mean x Scale Parameter (see Step 4b)

COMPLETED INCREMENTAL TRIANGLE

	12	24	36	48	60
2009	848	365	29	56	38
2010	656	133	90	25	26
2011	331	157	-43	17	13
2012	815	307	34	44	35
2013	1,022	371	43	55	43

RANDOMLY GENERATED PROCESS VARIANCE

	12	24	36	48	60	TOTAL
2009	848	365	29	56	38	1,335
2010	656	133	90	25	72	976
2011	331	157	-43	44	25	514
2012	815	307	33	75	27	1,258
2013	1,022	421	35	88	86	1,653
						5,736

Note that up to this point, the calculations for the incurred model have been identical to those in the paid simulation.

STEP 8.
CONVERT TO PAID LOSS DEVELOPMENT PATTERN.

Starting with the ultimate incurred values, convert to incremental paid losses using the paid loss development pattern from Step 7 of the paid simulation. This is necessary to provide a distribution of unpaid loss as opposed to IBNR.

Note: We are using the values from Step 7 in the paid example for illustration purposes only. The simulation process does not store the results from the paid model to use with the incurred model, but will in effect run a "new" paid model (using the paid parameters) in order to generate independent results for the incurred model.

INCREMENTAL PAID WITH PROCESS VARIANCE (FROM PAID METHOD, STEP 7)

	12	24	36	48	60	TOTAL
2009	405	463	174	104	128	1,274
2010	189	246	161	30	68	694
2011	259	425	155	58	80	977
2012	342	431	194	84	127	1,179
2013	390	469	235	79	160	1,334
						5,457

INCREMENTAL INCURRED CONVERTED TO PAID PATTERN

	12	24	36	48	60	TOTAL
2009	424	485	183	109	134	1,335
2010	266	346	227	42	96	976
2011	136	223	81	31	42	514
2012	365	460	208	90	136	1,258
2013	485	581	291	98	199	1,653
						5,736

STEP 9.
CALCULATE TOTAL UNPAID AMOUNTS.

Sum the future incremental values to estimate the total unpaid loss by period and in total (in the example, 1,771).

This provides *one* estimated possible outcome.

INCREMENTAL UNPAID WITH PROCESS VARIANCE

	12	24	36	48	60	TOTAL
2009						
2010					96	96
2011				31	42	73
2012			208	90	136	433
2013		581	291	98	199	1,169
						1,771

STEP 10.
REPEAT AND SUMMARIZE.

Repeat Steps 5 through 9 the specified number of times (e.g., 10,000 or more), capturing the resulting cash flows and the unpaid amounts by period and in total for each iteration. The resulting distribution of unpaid amounts from all the iterations can be used to calculate means, percentiles, etc.

TOTAL SIMULATED RESULTS

	MEAN UNPAID	STANDARD ERROR	COEFFICIENT OF VARIATION
2009	0	0	0.0%
2010	72	33	45.7%
2011	166	56	33.9%
2012	366	101	27.5%
2013	777	233	30.0%
	1,381	272	19.7%

A WALKTHROUGH OF THE BASIC BORNHUETTER-FERGUSON CALCULATION (PAID DATA)

Note that, for simplicity, we are using the data from Steps 1 to 5 of the paid chain ladder example. In the actual simulations the BF model is simulated independently of the chain ladder model.

STEP 6. COMPLETE THE NEW RANDOMLY-GENERATED TRIANGLE.

Calculate new VWAs and BF unpaid ratios. A priori loss ratios are simulated from a selected distribution with the selected mean and CoV. Use these to complete the bottom right of the triangle.

We are only illustrating the "Deterministic" option of allocating to incremental periods in Step 6.[60] The results of the allocation are shown in Step 7.

SAMPLE CUMULATIVE TRIANGLE

	12	24	36	48	60
2009	405	868	1,042	1,146	1,274
2010	189	434	596	626	
2011	259	684	838		
2012	342	773			
2013	390				

	12-24	24-36	36-48	48-60	60+
VWAs	2.310	1.247	1.082	1.112	1.000
CDF	3.464	1.500	1.203	1.112	1.000
BFUnpd	0.711	0.333	0.169	0.101	0.000

	ULTIMATE PREMIUM	MEAN L/R	COV	SIMULATED L/R	A PRIORI ULTIMATE	TOTAL UNPAID
	(1)	(2)	(3)	(4)	(5) (4) X (1)	(6) (5) X UNPD
2009	2,000	55.0%	20.0%	60.2%	1,204	0
2010	2,000	55.0%	20.0%	50.9%	1,018	102
2011	2,000	55.0%	20.0%	49.9%	999	168
2012	2,000	55.0%	20.0%	39.7%	794	265
2013	2,000	55.0%	20.0%	87.5%	1,750	1,245

60 In addition to the "Deterministic" approach, the model also offers a "Statistical" option based on a paper by Verrall, Richard J. "A Bayesian Generalized Linear Model for the Bornhuetter-Ferguson Method of Claims Reserving," North American Actuarial Journal, Vol. 8, No. 3, July 2004, pp. 67-89. The Statistical option essentially uses a Bayesian weighting of both the columns and rows in the allocation process.

STEP 7.
INTRODUCE PROCESS VARIANCE.

To add *process variance* in the simulation, replace every future incremental paid amount with a randomly selected point from a gamma distribution where:

Mean = the incremental paid loss amount

Variance = Mean x Scale Parameter (see Step 4b)

COMPLETED INCREMENTAL TRIANGLE

	12	24	36	48	60
2009	405	463	174	104	128
2010	189	246	161	30	102
2011	259	425	155	68	101
2012	342	431	131	54	80
2013	390	662	288	119	176

RANDOMLY GENERATED FUTURE INCREMENTAL PAYMENTS BASED ON ABOVE

	12	24	36	48	60
2009					
2010					101
2011				58	79
2012			133	58	89
2013		614	303	105	204

STEP 8.
CALCULATE TOTAL UNPAID AMOUNTS.

Sum the future incremental values to estimate the total unpaid loss by period and in total (in this example, 1,743).

This provides *one* estimated possible outcome.

TOTAL ESTIMATED FUTURE PAYMENTS

	12	24	36	48	60	TOTAL
2009						
2010					101	101
2011				58	79	136
2012			133	58	89	280
2013		614	303	105	204	1,225
						1,743

STEP 9.
REPEAT AND SUMMARIZE.

Repeat Steps 5 through 8 the specified number of times (e.g., 10,000 or more), capturing the resulting cash flows and the unpaid amounts by period and in total for each iteration. The resulting distribution of unpaid amounts from all the iterations can be used to calculate means, percentiles, etc.

TOTAL SIMULATED RESULTS

	MEAN UNPAID	STANDARD ERROR	COEFFICIENT OF VARIATION
2009	0	0	0.0%
2010	96	47	48.7%
2011	189	65	34.5%
2012	381	102	26.7%
2013	782	178	22.8%
	1,448	232	16.0%

A WALKTHROUGH OF THE BASIC CAPE COD CALCULATION (PAID DATA)

Note that, for simplicity, we are using the data from Steps 1 to 5 of the paid chain ladder example. In the actual simulations the Cape Cod model is simulated independently of the chain ladder model.

STEP 6. COMPLETE THE NEW RANDOMLY-GENERATED TRIANGLE.

Calculate new VWAs. Use the Cape Cod method to calculate the a priori ultimate loss ratios and use them to complete the bottom right of the triangle.

For simplicity, we are assuming that the trend is 5.0% for each period, the weight is 100.0% for all periods and the decay rate is 75.0%.

We are only illustrating the “Deterministic” option of allocating to incremental periods in Step 6.[61] The results of the allocation are shown in Step 7.

SAMPLE CUMULATIVE TRIANGLE

	12	24	36	48	60
2009	405	868	1,042	1,146	1,274
2010	189	434	596	626	
2011	259	684	838		
2012	342	773			
2013	390				

	12-24	24-36	36-48	48-60	60+
VWAs	2.310	1.247	1.082	1.112	1.000
CDF	3.464	1.500	1.203	1.112	1.000
Paid	0.289	0.667	0.831	0.899	1.000

	ULTIMATE PREMIUM	PREMIUM INDEX	ON-LEVEL PREMIUM	PAID	TREND INDEX	TRENDED PAID
	(1)	(2)	(3) (1) X (2)	(4)	(5)	(6) (4) X (5)
2009	2,000	1.420	2,840	1,274	1.252	1,595
2010	2,000	0.970	1,940	626	1.191	745
2011	2,000	1.050	2,100	838	1.133	950
2012	2,000	1.100	2,200	773	1.078	833
2013	2,000	1.000	2,000	390	1.025	400

	PERCENT PAID	ON-LEVEL RATIO	WEIGHTED RATIO	DE-TRENDED ULTIMATE	ULTIMATE
	(7)	(8) (6) ÷ (7) ÷ (3)	(9)	(10) (9) X (3) ÷ (5)	(11) (10) X [1–(7)] + (4)
2009	1.000	56.2%	53.4%	1,211	1,274
2010	0.899	42.7%	52.6%	857	712
2011	0.831	54.4%	53.6%	993	1,006
2012	0.667	56.8%	54.5%	1,113	1,143
2013	0.289	69.3%	55.4%	1,080	1,159

61 The “Deterministic” and “Statistical” options for allocating the total unpaid by year for the Cape Cod models are the same as for the Bornhuetter-Ferguson models.

STEP 7.
INTRODUCE PROCESS VARIANCE.

To add *process variance* in the simulation, replace every future incremental paid amount with a randomly selected point from a gamma distribution where:

Mean = the incremental paid loss amount

Variance = Mean x Scale Parameter (see Step 4b)

COMPLETED INCREMENTAL TRIANGLE

	12	24	36	48	60
2009	405	463	174	104	128
2010	189	246	161	30	86
2011	259	425	155	67	100
2012	342	431	183	76	112
2013	390	408	178	73	109

RANDOMLY GENERATED FUTURE INCREMENTAL PAYMENTS BASED ON ABOVE

	12	24	36	48	60
2009					
2010					85
2011				57	78
2012			187	81	123
2013		370	189	62	130

STEP 8.
CALCULATE TOTAL UNPAID AMOUNTS.

Sum the future incremental values to estimate the total unpaid loss by period and in total (in this example, 1,362).

This provides *one* estimated possible outcome.

TOTAL ESTIMATED FUTURE PAYMENTS

	12	24	36	48	60	TOTAL
2009						
2010					85	85
2011				57	78	135
2012			187	81	123	390
2013		370	189	62	130	752
						1,362

STEP 9.
REPEAT AND SUMMARIZE.

Repeat Steps 6 through 8 the specified number of times (e.g., 10,000 or more), capturing the resulting cash flows and the unpaid amounts by period and in total for each iteration. The resulting distribution of unpaid amounts from all the iterations can be used to calculate means, percentiles, etc.

TOTAL SIMULATED RESULTS

	MEAN UNPAID	STANDARD ERROR	COEFFICIENT OF VARIATION
2009	0	0	0.0%
2010	76	35	46.4%
2011	169	51	30.1%
2012	375	71	18.9%
2013	741	92	12.5%
	1,362	164	12.1%

USING THE ODP PROCESS ALGORITHM

A Walkthrough of the Chain Ladder Calculation (Paid Data)

Note that, for simplicity, we are using the data from Steps 1 to 7 of the paid chain ladder example, since the ODP Process algorithm builds on the basic model for the time horizon simulations.

STEP 8.
USE ORIGINAL DATA TRIANGLE AND N FUTURE DIAGONALS TO CALCULATE REMAINING UNPAID.

Calculate new VWAs and use them to complete the bottom right of the trapezoid. In this example N = 1.

NOTE: If a tail factor was used in the basic model, then the projected value would be used in the future diagonal and to calculate the VWAs. If the tail factor was extrapolated, then the remaining random factors would be applied – i.e., beyond 72 months in this example.

CUMULATIVE ONE-YEAR SAMPLE TRAPEZOID

	12	24	36	48	60
2009	352	783	1,045	1,183	1,295
2010	255	572	710	750	**818**
2011	279	638	767	**825**	
2012	311	717	**911**		
2013	308	**777**			

	12-24	24-36	36-48	48-60	60+
VWAs	2.317	1.267	1.094	1.093	1.000

COMPLETED CUMULATIVE TRAPEZOID WITH FUTURE EXPECTED PAID AMOUNTS

	12	24	36	48	60
2009	352	783	1,045	1,183	1,295
2010	255	572	710	750	**818**
2011	279	638	767	**825**	902
2012	311	717	**911**	997	1,090
2013	308	**777**	984	1,077	1,177

STEP 9.
CALCULATE TOTAL OF POSSIBLE OUTCOME AND EXPECTED UNPAID AMOUNTS.

Sum the future incremental values to estimate the total unpaid loss by period and in total (in this example, 1,445).

This provides *one* estimated possible outcome, with its conditional expected value.

TOTAL ESTIMATED ONE-YEAR FUTURE PAYMENTS AND REMAINING POINT ESTIMATE

	12	24	36	48	60	TOTAL
2009						
2010					**68**	68
2011				**58**	77	135
2012			**194**	85	93	373
2013		**469**	207	92	100	869
						1,445

STEP 10.
REPEAT AND SUMMARIZE.

Repeat Steps 5 through 9 the specified number of times (e.g., 10,000 or more), capturing the resulting cash flows and the unpaid amounts by period and in total for each iteration. The resulting distribution of unpaid amounts from all the iterations can be used to calculate means, percentiles, etc.

TOTAL SIMULATED RESULTS

	MEAN UNPAID	STANDARD ERROR	COEFFICIENT OF VARIATION
2009	0	0	0.0%
2010	71	30	42.3%
2011	158	35	22.4%
2012	378	61	16.3%
2013	761	140	18.5%
	1,369	184	13.4%

A Walkthrough of the Chain Ladder Calculation (Incurred Data)

Note that, for simplicity, we are using the data from Steps 1 to 8 of the incurred chain ladder example, since the ODP Process algorithm builds on the basic model for the time horizon simulations.

STEP 9.
USE ORIGINAL INCURRED DATA TRIANGLE AND N FUTURE DIAGONALS TO CALCULATE REMAINING IBNR.

Calculate new VWAs and use them to complete the bottom right of the trapezoid. In this example N=1.

NOTE: If a tail factor was used in the basic model, then the projected value would be used in the future diagonal and to calculate the VWAs. If the tail factor was extrapolated, then the remaining random factors would be applied – i.e., beyond 72 months in this example.

CUMULATIVE ONE-YEAR SAMPLE INCURRED TRAPEZOID

	12	24	36	48	60
2009	715	1,077	1,184	1,285	1,295
2010	654	794	804	835	**907**
2011	655	886	910	**954**	
2012	837	937	**970**		
2013	747	**1,168**			

	12-24	24-36	36-48	48-60	60+
VWAs	1.348	1.047	1.061	1.039	1.000

COMPLETED CUMULATIVE TRAPEZOID WITH FUTURE EXPECTED INCURRED AMOUNTS

	12	24	36	48	60
2009	715	1,077	1,184	1,285	1,295
2010	654	794	804	835	**907**
2011	655	886	910	**954**	991
2012	837	937	**970**	1,029	1,069
2013	747	**1,168**	1,223	1,297	1,348

STEP 10.
USE ORIGINAL PAID DATA TRIANGLE AND N FUTURE DIAGONALS TO CALCULATE REMAINING UNPAID.

Calculate new VWAs and use them to complete the bottom right of the trapezoid. In this example N=1.

NOTE: If a tail factor was used in the basic model, then the projected value would be used in the future diagonal and to calculate the VWAs. If the tail factor was extrapolated, then the remaining random factors would be applied – i.e., beyond 72 months in this example.

CUMULATIVE ONE-YEAR SAMPLE PAID TRAPEZOID

	12	24	36	48	60
2009	352	783	1,045	1,183	1,295
2010	255	572	710	750	**818**
2011	279	638	767	**825**	
2012	311	717	**911**		
2013	308	**777**			

	12-24	24-36	36-48	48-60	60+
VWAs	2.317	1.267	1.094	1.093	1.000

COMPLETED CUMULATIVE TRAPEZOID WITH FUTURE EXPECTED PAID AMOUNTS

	12	24	36	48	60
2009	352	783	1,045	1,183	1,295
2010	255	572	710	750	**818**
2011	279	638	767	**825**	902
2012	311	717	**911**	997	1,090
2013	308	**777**	984	1,077	1,177

STEP 11.
USE ORIGINAL PAID DATA TRIANGLE AND N FUTURE INCURRED PAID DIAGONALS. SUBTRACT THE CUMULATIVE PAID FROM INCURRED ULTIMATE VALUES. ALLOCATE REMAINING UNPAID USING PAID PATTERN.

Use the incurred converted to a random paid pattern from step 8 of the basic model for the trapezoid.

Calculate the remaining unpaid by subtracting the paid to date from the ultimate incurred by year from Step 9. Calculate the allocation percentages from the future paid in step 10.

Allocate the remaining unpaid to incremental period using the allocation percentages.

Sum the future incremental values to estimate the total unpaid loss by period and in total (in this example, 1,711).

This provides *one* estimated possible outcome, with its conditional expected value.

INCREMENTAL ONE-YEAR SAMPLE PAID TRAPEZOID

	12	24	36	48	60
2009	352	431	262	183	112
2010	255	317	138	40	**96**
2011	279	359	129	**31**	
2012	311	406	**208**		
2013	308	**581**			

	REMAINING UNPAID	36	48	60
2009	0			
2010	0			
2011	193			1.00
2012	144		0.48	0.52
2013	459	0.52	0.23	0.25

COMPLETED INCREMENTAL TRAPEZOID WITH FUTURE INCURRED CONVERTED TO PAID

	12	24	36	48	60	TOTAL
2009						
2010					**96**	96
2011				**31**	193	224
2012			**208**	69	75	352
2013		**581**	238	106	115	1,040
						1,711

STEP 12.
REPEAT AND SUMMARIZE.

Repeat Steps 5 through 11 the specified number of times (e.g., 10,000 or more), capturing the resulting cash flows and the unpaid amounts by period and in total for each iteration. The resulting distribution of unpaid amounts from all the iterations can be used to calculate means, percentiles, etc.

TOTAL SIMULATED RESULTS

	MEAN UNPAID	STANDARD ERROR	COEFFICIENT OF VARIATION
2009	0	0	0.0%
2010	72	33	45.7%
2011	213	70	33.0%
2012	344	67	19.5%
2013	786	151	19.2%
	1,415	210	14.8%

A Walkthrough of the Bornhuetter-Ferguson Calculation (Paid Data)

Note that, for simplicity, we are using the data from Steps 1 to 7 of the paid Bornhuetter-Ferguson example, since the ODP Process algorithm builds on the basic model for the time horizon simulations.

STEP 8.
USE ORIGINAL DATA TRIANGLE AND N FUTURE DIAGONALS TO CALCULATE REMAINING UNPAID.

Calculate new VWAs and BF unpaid ratios and use them to complete the bottom right of the trapezoid. In this example N=1.

Simulated a priori loss ratios from the basic model are used again without resampling. Use these to complete the bottom right of the triangle.

NOTE: If a tail factor was used in the basic model, then the projected value would be used in the future diagonal and to calculate the VWAs. If the tail factor was extrapolated, then the remaining random factors would be applied – i.e., beyond 72 months in this example.

CUMULATIVE ONE-YEAR SAMPLE TRAPEZOID

	12	24	36	48	60
2009	352	783	1,045	1,183	1,295
2010	255	572	710	750	**851**
2011	279	638	767	**825**	
2012	311	717	**850**		
2013	308	**922**			

	12-24	24-36	36-48	48-60	60+
VWAs	2.413	1.244	1.093	1.110	1.000
CDF	3.645	1.511	1.214	1.110	1.000
BFUnpd	0.726	0.338	0.176	0.099	0.000

	ULTIMATE PREMIUM	MEAN L/R	COV	SIMULATED L/R	A PRIORI ULTIMATE	TOTAL UNPAID
	(1)	(2)	(3)	(4)	(5) (4) X (1)	(6) (5) X UNPD
2009	2,000	55.0%	20.0%	60.2%	1,204	0
2010	2,000	55.0%	20.0%	50.9%	1,018	0
2011	2,000	55.0%	20.0%	49.9%	999	99
2012	2,000	55.0%	20.0%	39.7%	794	140
2013	2,000	55.0%	20.0%	87.5%	1,750	592

STEP 9.
CALCULATE TOTAL OF POSSIBLE OUTCOME AND EXPECTED UNPAID AMOUNTS.

Only the "Deterministic" option of allocating to incremental periods is available for the ODP Process algorithm.

Sum the future incremental values to estimate the total unpaid loss by period and in total (in this example, 1,737).

This provides one estimated possible outcome, with its conditional expected value.

TOTAL ESTIMATED ONE-YEAR FUTURE PAYMENTS AND REMAINING POINT ESTIMATE

	12	24	36	48	60	TOTAL
2009						
2010					**101**	101
2011				**58**	99	157
2012			**133**	61	79	273
2013		**614**	283	135	174	1,205
						1,737

STEP 10.
REPEAT AND SUMMARIZE.

Repeat Steps 5 through 9 the specified number of times (e.g., 10,000 or more), capturing the resulting cash flows and the unpaid amounts by period and in total for each iteration. The resulting distribution of unpaid amounts from all the iterations can be used to calculate means, percentiles, etc.

TOTAL SIMULATED RESULTS

	MEAN UNPAID	STANDARD ERROR	COEFFICIENT OF VARIATION
2009	0	0	0.0%
2010	96	47	48.7%
2011	200	58	28.9%
2012	394	94	23.9%
2013	794	172	21.6%
	1,484	232	15.6%

A Walkthrough of the Cape Cod Calculation (Paid Data)

Note that, for simplicity, we are using the data from Steps 1 to 7 of the paid Cape Cod example, since the ODP Process algorithm builds on the basic model for the time horizon simulations.

STEP 8.
USE ORIGINAL DATA TRIANGLE AND N FUTURE DIAGONALS TO CALCULATE REMAINING UNPAID.

Calculate new VWAs and CC paid ratios and use them to complete the bottom right of the trapezoid. In this example N=1.

Simulated a priori loss ratios from the basic model are used again without resampling. Use these to complete the bottom right of the triangle.

For simplicity, we are assuming that the trend is 5.0% for each period, the weight is 100.0% for all periods and the decay rate is 75.0%.

NOTE: If a tail factor was used in the basic model, then the projected value would be used in the future diagonal and to calculate the VWAs. If the tail factor was extrapolated, then the remaining random factors would be applied – i.e., beyond 72 months in this example.

CUMULATIVE ONE-YEAR SAMPLE TRAPEZOID

	12	24	36	48	60
2009	352	783	1,045	1,183	1,295
2010	255	572	710	750	**835**
2011	279	638	767	**824**	
2012	311	717	**904**		
2013	308	**678**			

	12-24	24-36	36-48	48-60	60+
VWAs	2.251	1.264	1.093	1.102	1.000
CDF	3.428	1.523	1.205	1.102	1.000
Paid	0.292	0.657	0.830	0.908	1.000

	ULTIMATE PREMIUM (1)	PREMIUM INDEX (2)	ON-LEVEL PREMIUM (3) (1) X (2)	PAID (4)	TREND INDEX (5)	TRENDED PAID (6) (4) X (5)
2009	2,000	1.420	2,840	1,295	1.252	1,621
2010	2,000	0.970	1,940	835	1.191	994
2011	2,000	1.050	2,100	824	1.133	934
2012	2,000	1.100	2,200	904	1.078	974
2013	2,000	1.000	2,000	678	1.025	696

	PERCENT PAID (7)	ON-LEVEL RATIO (8) (6) ÷ (7) ÷ (3)	WEIGHTED RATIO (9)	DE-TRENDED ULTIMATE (10) (9) X (3) ÷ (5)	ULTIMATE (11) (10) X [1–(7)] + (4)
2009	1.000	57.1%	53.8%	1,220	1,295
2010	1.000	51.2%	53.0%	863	835
2011	0.908	49.0%	52.5%	974	914
2012	0.830	53.3%	52.7%	1,075	1,086
2013	0.657	53.0%	52.7%	1,028	1,031

STEP 9.
CALCULATE TOTAL OF POSSIBLE OUTCOME AND EXPECTED UNPAID AMOUNTS.

Only the “Deterministic” option of allocating to incremental periods is available for the ODP Process algorithm.

Sum the future incremental values to estimate the total unpaid loss by period and in total (in this example, 1,325).

This provides one estimated possible outcome, with its conditional expected value.

TOTAL ESTIMATED ONE-YEAR FUTURE PAYMENTS AND REMAINING POINT ESTIMATE

	12	24	36	48	60	TOTAL
2009						
2010					**85**	85
2011				**57**	90	147
2012			**187**	83	99	369
2013		**370**	178	80	95	723
						1,325

STEP 10.
REPEAT AND SUMMARIZE.

Repeat Steps 5 through 9 the specified number of times (e.g., 10,000 or more), capturing the resulting cash flows and the unpaid amounts by period and in total for each iteration. The resulting distribution of unpaid amounts from all the iterations can be used to calculate means, percentiles, etc.

TOTAL SIMULATED RESULTS

	MEAN UNPAID	STANDARD ERROR	COEFFICIENT OF VARIATION
2009	0	0	0.0%
2010	76	35	46.4%
2011	171	39	22.8%
2012	379	55	14.5%
2013	743	73	9.8%
	1,369	145	10.6%

USING THE ODP RESIDUAL ALGORITHM

A Walkthrough of the Chain Ladder Calculation (Paid Data)

Note that, for simplicity, we are using the data from Steps 1 to 4 of the paid chain ladder example, since the ODP Residual algorithm builds on the basic model for the time horizon simulations.

STEP 5.
RANDOMLY CREATE A NEW TRAPEZOID OF "SAMPLE" DATA.

5a. Build a new trapezoid by randomly selecting (with replacement) from among the non-zero standardized Pearson residuals in Step 4a.

5b. Add the future expected values (fitted) to the fitted triangle from Step 2 using the appropriate VWAs from Step 1.

5c. Create a trapezoid of sample data based on the randomly selected residuals. For each cell:

$$C' = r_{SP}\sqrt{abs(\hat{m})} + \hat{m}$$

RANDOMLY SELECTED RESIDUALS

	12	24	36	48	60
2009	1.54	-0.50	-3.39	-0.50	1.54
2010	-3.17	-3.17	1.54	-4.69	**0.98**
2011	-0.53	4.69	-0.50	**-1.76**	
2012	1.40	1.54	**1.40**		
2013	4.69	**4.69**			

TRAPEZOID FITTED BACKWARDS FROM LATEST DIAGONAL & FUTURE EXPECTATION

	12	24	36	48	60
2009	375	849	1,074	1,183	1,295
2010	238	538	681	750	**821**
2011	268	606	767	**845**	
2012	317	717	**907**		
2013	308	**697**			

SAMPLE INCREMENTAL TRAPEZOID CALCULATED BASED ON THE RANDOM RESIDUALS

	12	24	36	48	60
2009	405	463	174	104	128
2010	189	246	161	30	**79**
2011	259	425	155	**62**	
2012	342	431	**210**		
2013	390	**482**			

STEP 6.
USE SAMPLE DATA TRAPEZOID (INCLUDING N FUTURE DIAGONALS) TO CALCULATE REMAINING UNPAID.

Calculate new VWAs and use them to complete the bottom right of the trapezoid. In this example N=1.

NOTE: If a tail factor was used in the basic model, then the projected value would be used in the future diagonal and to calculate the VWAs. If the tail factor was extrapolated, then the remaining random factors would be applied – i.e., beyond 72 months in this example.

SAMPLE CUMULATIVE TRAPEZOID

	12	24	36	48	60
2009	405	868	1,042	1,146	1,274
2010	189	434	596	626	705
2011	259	684	838	901	
2012	342	773	982		
2013	390	872			

	12-24	24-36	36-48	48-60	60+
VWAs	2.291	1.254	1.079	1.117	1.000

COMPLETED CUMULATIVE TRAPEZOID WITH FUTURE EXPECTED PAYMENTS

	12	24	36	48	60
2009	405	868	1,042	1,146	1,274
2010	189	434	596	626	**705**
2011	259	684	838	**901**	1,006
2012	342	773	**982**	1,060	1,184
2013	390	**872**	1,093	1,180	1,318

STEP 7.
CALCULATE TOTAL OF POSSIBLE OUTCOME AND EXPECTED UNPAID AMOUNTS.

Sum the future incremental values to estimate the total unpaid loss by period and in total (in this example, 1,587).

This provides *one* estimated possible outcome, with its conditional expected value.

TOTAL ESTIMATED ONE-YEAR FUTURE PAYMENTS AND REMAINING POINT ESTIMATE

	12	24	36	48	60	TOTAL
2009						
2010					**79**	79
2011				**62**	105	168
2012			**210**	78	124	412
2013		**482**	221	87	138	928
						1,587

STEP 8.
REPEAT AND SUMMARIZE.

Repeat Steps 5 through 7 the specified number of times (e.g., 10,000 or more), capturing the resulting cash flows and the unpaid amounts by period and in total for each iteration. The resulting distribution of unpaid amounts from all the iterations can be used to calculate means, percentiles, etc.

TOTAL SIMULATED RESULTS

	MEAN UNPAID	STANDARD ERROR	COEFFICIENT OF VARIATION
2009	0	0	0.0%
2010	72	22	31.2%
2011	159	31	19.7%
2012	379	54	14.2%
2013	759	91	12.0%
	1,368	136	9.9%

A Walkthrough of the Chain Ladder Calculation (Incurred Data)

Note that, for simplicity, we are using the data from Steps 1 to 4 of the incurred chain ladder example, since the ODP Residual algorithm builds on the basic model for the time horizon simulations.

STEP 5.
RANDOMLY CREATE A NEW TRAPEZOID OF "SAMPLE" DATA.

5a. Build a new trapezoid by randomly selecting (with replacement) from among the non-zero standardized Pearson residuals in Step 4a.

5b. Add the future expected values (fitted) to the fitted triangle from Step 2 using the appropriate VWAs from Step 1.

5c. Create a trapezoid of sample data based on the randomly selected residuals. For each cell:

$$C' = r_{SP}\sqrt{abs(\hat{m})} + \hat{m}$$

RANDOMLY SELECTED RESIDUALS

	12	24	36	48	60
2009	-1.34	6.61	-3.86	-2.71	8.71
2010	3.30	-2.71	8.47	-3.77	**-3.77**
2011	-13.12	-2.71	-13.12	**8.71**	
2012	3.30	6.61	**-2.71**		
2013	10.07	**3.30**			

TRAPEZOID FITTED BACKWARDS FROM LATEST DIAGONAL & FUTURE EXPECTATION

	12	24	36	48	60
2009	888	1,146	1,205	1,285	1,295
2010	577	745	783	835	**841**
2011	671	866	910	**970**	
2012	726	937	**985**		
2013	747	**964**			

SAMPLE INCREMENTAL TRAPEZOID CALCULATED BASED ON THE RANDOM RESIDUALS

	12	24	36	48	60
2009	848	365	29	56	38
2010	656	133	90	25	**-3**
2011	331	157	-43	**128**	
2012	815	307	**29**		
2013	1,022	**266**			

STEP 6.
USE SAMPLE DATA TRAPEZOID (INCLUDING N FUTURE DIAGONALS) TO CALCULATE REMAINING UNPAID.

Calculate new VWAs and use them to complete the bottom right of the trapezoid. In this example N=1.

NOTE: If a tail factor was used in the basic model, then the projected value would be used in the future diagonal and to calculate the VWAs. If the tail factor was extrapolated, then the remaining random factors would be applied – i.e., beyond 72 months in this example.

SAMPLE CUMULATIVE TRAPEZOID

	12	24	36	48	60
2009	848	1,213	1,242	1,297	1,335
2010	656	789	880	904	**901**
2011	331	488	445	**573**	
2012	815	1,122	**1,151**		
2013	1,022	**1,288**			

	12-24	24-36	36-48	48-60	60+
VWAs	1.335	1.029	1.081	1.016	1.000

COMPLETED CUMULATIVE TRAPEZOID WITH FUTURE EXPECTED INCURRED AMOUNTS

	12	24	36	48	60
2009	848	1,213	1,242	1,297	1,335
2010	656	789	880	904	**901**
2011	331	488	445	**573**	582
2012	815	1,122	**1,151**	1,245	1,264
2013	1,022	**1,288**	1,326	1,434	1,456

Note that up to this point, the calculations for the incurred model have been identical to those in the paid simulation.

STEP 7.
USE BASIC MODEL TO CONVERT TO PAID LOSS DEVELOPMENT PATTERN.

Use the results from Step 8 of the basic model. This is necessary to insure that the triangle results are identical to the basic model.

INCREMENTAL PAID WITH PROCESS VARIANCE (FROM BASIC MODEL)

	12	24	36	48	60	TOTAL
2009	405	463	174	104	128	1,274
2010	189	246	161	30	68	694
2011	259	425	155	58	80	977
2012	342	431	194	84	127	1,179
2013	390	469	235	79	160	1,334
						5,457

INCREMENTAL INCURRED CONVERTED TO PAID PATTERN (FROM BASIC MODEL)

	12	24	36	48	60	TOTAL
2009	424	485	183	109	134	1,335
2010	266	346	227	42	96	976
2011	136	223	81	31	42	514
2012	365	460	208	90	136	1,258
2013	485	581	291	98	199	1,653
						5,736

STEP 8.
RUN THE MODEL IN PARALLEL FOR PAID DATA.

Note: We are using the values from Step 6 in the paid example for illustration purposes only. The simulation process does not store the results from the paid model to use with the incurred model, but will in effect run a "new" paid model (using the paid parameters) in order to generate independent results for the incurred model.

COMPLETED CUMULATIVE PAID TRAPEZOID WITH FUTURE EXPECTED PAYMENTS

	12	24	36	48	60
2009	405	868	1,042	1,146	1,274
2010	189	434	596	626	**705**
2011	259	684	838	**901**	1,006
2012	342	773	**982**	1,060	1,184
2013	390	**872**	1,093	1,180	1,318

STEP 9.
USE BASIC MODEL SAMPLE TRIANGLE CONVERTED TO A PAID PATTERN. SUBTRACT THE CUMULATIVE PAID FROM INCURRED ULTIMATE VALUES. ALLOCATE REMAINING UNPAID USING PAID PATTERN.

Use the incurred converted to a random paid pattern from the basic model for the triangle.

Calculate the remaining unpaid by subtracting the paid to date from the ultimate incurred by year from Step 6. Calculate the allocation percentages from the future paid in step 8.

Allocate the remaining unpaid to incremental period using the allocation percentages.

Sum the future incremental values to estimate the total unpaid loss by period and in total (in this example, 1,575).

This provides one estimated possible outcome, with its conditional expected value.

INCREMENTAL SAMPLE PAID TRIANGLE

	12	24	36	48	60
2009	424	485	183	109	134
2010	266	346	227	42	
2011	136	223	81		
2012	365	460			
2013	485				

	REMAINING UNPAID	24	36	48	60
2009	0				
2010	21				1.00
2011	141			0.37	0.63
2012	440		0.51	0.19	0.30
2013	973	0.52	0.24	0.09	0.15

COMPLETED INCREMENTAL TRIANGLE WITH FUTURE INCURRED CONVERTED TO PAID

	12	24	36	48	60	TOTAL
2009						
2010					21	21
2011				52	89	141
2012			**224**	83	133	440
2013		**505**	232	91	145	973
						1,575

STEP 10.
REPEAT AND SUMMARIZE.

Repeat Steps 5 through 9 the specified number of times (e.g., 10,000 or more), capturing the resulting cash flows and the unpaid amounts by period and in total for each iteration. The resulting distribution of unpaid amounts from all the iterations can be used to calculate means, percentiles, etc.

TOTAL SIMULATED RESULTS

	MEAN UNPAID	STANDARD ERROR	COEFFICIENT OF VARIATION
2009	0	0	0.0%
2010	71	46	64.2%
2011	164	93	56.8%
2012	365	117	32.0%
2013	775	194	25.0%
	1,376	270	19.6%

A Walkthrough of the Bornhuetter-Ferguson Calculation (Paid Data)

Note that, for simplicity, we are using the data from Steps 1 to 4 of the paid Bornhuetter-Ferguson example, since the ODP Residual algorithm builds on the basic model for the time horizon simulations.

STEP 5. RANDOMLY CREATE A NEW TRAPEZOID OF "SAMPLE" DATA.

5a. Build a new trapezoid by randomly selecting (with replacement) from among the non-zero standardized Pearson residuals in Step 4a.

5b. Add the future expected values (fitted) to the fitted triangle from Step 2 using the appropriate VWAs from Step 1 and the mean BF assumptions.

5c. Create a trapezoid of sample data based on the randomly selected residuals. For each cell:

$$C' = r_{SP}\sqrt{abs(\hat{m})} + \hat{m}$$

RANDOMLY SELECTED RESIDUALS

	12	24	36	48	60
2009	1.54	-0.50	-3.39	-0.50	1.54
2010	-3.17	-3.17	1.54	-4.69	**0.98**
2011	-0.53	4.69	-0.50	**-1.76**	
2012	1.40	1.54	**1.40**		
2013	4.69	**4.69**			

TRAPEZOID FITTED BACKWARDS FROM LATEST DIAGONAL & FUTURE EXPECTATION

	12	24	36	48	60
2009	375	849	1,074	1,183	1,295
2010	238	538	681	750	**845**
2011	268	606	767	**860**	
2012	317	717	**908**		
2013	308	**711**			

SAMPLE INCREMENTAL TRAPEZOID CALCULATED BASED ON THE RANDOM RESIDUALS

	12	24	36	48	60
2009	405	463	174	104	128
2010	189	246	161	30	**105**
2011	259	425	155	**76**	
2012	342	431	**211**		
2013	390	**497**			

STEP 6. USE ORIGINAL DATA TRIANGLE AND N FUTURE DIAGONALS TO CALCULATE REMAINING UNPAID.

Calculate new VWAs and BF unpaid ratios and use them to complete the bottom right of the trapezoid. In this example N=1.

Simulated a priori loss ratios from the basic model are used again without resampling. Use these to complete the bottom right of the triangle.

CUMULATIVE ONE-YEAR SAMPLE TRAPEZOID

	12	24	36	48	60
2009	405	868	1,042	1,146	1,274
2010	189	434	596	626	**730**
2011	259	684	838	**914**	
2012	342	773	**983**		
2013	390	**877**			

	12-24	24-36	36-48	48-60	60+
VWAs	2.301	1.254	1.085	1.131	1.000
CDF	3.541	1.539	1.227	1.131	1.000
BFUnpd	0.718	0.350	0.185	0.116	0.000

NOTE: If a tail factor was used in the basic model, then the projected value would be used in the future diagonal and to calculate the VWAs. If the tail factor was extrapolated, then the remaining random factors would be applied – i.e., beyond 72 months in this example.

	ULTIMATE PREMIUM	MEAN L/R	COV	SIMULATED L/R	A PRIORI ULTIMATE	TOTAL UNPAID
	(1)	(2)	(3)	(4)	(5)	(6)
					(4) X (1)	(5) X UNPD
2009	2,000	55.0%	20.0%	60.2%	1,204	0
2010	2,000	55.0%	20.0%	50.9%	1,018	0
2011	2,000	55.0%	20.0%	49.9%	999	116
2012	2,000	55.0%	20.0%	39.7%	794	147
2013	2,000	55.0%	20.0%	87.5%	1,750	613

STEP 7.
CALCULATE TOTAL OF POSSIBLE OUTCOME AND EXPECTED UNPAID AMOUNTS.

Only the "Deterministic" option of allocating to incremental periods is available for the ODP Process algorithm.

Sum the future incremental values to estimate the total unpaid loss by period and in total (in this example, 1,764).

This provides *one* estimated possible outcome, with its conditional expected value.

TOTAL ESTIMATED ONE-YEAR FUTURE PAYMENTS AND REMAINING POINT ESTIMATE

	12	24	36	48	60	TOTAL
2009						
2010					**105**	105
2011				**76**	116	192
2012			**211**	55	92	358
2013		**497**	289	121	203	1,110
						1,764

STEP 8.
REPEAT AND SUMMARIZE.

Repeat Steps 5 through 7 the specified number of times (e.g., 10,000 or more), capturing the resulting cash flows and the unpaid amounts by period and in total for each iteration. The resulting distribution of unpaid amounts from all the iterations can be used to calculate means, percentiles, etc.

TOTAL SIMULATED RESULTS

	MEAN UNPAID	STANDARD ERROR	COEFFICIENT OF VARIATION
2009	0	0	0.0%
2010	95	25	26.7%
2011	198	38	19.4%
2012	394	59	15.0%
2013	795	99	12.4%
	1,483	143	9.6%

A Walkthrough of the Cape Cod Calculation (Paid Data)

Note that, for simplicity, we are using the data from Steps 1 to 4 of the paid Cape Cod example, since the ODP Residual algorithm builds on the basic model for the time horizon simulations.

STEP 5.
RANDOMLY CREATE A NEW TRAPEZOID OF "SAMPLE" DATA.

5a. Build a new trapezoid by randomly selecting (with replacement) from among the non-zero standardized Pearson residuals in Step 4a.

5b. Add the future expected values (fitted) to the fitted triangle from Step 2 using the appropriate VWAs from Step 1 and the mean CC assumptions.

5c. Create a trapezoid of sample data based on the randomly selected residuals. For each cell:

$$C' = r_{SP}\sqrt{abs(\hat{m})} + \hat{m}$$

RANDOMLY SELECTED RESIDUALS

	12	24	36	48	60
2009	1.54	-0.50	-3.39	-0.50	1.54
2010	-3.17	-3.17	1.54	-4.69	0.98
2011	-0.53	4.69	-0.50	-1.76	
2012	1.40	1.54	1.40		
2013	4.69	4.69			

TRAPEZOID FITTED BACKWARDS FROM LATEST DIAGONAL & FUTURE EXPECTATION

	12	24	36	48	60
2009	375	849	1,074	1,183	1,295
2010	238	538	681	750	826
2011	268	606	767	848	
2012	317	717	894		
2013	308	650			

SAMPLE INCREMENTAL TRAPEZOID CALCULATED BASED ON THE RANDOM RESIDUALS

	12	24	36	48	60
2009	405	463	174	104	128
2010	189	246	161	30	84
2011	259	425	155	65	
2012	342	431	196		
2013	390	429			

STEP 6.
USE ORIGINAL DATA TRIANGLE AND N FUTURE DIAGONALS TO CALCULATE REMAINING UNPAID.

Calculate new VWAs and CC paid ratios and use them to complete the bottom right of the trapezoid. In this example N=1.

Simulated a priori loss ratios from the basic model are used again without resampling. Use these to complete the bottom right of the triangle.

CUMULATIVE ONE-YEAR SAMPLE TRAPEZOID

	12	24	36	48	60
2009	405	868	1,042	1,146	1,274
2010	189	434	596	626	710
2011	259	684	838	903	
2012	342	773	969		
2013	390	819			

	12-24	24-36	36-48	48-60	60+
VWAs	2.258	1.249	1.080	1.120	1.000
CDF	3.411	1.511	1.210	1.120	1.000
Paid	0.293	0.662	0.827	0.893	1.000

For simplicity, we are assuming that the trend is 5.0% for each period, the weight is 100.0% for all periods and the decay rate is 75.0%.

NOTE: If a tail factor was used in the basic model, then the projected value would be used in the future diagonal and to calculate the VWAs. If the tail factor was extrapolated, then the remaining random factors would be applied – i.e., beyond 72 months in this example.

	ULTIMATE PREMIUM	PREMIUM INDEX	ON-LEVEL PREMIUM	PAID	TREND INDEX	TRENDED PAID
	(1)	(2)	(3) (1) X (2)	(4)	(5)	(6) (4) X (5)
2009	2,000	1.420	2,840	1,274	1.252	1,595
2010	2,000	0.970	1,940	710	1.191	845
2011	2,000	1.050	2,100	903	1.133	1,023
2012	2,000	1.100	2,200	969	1.078	1,044
2013	2,000	1.000	2,000	819	1.025	840

	PERCENT PAID	ON-LEVEL RATIO	WEIGHTED RATIO	DE-TRENDED ULTIMATE	ULTIMATE
	(7)	(8) (6) ÷ (7) ÷ (3)	(9)	(10) (9) X (3) ÷ (5)	(11) (10) X [1–(7)] + (4)
2009	1.000	56.2%	53.7%	1,219	1,274
2010	1.000	43.6%	53.1%	866	710
2011	0.893	54.6%	54.1%	1,004	1,011
2012	0.827	57.4%	55.2%	1,126	1,164
2013	0.662	63.4%	56.0%	1,093	1,189

STEP 7.
CALCULATE TOTAL OF POSSIBLE OUTCOME AND EXPECTED UNPAID AMOUNTS.

Only the “Deterministic” option of allocating to incremental periods is available for the ODP Process algorithm.

Sum the future incremental values to estimate the total unpaid loss by period and in total (in this example, 1,446).

This provides *one* estimated possible outcome, with its conditional expected value.

TOTAL ESTIMATED ONE-YEAR FUTURE PAYMENTS AND REMAINING POINT ESTIMATE

	12	24	36	48	60	TOTAL
2009						
2010					**84**	84
2011				**65**	108	172
2012			**196**	75	121	391
2013		**429**	180	72	117	798
						1,446

STEP 8.
REPEAT AND SUMMARIZE.

Repeat Steps 5 through 7 the specified number of times (e.g., 10,000 or more), capturing the resulting cash flows and the unpaid amounts by period and in total for each iteration. The resulting distribution of unpaid amounts from all the iterations can be used to calculate means, percentiles, etc.

TOTAL SIMULATED RESULTS

	MEAN UNPAID	STANDARD ERROR	COEFFICIENT OF VARIATION
2009	0	0	0.0%
2010	75	23	30.4%
2011	170	30	17.7%
2012	379	46	12.2%
2013	742	63	8.5%
	1,367	118	8.6%

USING THE MACK PROCESS ALGORITHM

A Walkthrough of the Chain Ladder Calculation (Paid Data)

Note that since the Mack Process algorithm is different than the basic model so all steps are different.

STEP 1.
BUILD A BASIC DEVELOPMENT MODEL.

Use the standard chain-ladder method and the all-period volume-weighted average (VWA) to calculate age-to-age factors.

CUMULATIVE PAID LOSS DATA

	12	24	36	48	60
2009	352	783	1,045	1,183	1,295
2010	255	572	710	750	
2011	279	638	767		
2012	311	717			
2013	308				

AGE-TO-AGE FACTORS

	12-24	24-36	36-48	48-60
2009	2.224	1.335	1.132	1.095
2010	2.243	1.241	1.056	
2011	2.287	1.202		
2012	2.306			

	12-24	24-36	36-48	48-60	60+
VWAs	2.264	1.265	1.101	1.095	1.000

STEP 2.
CALCULATE UNSCALED RESIDUALS.

Using the age-to-age factors, subtract the average from the actual amount for each cell. Divide each result by the square root of the cumulative paid amount. This results in “unscaled residuals.”

$r_u = (C - A)\ x\ \sqrt{m}$

C = actual age-to-age factor

A = average age-to-age factor

m = cumulative amount from triangle of data

UNSCALED RESIDUALS

	12-24	24-36	36-48	48-60
2009	-0.742	1.936	0.990	0.000
2010	-0.333	-0.578	-1.201	
2011	0.380	-1.597		
2012	0.731			

STEP 3.
STANDARDIZE THE RESIDUALS.

Divide each unscaled residual by the standard deviation of the unscaled residuals in that column.[62]

The last standard deviation is the minimum of the two prior or the prior squared divided by the second prior, whichever is less.

$$r_S = \frac{r_U}{sd}$$

sd = standard deviation of unscaled residuals

STANDARD DEVIATION OF RESIDUALS

	12-24	24-36	36-48	48-60
2009	0.669	1.821	1.557	1.331

STANDARDIZED RESIDUALS

	12-24	24-36	36-48	48-60
2009	-1.110	1.063	0.636	0.00
2010	-0.498	-0.317	-0.772	
2011	0.568	-0.877		
2012	1.094			

SIMULATION STEPS

STEP 4.
RANDOMLY CREATE A NEW TRIANGLE OF "SAMPLE" AGE-TO-AGE FACTORS.

4a. Build a new triangle by randomly selecting (with replacement) from among the non-zero standardized residuals in Step 3.[63]

4b. Create a triangle of sample data based on the randomly selected residuals. For each cell:

$$C' = A + \frac{r'_S \times sd}{\sqrt{m}}$$

4c. Calculate the all-period volume-weighted average (VWA) of the sample age-to-age factors.

RANDOMLY SELECTED RESIDUALS

	12-24	24-36	36-48	48-60
2009	-0.772	-0.877	-0.877	-0.317
2010	0.568	1.063	0.636	
2011	0.568	0.636		
2012	-0.877			

SAMPLE TRIANGLE OF FACTORS CALCULATED BASED ON THE RANDOM RESIDUALS

	12-24	24-36	36-48	48-60
2009	2.237	1.208	1.059	1.082
2010	2.288	1.346	1.139	
2011	2.287	1.311		
2012	2.231			

	12-24	24-36	36-48	48-60	60+
VWAS	2.258	1.281	1.091	1.082	1.000

62 The standardized residuals can also be adjusted so that the average residual is zero, but the zero-adjusted residuals are not shown in the example. Also, the standard deviations of the unscaled residuals are calculated under the assumption that the average unscaled residual is zero.

63 As noted in Section 3, Step 4, this is the only option for randomly simulating a new triangle. Other options used by the ODP Bootstrap model are not used.

STEP 5.
COMPLETE THE NEXT DEVELOPMENT PERIOD FOR THE TRIANGLE.

Use the new VWAs to complete the next diagonal to the right of the data triangle.

NOTE: A randomly generated tail factor could also be applied here to extrapolate future development periods – i.e., beyond 60 months in this example.

CUMULATIVE TRIANGLE WITH FUTURE EXPECTED PAYMENTS

	12	24	36	48	60
2009	352	783	1,045	1,183	1,295
2010	255	572	710	750	812
2011	279	638	767	837	
2012	311	717	918		
2013	308	695			

STEP 6.
INTRODUCE PROCESS VARIANCE.

To add process variance in the simulation, replace every future cumulative paid amount with a randomly selected point from a gamma distribution[64] where:

Mean = the cumulative paid loss amount

Variance = Prior Mean x Variance of the Residuals (see Step 3)

CUMULATIVE TRIANGLE WITH FUTURE POSSIBLE OUTCOMES

	12	24	36	48	60
2009	352	783	1,045	1,183	1,295
2010	255	572	710	750	765
2011	279	638	767	911	
2012	311	717	863		
2013	308	698			

STEP 7.
USE DATA TRAPEZOID (INCLUDING N FUTURE DIAGONALS) TO CALCULATE REMAINING UNPAID.

Calculate new VWAs and use them to complete the bottom right of the trapezoid. In this example N = 1.

NOTE: If a tail factor was used in the basic model, then the projected value would be used in the future diagonal and to calculate the VWAs. If the tail factor was extrapolated, then the remaining random factors would be applied – i.e., beyond 72 months in this example.

	12-24	24-36	36-48	48-60	60+
VWAs	2.265	1.249	1.128	1.066	1.000

COMPLETED CUMULATIVE TRAPEZOID WITH FUTURE EXPECTED PAYMENTS

	12	24	36	48	60
2009	352	783	1,045	1,183	1,295
2010	255	572	710	750	765
2011	279	638	767	911	971
2012	311	717	863	973	1,037
2013	308	698	873	984	1,048

64 As a technical note, this is based on the "distribution free" Mack method so any distribution can be used. The gamma distribution is used here only for consistency with the ODP Bootstrap model.

STEP 8.
CALCULATE TOTAL UNPAID AMOUNTS.

Sum the future incremental values to estimate the total unpaid loss by period and in total (in this example, 1,279).

This provides one estimated possible outcome, with its conditional expected value.

TOTAL ESTIMATED ONE-YEAR FUTURE PAYMENTS AND REMAINING POINT ESTIMATE

	12	24	36	48	60	TOTAL
2009						
2010					**15**	15
2011				**144**	60	204
2012			**146**	110	64	320
2013		390	174	111	65	740
						1,279

STEP 9.
REPEAT AND SUMMARIZE.

Repeat Steps 5 through 8 the specified number of times (e.g., 10,000 or more), capturing the resulting cash flows and the unpaid amounts by period and in total for each iteration. The resulting distribution of unpaid amounts from all the iterations can be used to calculate means, percentiles, etc.

TOTAL SIMULATED RESULTS

	MEAN UNPAID	STANDARD ERROR	COEFFICIENT OF VARIATION
2009	0	0	0.0%
2010	71	44	61.3%
2011	157	57	36.3%
2012	376	73	19.3%
2013	755	39	5.1%
	1,359	165	12.2%

USING THE HAYNE MLE ALGORITHM

A Walkthrough of the Berquist Sherman Calculation (Paid Data)

Note that the Hayne MLE algorithm is the same for either claim frequency or claim severity, so we will only illustrate the model for claim severity.

STEP 1.
CONVERT THE CUMULATIVE PAID DATA INTO AVERAGE INCREMENTAL SEVERITIES.

Divide the cumulative paid loss triangle by the ultimate claim count, then subtract to get incremental severity by development period.

CUMULATIVE PAID LOSS DATA

	12	24	36	48	60
2009	352	783	1,045	1,183	1,295
2010	255	572	710	750	
2011	279	638	767		
2012	311	717			
2013	308				

ULTIMATE CLAIM COUNT

2009	19
2010	14
2011	16
2012	18
2013	19

INCREMENTAL PAID LOSS SEVERITY DATA

	12	24	36	48	60
2009	18.53	22.68	13.79	7.26	5.89
2010	18.21	22.64	9.86	2.86	
2011	17.44	22.44	8.06		
2012	17.28	22.56			
2013	16.21				

STEP 2.
FIT MODEL TO THE DATA.

Using the incremental severity data, fit the Berquist-Sherman algorithm using Maximum Likelihood Estimation.

DEVELOPMENT PERIOD PARAMETERS

	12	24	36	48	60
Mean	18.08	23.14	10.67	8.04	8.34
Std Dev	0.32	0.27	1.10	1.88	2.22

TREND & VARIANCE PARAMETERS

	TREND	K	P	AIC
Mean	-0.100	16.71	-2.66	18.85
Std Dev	0.004	3.73	0.66	

STEP 3. CALCULATE PREDICTED MEAN AND STANDARD DEVIATION.

Use the fitted parameters to calculate the estimated mean and standard deviation for each incremental cell.

$$\mu_{i,j} = D_j e^{iT}$$

$$\sigma_{i,j} = e^{K-C_j} \mu_{i,j}^{2p}$$

D = Development Parameters

T = Trend Parameter

C = Claim Count

PREDICTED MEANS

	12	24	36	48	60	TOTAL
2009	17.90	22.91	10.56	7.96	8.24	0.00
2010	17.72	22.68	10.46	7.88	8.16	8.16
2011	17.55	22.45	10.35	7.81	8.08	15.88
2012	17.37	22.23	10.25	7.73	8.00	25.97
2013	17.20	22.01	10.15	7.65	7.92	47.72
TOTAL						97.74

PREDICTED STANDARD DEVIATIONS

	12	24	36	48	60	TOTAL
2009	0.45	0.24	1.85	3.91	3.57	0.00
2010	0.54	0.28	2.21	4.68	4.28	4.28
2011	0.52	0.27	2.12	4.50	4.11	6.09
2012	0.51	0.26	2.05	4.36	3.98	6.25
2013	0.50	0.26	2.05	4.36	3.98	6.25
TOTAL						11.55

SIMULATION STEPS

STEP 4. SAMPLE RANDOM PARAMETERS.

Using the variance / covariance matrix and the fitted parameters, sample new parameters using the multivariate normal distribution.

DEVELOPMENT PARAMETERS

	12	24	36	48	60
Sample	18.52	23.30	12.19	8.08	11.07

TREND & VARIANCE PARAMETERS

	TREND	K	P
Sample	-0.008	13.53	-1.84

STEP 5. CALCULATE SAMPLE MEAN AND STANDARD DEVIATION.

Use the sampled parameters to calculate the sample mean and standard deviation for each incremental cell.

SAMPLE MEANS

	12	24	36	48	60
2009	18.11	23.12	12.09	8.02	10.98
2010	17.96	22.93	11.99	7.95	10.89
2011	17.82	22.75	11.90	7.89	10.80
2012	17.68	22.57	11.80	7.83	10.72
2013	17.53	22.39	11.71	7.77	10.63

SAMPLE STANDARD DEVIATIONS

	12	24	36	48	60
2009	0.96	0.61	2.02	4.29	2.41
2010	1.13	0.72	2.38	5.08	2.85
2011	1.07	0.69	2.26	4.82	2.70
2012	1.03	0.66	2.16	4.61	2.58
2013	1.02	0.65	2.14	4.56	2.55

STEP 6.
INTRODUCE PROCESS VARIANCE.

Sample each incremental cell using the normal distribution and the mean and standard deviations from Step 5.

SAMPLED PAID LOSS SEVERITY WITH PROCESS VARIANCE

	12	24	36	48	60
2009	17.92	23.38	10.45	5.41	12.66
2010	19.40	23.35	14.65	5.79	10.00
2011	18.07	22.95	13.74	-3.49	15.86
2012	18.01	22.94	14.82	17.99	9.95
2013	19.41	22.16	10.83	13.26	9.77

STEP 7.
CALCULATE TOTAL UNPAID AMOUNTS.

Multiply severities by the claim counts and sum the future incremental values to estimate the total unpaid loss by period and in total (in this example, 2,172).

This provides one estimated possible outcome.

TOTAL ESTIMATED INCREMENTAL PAYMENTS

	12	24	36	48	60	TOTAL
2009						
2010					140	140
2011				-56	254	198
2012			267	324	179	770
2013		421	206	252	186	1,064
						2,172

STEP 8.
REPEAT AND SUMMARIZE.

Repeat Steps 4 through 7 the specified number of times (e.g., 10,000 or more), capturing the resulting cash flows and the unpaid amounts by period and in total for each iteration. The resulting distribution of unpaid amounts from all the iterations can be used to calculate means, percentiles, etc.

TOTAL SIMULATED RESULTS

	MEAN UNPAID	STANDARD ERROR	COEFFICIENT OF VARIATION
2009	0	0	0.0%
2010	112	481	428.5%
2011	251	492	195.7%
2012	465	279	60.0%
2013	907	341	37.5%
	1,734	898	51.8%

A Walkthrough of the Cape Cod Calculation (Paid Data)

Note that the Hayne MLE algorithm is the same for either claim frequency or claim severity, so we will only illustrate the model for claim severity.

STEP 1.
CONVERT THE CUMULATIVE PAID DATA INTO AVERAGE INCREMENTAL SEVERITIES.

Divide the cumulative paid loss triangle by the ultimate claim count, then subtract to get incremental severity by development period.

CUMULATIVE PAID LOSS DATA

	12	24	36	48	60
2009	352	783	1,045	1,183	1,295
2010	255	572	710	750	
2011	279	638	767		
2012	311	717			
2013	308				

ULTIMATE CLAIM COUNT

2009	19
2010	14
2011	16
2012	18
2013	19

INCREMENTAL PAID LOSS SEVERITY DATA

	12	24	36	48	60
2009	18.53	22.68	13.79	7.26	5.89
2010	18.21	22.64	9.86	2.86	
2011	17.44	22.44	8.06		
2012	17.28	22.56			
2013	16.21				

STEP 2.
FIT MODEL TO THE DATA.

Using the incremental severity data, fit the Cape Cod algorithm using Maximum Likelihood Estimation.

DEVELOPMENT PERIOD PARAMETERS

	24	36	48	60
Mean	1.26	0.61	0.51	0.52
Std Dev	0.01	0.07	0.09	0.11

ACCIDENT PERIOD PARAMETERS

	LEVEL	2010	2011	2012	2013
Mean	17.99	1.00	0.99	0.99	0.90
Std Dev	0.16	0.01	0.01	0.01	0.02

VARIANCE PARAMETERS

	K	P
Mean	24.85	-4.25
Std Dev	5.24	0.90

AIC
13.92

STEP 3.
CALCULATE PREDICTED MEAN AND STANDARD DEVIATION.

Use the fitted parameters to calculate the estimated mean and standard deviation for each incremental cell.

$$\mu_{i,j} = LA_iD_j$$

$$\sigma_{i,j} = e^{K-C_j}\mu_{i,j}^{2p}$$

L = Level Parameter

A = Accident Parameters

D = Development Parameters

C = Claim Count

PREDICTED MEANS

	12	24	36	48	60	TOTAL
2009	17.99	22.74	10.95	9.12	9.30	0.00
2010	17.94	22.67	10.92	9.09	9.27	9.27
2011	17.72	22.41	10.79	8.99	9.16	18.15
2012	17.80	22.50	10.84	9.02	9.20	29.06
2013	16.25	20.55	9.90	8.24	8.40	47.09
TOTAL						103.57

PREDICTED STANDARD DEVIATIONS

	12	24	36	48	60	TOTAL
2009	0.26	0.10	2.18	4.74	4.36	0.00
2010	0.31	0.11	2.57	5.59	5.15	5.15
2011	0.31	0.11	2.52	5.50	5.06	7.48
2012	0.28	0.10	2.34	5.10	4.69	7.32
2013	0.41	0.15	3.35	7.30	6.72	10.47
TOTAL						15.67

SIMULATION STEPS:

STEP 4.
SAMPLE RANDOM PARAMETERS.

Using the variance / covariance matrix and the fitted parameters, sample new parameters using the multivariate normal distribution.

DEVELOPMENT PARAMETERS

	24	36	48	60
Sample	1.29	0.61	0.47	0.55

ACCIDENT PERIOD PARAMETERS

	LEVEL	2010	2011	2012	2013
Sample	18.30	0.98	0.99	0.99	0.92

VARIANCE PARAMETERS

	K	P
Sample	19.63	-3.62

STEP 5.
CALCULATE SAMPLE MEAN AND STANDARD DEVIATION.

Use the sampled parameters to calculate the sample mean and standard deviation for each incremental cell.

SAMPLE MEANS

	12	24	36	48	60
2009	18.30	23.53	11.11	8.56	9.96
2010	17.98	23.12	10.92	8.41	9.79
2011	18.07	23.24	10.98	8.45	9.84
2012	18.11	23.28	11.00	8.47	9.86
2013	16.87	21.69	10.25	7.89	9.19

SAMPLE STANDARD DEVIATIONS

	12	24	36	48	60
2009	0.11	0.05	0.69	1.78	1.03
2010	0.14	0.06	0.86	2.21	1.28
2011	0.13	0.05	0.79	2.03	1.17
2012	0.12	0.05	0.74	1.90	1.10
2013	0.15	0.06	0.93	2.39	1.38

STEP 6.
INTRODUCE PROCESS VARIANCE.

Sample each incremental cell using the normal distribution and the mean and standard deviations from Step 5.

SAMPLED PAID LOSS SEVERITY WITH PROCESS VARIANCE

	12	24	36	48	60
2009	18.20	23.54	11.07	7.50	11.13
2010	17.83	23.06	11.43	8.48	11.16
2011	18.37	23.22	11.45	7.83	10.95
2012	17.96	23.34	10.31	8.95	8.15
2013	16.86	21.66	10.88	8.48	11.04

STEP 7.
CALCULATE TOTAL UNPAID AMOUNTS.

Multiply severities by the claim counts and sum the future incremental values to estimate the total unpaid loss by period and in total (in this example, 1,939).

This provides one estimated possible outcome.

TOTAL ESTIMATED INCREMENTAL PAYMENTS

	12	24	36	48	60	TOTAL
2009						
2010					156	156
2011				125	175	300
2012			186	161	147	493
2013		411	207	161	210	989
						1,939

STEP 8.
REPEAT AND SUMMARIZE.

Repeat Steps 4 through 7 the specified number of times (e.g., 10,000 or more), capturing the resulting cash flows and the unpaid amounts by period and in total for each iteration. The resulting distribution of unpaid amounts from all the iterations can be used to calculate means, percentiles, etc.

TOTAL SIMULATED RESULTS

	MEAN UNPAID	STANDARD ERROR	COEFFICIENT OF VARIATION
2009	0	0	0.0%
2010	130	145	119.2%
2011	291	212	73.0%
2012	521	243	46.6%
2013	891	574	64.4%
	1,832	790	43.1%

A Walkthrough of the Chain Ladder Calculation (Paid Data)

Note that the Hayne MLE algorithm is the same for either claim frequency or claim severity, so we will only illustrate the model for claim severity.

STEP 1.
CONVERT THE CUMULATIVE PAID DATA INTO AVERAGE INCREMENTAL SEVERITIES.

Divide the cumulative paid loss triangle by the ultimate claim count, then subtract to get incremental severity by development period.

CUMULATIVE PAID LOSS DATA

	12	24	36	48	60
2009	352	783	1,045	1,183	1,295
2010	255	572	710	750	
2011	279	638	767		
2012	311	717			
2013	308				

ULTIMATE CLAIM COUNT

2009	19
2010	14
2011	16
2012	18
2013	19

INCREMENTAL PAID LOSS SEVERITY DATA

	12	24	36	48	60
2009	18.53	22.68	13.79	7.26	5.89
2010	18.21	22.64	9.86	2.86	
2011	17.44	22.44	8.06		
2012	17.28	22.56			
2013	16.21				

STEP 2.
FIT MODEL TO THE DATA.

Using the incremental severity data, fit the Chain Ladder algorithm using Maximum Likelihood Estimation.

DEVELOPMENT PARAMETERS

	12	24	36	48	60
Mean	0.280	0.354	0.169	0.088	0.108
Std Dev	0.010	0.010	0.012	0.016	0.012

VARIANCE PARAMETERS

	K	P
Mean	5.69	-0.46
Std Dev	2.05	0.39

AIC
24.97

STEP 3.
CALCULATE PREDICTED MEAN AND STANDARD DEVIATION.

Use the fitted parameters to calculate the estimated mean and standard deviation for each incremental cell.

$$\mu_{i,j} = S_{i,i} D_j \div U_j$$

$$\sigma_{i,j} = e^{K-C_j} \mu_{i,j}^{2p}$$

S = Cumulative Severity

D = Development Parameters

U = Ultimate (Age-to-Ultimate) Development Parameters

C = Claim Count

PREDICTED MEANS

	12	24	36	48	60	TOTAL
2009	19.10	24.15	11.52	6.01	7.37	0.00
2010	16.83	21.28	10.16	5.30	6.49	6.49
2011	16.72	21.14	10.09	5.26	6.45	11.71
2012	17.59	22.24	10.61	5.54	6.78	22.93
2013	16.21	20.50	9.78	5.10	6.25	41.63
TOTAL						82.76

PREDICTED STANDARD DEVIATIONS

	12	24	36	48	60	TOTAL
2009	1.03	0.93	1.30	1.91	1.59	0.00
2010	1.27	1.14	1.60	2.15	1.96	1.96
2011	1.19	1.07	1.50	2.02	1.84	2.73
2012	1.10	0.99	1.38	1.86	1.70	2.87
2013	1.11	1.00	1.40	1.88	1.71	3.07
TOTAL						5.39

SIMULATION STEPS:

STEP 4.
SAMPLE RANDOM PARAMETERS.

Using the variance / covariance matrix and the fitted parameters, sample new parameters using the multivariate normal distribution.

DEVELOPMENT PARAMETERS

	12	24	36	48	60
Sample	0.271	0.350	0.188	0.104	0.100

VARIANCE PARAMETERS

	K	P
Sample	4.96	-0.23

STEP 5.
CALCULATE SAMPLE MEAN AND STANDARD DEVIATION.

Use the sampled parameters to calculate the sample mean and standard deviation for each incremental cell.

SAMPLE MEANS

	12	24	36	48	60
2009	18.47	23.85	12.81	7.06	6.81
2010	15.91	20.54	11.04	6.08	5.86
2011	16.06	20.74	11.14	6.14	5.92
2012	17.39	22.45	12.06	6.64	6.41
2013	16.21	20.93	11.24	6.19	5.97

SAMPLE STANDARD DEVIATIONS

	12	24	36	48	60
2009	1.39	1.31	1.02	1.74	1.75
2010	1.68	1.58	1.83	2.10	2.11
2011	1.56	1.47	1.70	1.96	1.97
2012	1.45	1.36	1.58	1.81	1.83
2013	1.43	1.35	1.56	1.79	1.81

STEP 6.
INTRODUCE PROCESS VARIANCE.

Sample each incremental cell using the normal distribution and the mean and standard deviations from Step 5.

SAMPLED PAID LOSS SEVERITY WITH PROCESS VARIANCE

	12	24	36	48	60
2009	18.30	23.44	13.31	8.97	9.32
2010	14.75	24.05	11.52	3.35	5.66
2011	14.38	19.17	10.92	2.85	7.82
2012	17.59	22.61	11.09	9.97	5.68
2013	15.83	19.19	12.92	7.18	7.90

STEP 7.
CALCULATE TOTAL UNPAID AMOUNTS.

Multiply severities by the claim counts and sum the future incremental values to estimate the total unpaid loss by period and in total (in this example, 1,628).

This provides one estimated possible outcome.

TOTAL ESTIMATED INCREMENTAL PAYMENTS

	12	24	36	48	60	TOTAL
2009						
2010					79	79
2011				46	125	171
2012			200	179	102	481
2013		365	245	136	150	896
						1,628

STEP 8.
REPEAT AND SUMMARIZE.

Repeat Steps 4 through 7 the specified number of times (e.g., 10,000 or more), capturing the resulting cash flows and the unpaid amounts by period and in total for each iteration. The resulting distribution of unpaid amounts from all the iterations can be used to calculate means, percentiles, etc.

TOTAL SIMULATED RESULTS

	MEAN UNPAID	STANDARD ERROR	COEFFICIENT OF VARIATION
2009	0	0	0.0%
2010	91	37	40.1%
2011	187	57	30.4%
2012	413	66	15.9%
2013	792	78	9.8%
	1,484	142	9.6%

A Walkthrough of the Hoerl Curve Calculation (Paid Data)

Note that the Hayne MLE algorithm is the same for either claim frequency or claim severity, so we will only illustrate the model for claim severity.

STEP 1.
CONVERT THE CUMULATIVE PAID DATA INTO AVERAGE INCREMENTAL SEVERITIES.

Divide the cumulative paid loss triangle by the ultimate claim count, then subtract to get incremental severity by development period.

CUMULATIVE PAID LOSS DATA

	12	24	36	48	60
2009	352	783	1,045	1,183	1,295
2010	255	572	710	750	
2011	279	638	767		
2012	311	717			
2013	308				

ULTIMATE CLAIM COUNT

2009	19
2010	14
2011	16
2012	18
2013	19

INCREMENTAL PAID LOSS SEVERITY DATA

	12	24	36	48	60
2009	18.53	22.68	13.79	7.26	5.89
2010	18.21	22.64	9.86	2.86	
2011	17.44	22.44	8.06		
2012	17.28	22.56			
2013	16.21				

STEP 2.
FIT MODEL TO THE DATA.

Using the incremental severity data, fit the Hoerl Curve algorithm using Maximum Likelihood Estimation.

DEVELOPMENT PARAMETERS

	LEVEL	D	D^2	LN(D)
Mean	9.05	-6.74	0.58	7.55
Std Dev	0.88	0.98	0.10	0.98

TREND & VARIANCE PARAMETERS

	TREND	K	P	AIC
Mean	-0.011	14.78	-2.31	19.20
Std Dev	0.005	2.94	0.53	

STEP 3.
CALCULATE PREDICTED MEAN AND STANDARD DEVIATION.

Use the fitted parameters to calculate the estimated mean and standard deviation for each incremental cell.

$$\mu_{i,j} = e^{(L+jD+j^2D^2+\ln(j)\ln(D)+iT)}$$

$$\sigma_{i,j} = e^{K-C_j}\mu_{i,j}^{2p}$$

L = Level Parameter

D = Development Parameter

D^2 = Development Parameter

Ln(D) = Development Parameter

T = Trend Parameter

C = Claim Count

PREDICTED MEANS

	12	24	36	48	60	TOTAL
2009	17.95	22.95	10.74	6.63	8.08	0.00
2010	17.75	22.69	10.62	6.56	7.99	7.99
2011	17.55	22.44	10.50	6.48	7.90	14.38
2012	17.36	22.19	10.38	6.41	7.81	24.60
2013	17.16	21.94	10.26	6.34	7.72	46.26
TOTAL						93.23

PREDICTED STANDARD DEVIATIONS

	12	24	36	48	60	TOTAL
2009	0.47	0.27	1.55	4.71	2.98	0.00
2010	0.56	0.32	1.85	5.63	3.57	3.57
2011	0.54	0.31	1.77	5.40	3.42	6.40
2012	0.52	0.30	1.72	5.23	3.31	6.42
2013	0.52	0.30	1.72	5.22	3.31	6.42
TOTAL						11.67

SIMULATION STEPS:

STEP 4.
SAMPLE RANDOM PARAMETERS.

Using the variance / covariance matrix and the fitted parameters, sample new parameters using the multivariate normal distribution.

DEVELOPMENT PARAMETERS

	LEVEL	D	D^2	LN(D)
Sample	8.69	-6.94	0.69	7.00

TREND & VARIANCE PARAMETERS

	TREND	K	P
Sample	-0.022	12.16	-2.01

STEP 5.
CALCULATE SAMPLE MEAN AND STANDARD DEVIATION.

Use the sampled parameters to calculate the sample mean and standard deviation for each incremental cell.

SAMPLE MEANS

	12	24	36	48	60
2009	11.10	10.76	5.51	4.89	10.90
2010	10.85	10.53	5.39	4.78	10.65
2011	10.61	10.29	5.27	4.67	10.42
2012	10.37	10.06	5.15	4.57	10.19
2013	10.14	9.84	5.04	4.47	9.96

SAMPLE STANDARD DEVIATIONS

	12	24	36	48	60
2009	0.80	0.85	3.26	4.15	0.83
2010	0.97	1.04	3.97	5.06	1.01
2011	0.95	1.01	3.89	4.95	0.99
2012	0.94	1.00	3.83	4.88	0.98
2013	0.96	1.02	3.90	4.97	0.99

STEP 6.
INTRODUCE PROCESS VARIANCE.

Sample each incremental cell using the normal distribution and the mean and standard deviations from Step 5.

SAMPLED PAID LOSS SEVERITY WITH PROCESS VARIANCE

	12	24	36	48	60
2009	11.86	11.31	6.64	6.55	10.84
2010	10.96	11.28	1.15	-0.47	12.09
2011	9.32	11.10	0.88	1.93	10.07
2012	9.63	9.78	1.35	8.84	9.53
2013	11.08	10.20	10.26	-2.76	9.52

STEP 7.
CALCULATE TOTAL UNPAID AMOUNTS.

Multiply severities by the claim counts and sum the future incremental values to estimate the total unpaid loss by period and in total (in this example, 1,234).

This provides one estimated possible outcome.

TOTAL ESTIMATED INCREMENTAL PAYMENTS

	12	24	36	48	60	TOTAL
2009						
2010					169	169
2011				31	161	192
2012			24	159	172	355
2013		194	195	-52	181	517
						1,234

STEP 8.
REPEAT AND SUMMARIZE.

Repeat Steps 4 through 7 the specified number of times (e.g., 10,000 or more), capturing the resulting cash flows and the unpaid amounts by period and in total for each iteration. The resulting distribution of unpaid amounts from all the iterations can be used to calculate means, percentiles, etc.

TOTAL SIMULATED RESULTS

	MEAN UNPAID	STANDARD ERROR	COEFFICIENT OF VARIATION
2009	0	0	0.0%
2010	116	72	61.5%
2011	234	128	54.8%
2012	449	148	32.9%
2013	886	155	17.5%
	1,686	311	18.4%

[THIS PAGE INTENTIONALLY LEFT BLANK]

B. Using Diagnostics to Improve Your Model

GENERAL BACKGROUND

Even though the actual processes underlying the settlement of individual claims are extremely complex, and in fact unknowable, we can still observe the results of those processes and estimate possible outcomes either for individual claims or in the aggregate for all claims within a specific cohort of exposures.

The traditional deterministic process of evaluating insurance liabilities typically includes a variety of assumptions and methods used to generate a number of point estimates. From a statistical point of view, this estimation process can be characterized as a search for “the” mean loss development pattern, which leads to a “best” **central estimate**.[65]

As we move into the stochastic modeling world, with models such as the ODP bootstrap, that “mean” pattern is still relevant, but it becomes far more important to “build” a model that “captures” all of the statistical features in the data. As such, we are now concerned with estimating “all” possible outcomes and not just the “mean” path. From a statistical point of view, this estimation process can be characterized as a search for “the” loss development model, which leads to an estimate of the “best” **distribution of possible outcomes**.

TAILORING THE MODEL

As noted in Section 1, an advantage of most models is that it can be specifically “tailored” to the statistical features found in the data under analysis. This is particularly important as the results of any simulation model are only as good as the model used in the simulation process. If the model does not “fit” the data then the results of the simulation may not be a very good estimate of a distribution of possible outcomes.

Like all models and methods, the quality of a model’s results depends on the quality of the assumptions. Thus, we need a variety of diagnostic tools to help us judge the quality of those assumptions and to change or adjust the parameters of the model depending on what statistical features we find in the data. In essence, we need the tools to find the model that provides the best fit to the data.[66]

THE DIAGNOSTICS

The diagnostic tests included in the Arius system are designed to either test various assumptions in the model, to allow the user to gauge the quality of the model fit, or to help guide the user in adjusting the model parameters. Most, if not all, of these tests are relative in nature, which means that the analyst can use them to “improve” the fit of the model by comparing the tests from one set of model parameters to another. The model will generally run properly with any set of parameters, so the question the analyst is trying to answer is: “which set of model parameters will give me simulations

65 Using the mean development factors does not imply that the mean or expected value of the distribution is being estimated, since no distribution of possible outcomes is being calculated from which to calculate the expected value. If the aggregate distribution is normally distributed then the mean development factors may lead to a reasonable approximation of the mean of the distribution, but to the extent that the aggregate distribution is skewed then the mean development factors may result in an estimate that differs (perhaps significantly) from the mean of the aggregate distribution.

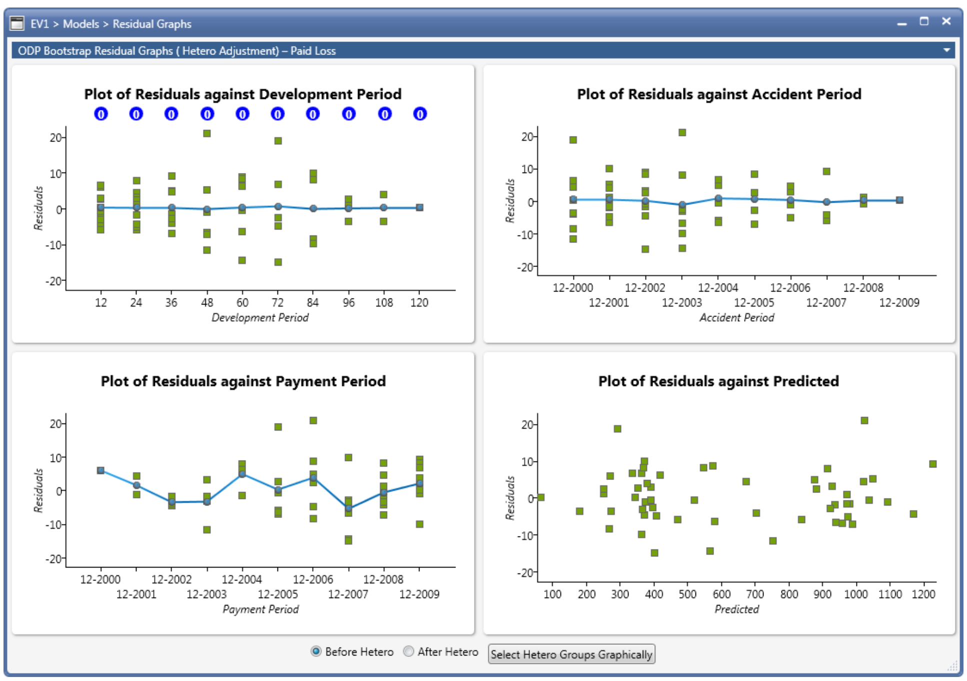
66 For ease of illustration, throughout Appendix B we will discuss these concepts using the ODP Bootstrap model, but they can be similarly applied to the other Arius models.

that are most realistic and most consistent with the statistical features I found in the data (or that I believe are in the data)?"

Also included with some of the diagnostics are statistical tests, which can be viewed as a type of pass/fail test for some aspect of the model assumptions. However, for these statistical tests it is important to note that a "fail" result for a given test does not generally invalidate the entire model. A "fail" signal should only be interpreted as the possibility that a better model (or model parameterization) might be possible, although it is also fair to say that the more "fail" signals you see the more likely it is that the distribution of possible outcomes from that model may not be reasonable.

RESIDUAL GRAPHS

Since the simulations in the bootstrap model use the residual values to create sample triangles of historical data (for the default **Residuals** option), an important diagnostic test is to review graphs of the residuals to make sure that they are indeed random. Moreover, rather than simply looking at the residuals compared to the predicted (or actual) values it can be very useful to also look at them by development period, accident period and calendar period.[67] Consider the residual plots in Graph B-1.



Graph B-1: Residual Graphs *Prior to* Heteroscedasticity Adjustment

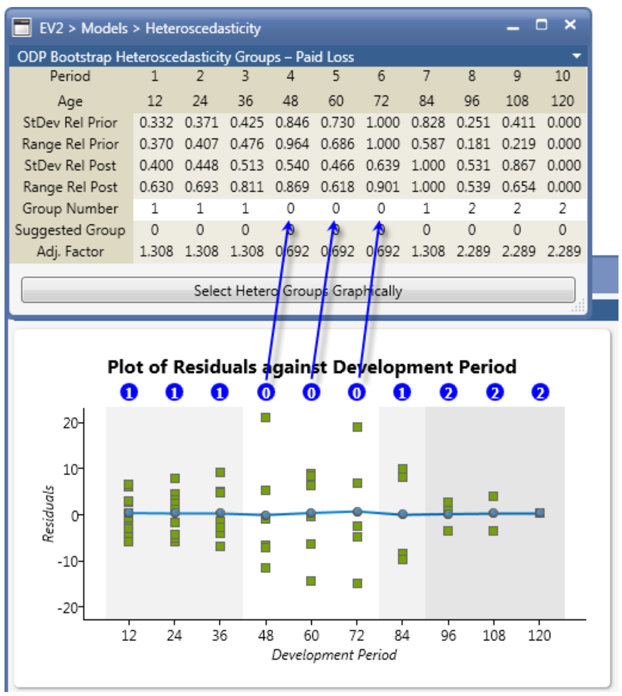
At first glance, the residuals in these graphs appear reasonably random as this model is providing a reasonably good fit of the data. In addition to the residual plots (green dots), the development, accident and calendar graphs also include a trend line (blue line) linking the averages for each period. These graphs are important as they may reveal potential features in the data that we can account for to provide an improved model fit.

67 Note that for each of these four graphs the same residuals are being plotted. In other words, this is four different views of the same data.

In this example there do not appear to be any issues with the trends, but the development period graph does illustrate a common issue with insurance triangle data. In this graph (upper left), you can see that the range of the residuals in the first three periods is not the same as the middle three nor the last two. What this means is the residuals in these three different sub-groups may be representative of three separate standard deviations.

While the ODP bootstrap model does not require a specific type of distribution for these residuals, it is nevertheless important that all of the residuals are independent and identically distributed. This means that, at a minimum, they must share a common mean and standard deviation. This is important since we will be sampling with replacement from all residuals during the simulations. If the residuals do have different standard deviations (as shown in Graph B-1) then this is referred to in statistical terms as *heteroscedasticity*.[68]

In order to adjust for this heteroscedasticity, the ODP bootstrap model allows the analyst to identify groups of development periods with potentially similar standard deviations, and then adjusts the residuals to a common standard deviation value. A set of tools for identifying these groups is included with the model. In the **Heteroscedasticity** table there are sets of relativities for each development period based on the standard deviation and the range of each period, respectively.



EV2 > Models > Heteroscedasticity

ODP Bootstrap Heteroscedasticity Groups – Paid Loss

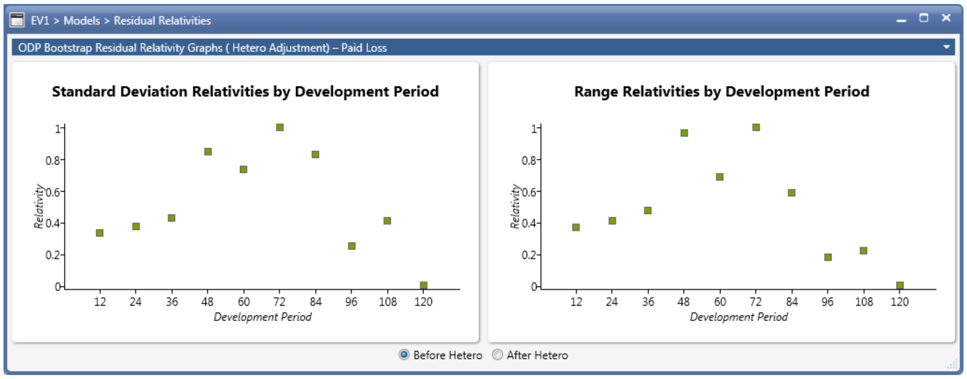
Period	1	2	3	4	5	6	7	8	9	10
Age	12	24	36	48	60	72	84	96	108	120
StDev Rel Prior	0.332	0.371	0.425	0.846	0.730	1.000	0.828	0.251	0.411	0.000
Range Rel Prior	0.370	0.407	0.476	0.964	0.686	1.000	0.587	0.181	0.219	0.000
StDev Rel Post	0.400	0.448	0.513	0.540	0.466	0.639	1.000	0.531	0.867	0.000
Range Rel Post	0.630	0.693	0.811	0.869	0.618	0.901	1.000	0.539	0.654	0.000
Group Number	1	1	1	0	0	0	1	2	2	2
Suggested Group	0	0	0	0	0	0	0	0	0	0
Adj. Factor	1.308	1.308	1.308	0.692	0.692	0.692	1.308	2.289	2.289	2.289

Select Hetero Groups Graphically

Table B-1: Heteroscedasticity Relativities and Groups

68 While some papers on ODP bootstrap modeling have discussed whether the mean of the residuals should be zero or not, this is usually not as important a consideration as the standard deviations and adjusting for heteroscedasticity. Even after adjusting for heteroscedasticity, the mean of the residuals will usually not equal zero. One of the model options available is to adjust the residuals so that the mean is equal to zero, but this might remove a useful feature of the data so we would generally advise the analyst to test the impact of this option.

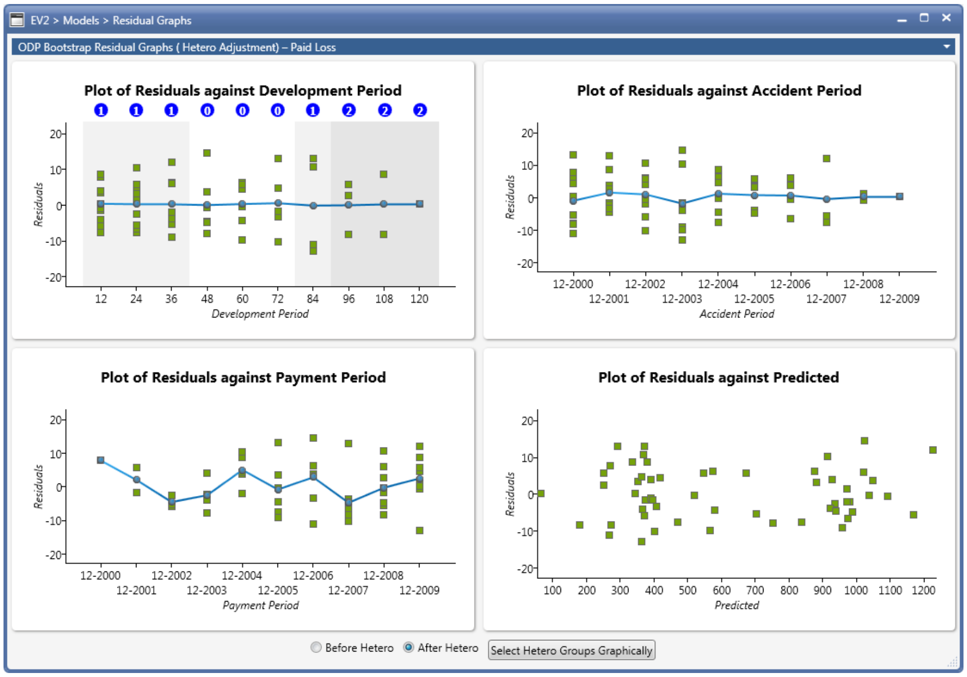
Also in the DIAGNOSTICS collection, there are **Residual Relativities** graphs that can be used to help visualize the likely groupings of periods. These are illustrated in Graph B-2 (below) for the residuals shown in Graph B-1 and Table B-1.



Graph B-2:
Residual Relativities *Prior to* Heteroscedasticity Adjustment

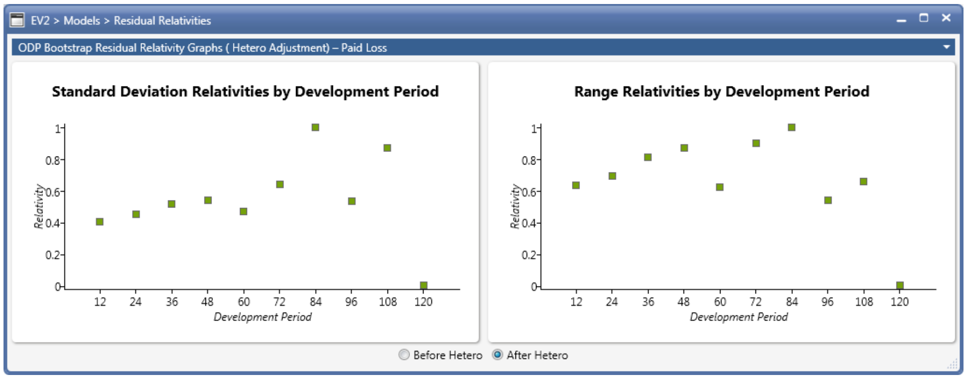
The relativities illustrated in Graph B-2 confirm that the residuals in the first three periods are not the same as the middle four or the last two, but a little testing is required to determine the optimal groups using the other diagnostic tests noted below. For example, looking at the standard deviation relativities period 7 looks like it could belong with periods 4, 5 and 6, but looking at the range relativities it looks like period 7 may belong with periods 1, 2 and 3. In addition, looking at periods 8 and 9 the standard deviation relativities look like they might be grouped with periods 1, 2 and 3 while the range relativities look like they might be a separate group.

Consider the residual plots in Graph B-3, which are from the same data model after the first three development periods (plus period 7) and last two development periods, respectively, have been adjusted (separately). This grouping resulted in better improvement in the other tests described below compared to other groupings that were tested by grouping by eye.



Graph B-3:
Residual Graphs *after* Heteroscedasticity Adjustment

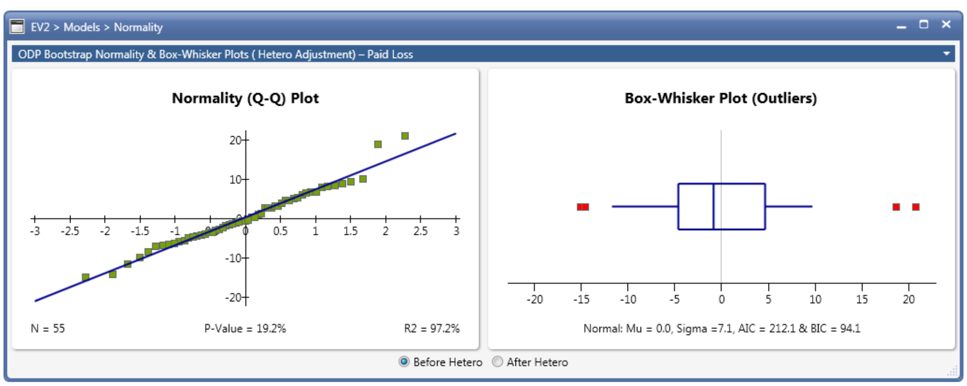
Comparing the residual plots in Graphs B-1 & B-3, you can see that the general "shape" of the residuals has not changed and the "randomness" is still consistent, but now the residuals do appear to exhibit the same standard deviation (or homoscedasticity).[69] Comparing the residual relativities in Graphs B-2 & B-4, you can also see that the relativities are also more consistent.



Graph B-4: Residual Relativities *after* Heteroscedasticity Adjustment

NORMALITY TEST

Another test of the residuals is whether they come from a normal distribution or not. As noted earlier, the ODP bootstrap model does not depend on the residuals being normally distributed, but this is still a useful test for comparing parameter sets and for gauging the skewness of the residuals. For this test, we can use both graphs and calculated test values. Consider the normality plot (sometimes called Q-Q plot) in Graph B-5 for the heteroscedasticity (hereafter "hetero") groups noted above.



Graph B-5: Normality & Box-Whisker Plots *Prior to* Heteroscedasticity Adjustment

Even before the hetero adjustment, the Normality Plot looks quite good. In addition to the graph, the P-Value is a statistical test for normality which can be thought of as a pass/fail test with a value greater than 5.0% generally considered a "passing" score (of the normality test, not whether the ODP

69 During the simulation process, after the residuals are randomly selected from this homoscedastic group they are adjusted back to their original heteroscedastic group sizes to insure than the simulation process produces sample triangles which exhibit the same statistical properties as the original data.

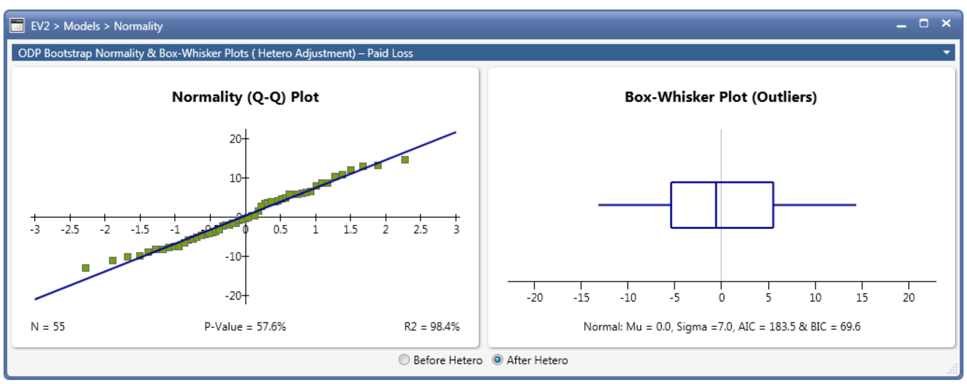
Bootstrap model passes or fails).[70] In addition to the pass/fail test, the P-Value (prior to the hetero adjustment) can be thought of as a gauge for the Process Variance distribution since a low value would be more indicative of the gamma or lognormal distribution and a higher value is more indicative of the normal distribution.[71] Also shown with these graphs is N, which is the number of data points, and the well-known R^2 test.

In the Box-Whisker Plot of Graph B-5 the parameters of the normal fit are shown[72] as well as the Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC). The AIC and BIC are statistical tests similar to the R^2 test, except that they include a penalty for adding more parameters and a lower score indicates a better model fit. It should also be noted that the AIC and BIC tests are only relative to the model parameters being tested and cannot be compared between one model and another. As you can see in Graph B-6, the AIC and BIC increased for the hetero groups being tested which could indicate that the hetero groups are not optimal or that we over-parameterized the model.

OUTLIERS

Another useful test is to check for outliers, or values outside of the "typical" range, in the data. A very useful graphical test for this purpose is a Box-Whiskers plot. Consider the Graph B-5, again for the hetero groups noted above.

Before the hetero adjustment, you can see that there were four outliers in the data model, which correspond to the two highest and two lowest values in all the previous graphs. The basic design of the Box-Whiskers plot is to use the box to show the inter-quartile range (the 25th to 75th percentiles) and the median (50th percentile) of the residuals. The whiskers then extend to the largest values that are less than three times the inter-quartile range. Any values beyond the whiskers are generally considered outliers and are identified individually with a point (in red).



Graph B-6:
Normality & Box-Whisker Plots *after* Heteroscedasticity Adjustment

After the hetero adjustment, in Graph B-6, the residuals do not contain any outliers. If the data continued to contain outliers, the ODP bootstrap model contains an option to remove specific outliers

70 For the interested reader, this is known as the Shapiro-Francia Test. The Null Hypothesis is typically considered to be invalid if the value is less than 5.0%, but as noted earlier failing this test does not generally invalidate the ODP bootstrap model.

71 There are no definitive values here for changing from one distribution to another. But for those looking for guidance on when a different distribution may be beneficial in improving the model fit to the underlying data, this test can be useful for both the **Residual Sampling Distribution** and **Process Variance Distribution** options.

72 If the **Normal Fit** option is used for **Residual Sampling Distribution**, then these parameters are used in the simulation process.

(i.e., cells or observations) from the model.[73] However, if there are still a lot of outliers then this typically means that there is significant skewness in the data which will be replicated in the ODP bootstrap simulations or the model is not sufficiently “optimized” yet—i.e., some other aspect of the model should be adjusted first. Finally, removing only one or a few outliers should be done with caution as they may still represent realistic “extreme” or “skewed” values in the simulations.

SUGGESTED HETERO GROUPS

Even for smaller data sets the process of testing various hetero groups can be a long process and there is no guarantee that the optimal hetero groups[74] have been found. In order to help you find the optimal hetero groups you can use the **SUGGEST HETERO GROUPS** icon in the **HOME** ribbon to run an algorithm to search for the “optimal” groups. The algorithm depends on the size of the data; all combinations of small data sets can be tested, but the number of combinations increases exponentially with size, so with larger triangles different algorithms and time constraints come into play. Essentially, the algorithm searches through the possible combinations of group numbers until it finds the optimal groups for small data sets, but as the size increases it will suggest groups that should be very close to optimal.

EV3 > Models > Heteroscedasticity

ODP Bootstrap Heteroscedasticity Groups – Paid Loss

Period	1	2	3	4	5	6	7	8	9	10
Age	12	24	36	48	60	72	84	96	108	120
StDev Rel Prior	0.332	0.371	0.425	0.846	0.730	1.000	0.828	0.251	0.411	0.000
Range Rel Prior	0.370	0.407	0.476	0.964	0.686	1.000	0.587	0.181	0.219	0.000
StDev Rel Post	0.331	0.596	0.682	0.617	0.729	0.998	1.000	0.303	0.496	0.000
Range Rel Post	0.370	0.655	0.766	0.705	0.686	1.000	0.710	0.219	0.265	0.000
Group Number	1	3	3	0	1	1	2	2	2	2
Suggested Group	0	1	1	3	0	0	2	2	2	2
Adj. Factor	0.898	1.445	1.445	0.657	0.898	0.898	1.087	1.087	1.087	1.087

Select Hetero Groups Graphically

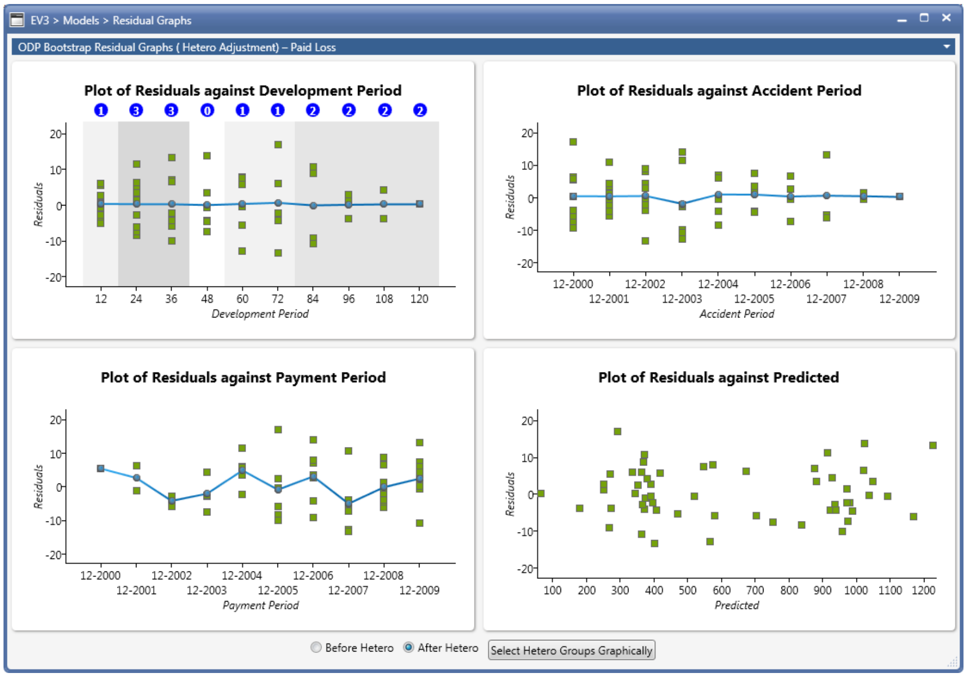
Table B-2: Suggested and Selected Heteroscedasticity Relativities and Groups

After running the SUGGEST HETERO GROUPS algorithm, the calculated hetero groups are shown in the **Suggested Group** row of the **Heteroscedasticity** table, as shown in Table B-2 for the data used in this Appendix. While the group numbers for the Suggested Groups will be optimal they are not necessarily in a particular order. For example, in Table B-2 the Suggested Groups sorted from the smallest to largest Adjustment Factor are 3, 0, 2 and 1. Thus, the entered Group Numbers were switched so that they are in ascending order, but this is not a requirement as you can simply copy and paste the **Suggested Group** row into the **Group Number** row.

Consider the residual plots in Graph B-7, which use the optimal hetero groups from Table B-2. Comparing the residual plots in Graphs B-3 & B-7, you can see that the general “shape” of the residuals has not changed and the “randomness” is still consistent, but interestingly the residuals appear to exhibit different standard deviations like we saw in Graph B-1 prior to adjusting for heteroscedasticity.

73 The data represented by an outlier is not actually removed from the model. Essentially that value is given zero weight in the calculations described in more detail in Section 3 and Appendix A. In addition, a “removed” outlier is never selected as part of the sampling process in the simulations.

74 Optimal in this case means maximum improvement of as many statistical tests as possible with the fewest parameters (i.e., hetero groups).

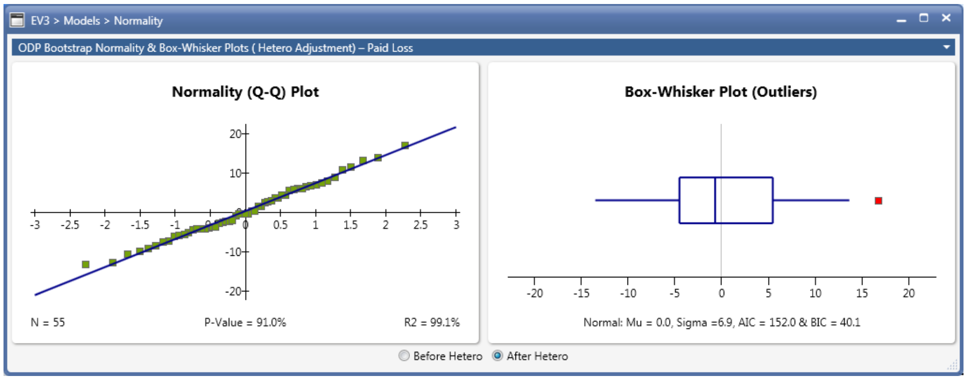


Graph B-7:
Residual Graphs *after* Optimal Heteroscedasticity Adjustment

The reason for this apparent reversal of the "homoscedasticity" of the residuals in Graph B-7 compared to Graph B-3 is related to the discussion of the issues regarding the groups following Graph B-2. By testing different combinations of groups statistically the algorithm checks the significance of each combination for each development period. For example, development period 4 continues to be in group 0 (with the largest variance), but most of the other development periods grouped by eye were moved to different groups considering both the variance and number of observations in each period.

Even though the residuals in Graph B-7 don't look like they are "better" than the residuals in Graph B-3, the other tests tell a different story. Consider the Normality Plot in Graph B-8 for the optimal hetero groups shown in Table B-2.

Compared to the test values in Graph B-6, the P-Value has significantly improved from 57.6% to 91.0% and the R^2 test showed more improvement increasing from 98.4% to 99.1%. More importantly both the AIC and BIC decreased, indicating that the improvement in the other statistics did not come at the expense of over-parameterizing the model. Even though the optimal number of groups is more than we selected by setting the groups by eye, the combination used is more statistically significant. Quite often the optimization algorithm will find a combination with fewer groups than you might use by eye, but that is not necessary true and, in fact, it is possible for the algorithm to use more groups.



Graph B-8:
Normality & Box-Whisker Plots *after* Optimal Heteroscedasticity Adjustment

Finally, we can check the outliers for the optimal hetero groups as shown in Graph B-8. At first glance comparing Graph B-8 to Graph B-6 you could get the impression that the model improvement was not too significant since we still have one outlier and we had previously removed all of them. However, a closer look reveals that the optimal groups continued to produce much more symmetry compared to using no hetero groups. Indeed, as noted earlier removing all outliers is not really the goal since including some “extreme” values in the simulation process will tend to benefit the tails of the overall distribution of possible outcomes. This does not mean that you should never remove outliers, just that you should be confident that they would represent unrealistic “extreme” values that you do not expect to ever reoccur.

C. Correlating Multiple Lines Together

GENERAL BACKGROUND

Correlation is a way of measuring the strength and direction of a relationship between two or more sets of numbers. Essentially, this is a way to measure the tendency of two variables to "move" in the same or opposite directions.[75]

The **correlation coefficient** between two variables is indicated (and calculated) using a range of values from -100% to 100%. With **positive correlation**, the two variables tend to "change" in the same direction – i.e., when the X variable is high, the Y variable tends to be high; if X is low, Y tends to be low. The closer the measured correlation is to 100%, the stronger the tendency is to "move" in the same direction.

Conversely, with **negative correlation**, the two variables tend to "move" in opposite directions – i.e., as the X variable increases, the Y variable tends to decrease and vice versa. The closer the measured correlation is to -100%, the stronger the tendency is to "move" in the opposite direction.

A correlation of zero indicates that no relationship is anticipated between the two variables – i.e., you would not expect to use the values or movements in one variable to tell you anything about the value or movements in the second variable. In statistical terms, this is referred to as being **independent**.

HOW DOES THIS AFFECT UNPAID CLAIM ESTIMATES?

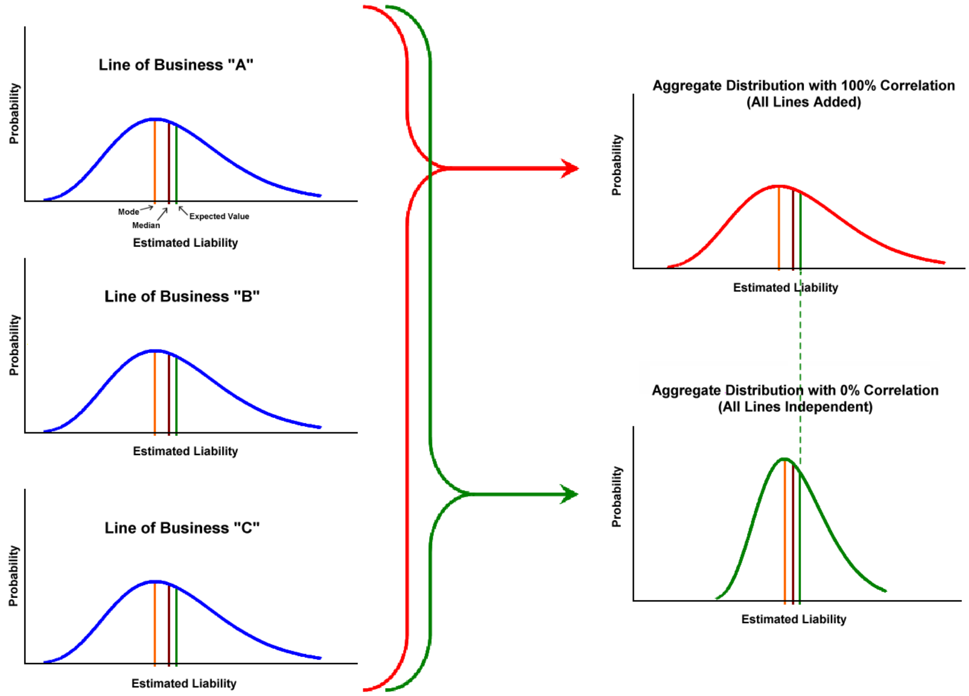
As noted in Section 1, the various models are used to estimate a distribution of possible outcomes for a single block or segment of business. The models are used separately for each segment, without regard to other segments, but each segment is usually only part of an entire book of business for a company. However, rather than simply "add up" the distributions for each segment, in order to estimate a distribution of possible outcomes for the company as a whole we need to take correlation into account.

For most insurance coverages, insurance claims tend to happen independently of each other. For example, Homeowners claims are usually not related to Auto Liability claims. However, one can usually find examples of positive correlation – e.g., catastrophes would tend to cause claims for multiple coverages at the same time – as well as examples of negative correlation – e.g., Workers' Compensation and Unemployment claims tend to move in opposite directions such as when an increase in unemployment causes more Unemployment claims to be filed while there are correspondingly fewer employed workers that can file a Workers' Compensation claim.

As such, these "correlations" between multiple segments will have an impact on a distribution of possible outcomes for all segments combined. If we are only concerned with the expected value of the aggregate distribution, then we can calculate the expected value for each segment separately and add all the expectations together. However, if we are concerned about the distribution or trying to quantify a value other than the mean, such as the 75th percentile, we cannot simply sum the segments. The only time the sum of the distributions would be appropriate for the aggregate is when all segments are 100% correlated with each other – a highly unlikely situation!

75 Movement in this context is relative and does not necessarily imply that the average values of the observations are moving (i.e., the averages may or may not be increasing or decreasing). More simply stated, we are describing the movements between one observation and the next and whether the corresponding movements from one observation to the next for another variable are similar or not. Finally, these movements are also relative in size and shape as, for example, one variable could be very small and stable and the other could be very large and volatile.

The impact of correlation on the aggregate distribution can be illustrated graphically.



Graph C-1:
The Impact of Correlation on Aggregate Distributions

In Graph C-1 we have kept the example simple by starting with 3 identical distributions. Then, the sum of the 3 segments (which is equivalent to assuming 100% correlation) results in the same identical distribution, except that all numbers along the Estimated Liability axis are 3 times as large.

Alternatively, if we assume 0% correlation (i.e., independence) between the 3 segments then the expected value (or mean) would be the same as the sum, but the resulting aggregate distribution is narrower since some positive outcomes would be offset by some negative outcomes, and vice versa, which means that other parts of the distribution will not be the same as the sum. Thus, the values for every part of the aggregate distribution, except the mean, would not equal 3 times the corresponding value for one of the individual distributions.

The degree to which the segments are correlated will influence the shape of the aggregate distribution. How significant will this impact be? That primarily depends upon three factors – how volatile (i.e., wide) the distributions are for the individual segments, the relative values of the amounts and how strongly correlated the segments are with each other. All else being equal, if there is not much volatility then the strength of the correlation will not matter that much. If, however, there is considerable volatility, the strength of correlations (or lack thereof) will produce differences that could be significant.

It is important to note that the correlation between individual segments does not affect the distribution of either segment. It only influences the aggregate distribution of both segments combined.

THE CORRELATION MATRIX

Since we are usually concerned about multiple segments, the best way to quickly see and understand the relationships between the various segments is to use a correlation matrix. This is a symmetric matrix of the various correlation coefficients between each pair of segments.[76] It lists each segment down the first column and across the top row. At each intersection of two segments you will find the correlation coefficient describing the expected relationship between those two segments.

	1	2	3	4	5
1	1.00	0.25	0.11	0.34	0.42
2	0.25	1.00	0.15	0.15	0.27
3	0.11	0.15	1.00	-0.19	0.02
4	0.34	0.15	-0.19	1.00	-0.36
5	0.42	0.27	0.02	-0.36	1.00

Table C-1:
Sample Correlation Matrix

In Table C-1 above, segments 1 and 5 are expected to exhibit the strongest positive correlation with each other, while 4 and 5 are expected to exhibit the strongest negative correlation with each other. In addition to correlation coefficients, an additional output is a matrix of the p-values that correspond to each correlation coefficient in the correlation matrix. For each correlation coefficient, the p-value is a measure of the significance of the correlation coefficient, with a low value (less than 5.0%) indicating the coefficient is significant (i.e., likely to be correct or very close to correct) and a high value (greater than 5.0%) indicating the coefficient is not significantly different from zero.

	1	2	3	4	5
1	0.00	0.03	0.81	0.04	0.02
2	0.03	0.00	0.45	0.63	0.07
3	0.81	0.45	0.00	0.29	0.92
4	0.04	0.63	0.29	0.00	0.06
5	0.02	0.07	0.92	0.06	0.00

Table C-2:
P-Values for Sample Correlation Matrix

In Table C-2 above, segments 1 and 5 exhibit the strongest p-value, while 3 and 5 are not likely to exhibit correlation significantly different from zero.

MEASURING CORRELATION

There are several ways to measure correlation, both parametric (for example, Pearson's R) and non-parametric (Spearman's Rank Order, or Kendall's Tau). Pearson's correlation coefficient may be more helpful if the underlying values are "normally" distributed, whereas Spearman's and Kendall's formulas may be more useful when distributions are not normal.

Our model uses Spearman's Rank Order calculation to assess the correlation between each pair of segments in the model. Rather than calculating the correlation of the residuals (or incremental values) themselves, Spearman's formula calculates the correlation of the ranks of those residuals.[77] The residuals to be correlated are converted to a rank order, and the differences between the ranks of

76 In other words, the top right and bottom left triangles of the matrix are mirror images. In addition, the center diagonal, where each segment intersects with itself, is always filled with 1's because any variable is always perfectly correlated with itself.

77 In point of fact, the model calculates the correlation coefficients for the unadjusted standardized Pearson residuals and the heteroscedasticity adjusted residuals. This allows the user to see the impact of the Hetero adjustment on the correlation of the residuals.

each observation of the two variables, D, are calculated. The correlation coefficient (ρ) is then calculated as:

$$\rho = 1 - \frac{6\sum D^2}{N(N^2 - 1)} \qquad \text{(C.1)}$$

Two examples should prove useful at this point. Consider the residuals for the following two segments:

Table C-3: Sample Residuals for Two Segments

SEGMENT A

	12	24	36	48	60
2009	-1.90	-3.07	3.85	4.38	0.00
2010	1.77	1.51	-0.65	-5.50	
2011	1.10	1.74	-3.95		
2012	-0.48	0.43			
2013	0.00				

SEGMENT B

	12	24	36	48	60
2009	-1.23	2.57	-0.98	-1.64	0.00
2010	-1.14	3.19	-4.74	1.56	
2011	0.16	-4.28	6.94		
2012	2.78	-2.74			
2013	0.00				

From these residuals (resid.), we can calculate the ranks and differences (diff.) as follows for each observation (obs.):

Table C-4: Calculation of Residual Ranks

LOB A			LOB B			RANKS	
OBS.	RESID.	RANK	OBS.	RESID.	RANK	DIFF.	DIFF. 2
1	-1.90	4	1	-1.23	5	-1	1
2	1.77	11	2	-1.14	6	5	25
3	1.10	8	3	0.16	8	0	0
4	-0.48	6	4	2.78	11	-5	25
5	-3.07	3	5	2.57	10	-7	49
6	1.51	9	6	3.19	12	-3	9
7	1.74	10	7	-4.28	2	8	64
8	0.43	7	8	-2.74	3	4	16
9	3.85	12	9	-0.98	7	5	25
10	-0.65	5	10	-4.74	1	4	16
11	-3.95	2	11	6.94	13	-11	121
12	4.38	13	12	-1.64	4	9	81
13	-5.50	1	13	1.56	9	-8	64
							496

The Spearman Rank Order correlation coefficient is calculated as:

$$\rho = 1 - \frac{6\sum D^2}{N(N^2 - 1)} = 1 - \frac{6 \times 496}{13 \times (13^2 - 1)} = -0.363$$

For the two segments represented above, there is negative correlation between them. As a second example, consider the following residuals:

SEGMENT C

	12	24	36	48	60
2009	-1.90	-3.07	3.85	4.38	0.00
2010	1.77	1.51	-0.65	-5.50	
2011	1.10	1.74	-3.95		
2012	-0.48	0.43			
2013	0.00				

SEGMENT D

	12	24	36	48	60
2009	-1.64	-2.74	3.19	6.94	0.00
2010	2.78	1.56	-1.23	-4.74	
2011	0.16	2.57	-4.28		
2012	-1.14	-0.98			
2013	0.00				

From these residuals, we can calculate the ranks and differences as follows:

Table C-6: Calculation of Residual Ranks

LOB C			LOB D			RANKS	
OBS.	RESID.	RANK	OBS.	RESID.	RANK	DIFF.	DIFF. 2
1	-1.90	4	1	-1.64	4	0	0
2	1.77	11	2	2.78	11	0	0
3	1.10	8	3	0.16	8	0	0
4	-0.48	6	4	-1.14	6	0	0
5	-3.07	3	5	-2.74	3	0	0
6	1.51	9	6	1.56	9	0	0
7	1.74	10	7	2.57	10	0	0
8	0.43	7	8	-0.98	7	0	0
9	3.85	12	9	3.19	12	0	0
10	-0.65	5	10	-1.23	5	0	0
11	-3.95	2	11	-4.28	2	0	0
12	4.38	13	12	6.94	13	0	0
13	-5.50	1	13	-4.74	1	0	0
							0

The Spearman Rank Order correlation coefficient is calculated as:

$$\rho = 1 - \frac{6\sum D^2}{N(N^2-1)} = 1 - \frac{6 \times 0}{13 \times (13^2 - 1)} = 1.000$$

Thus, for this second pair the business is perfectly positively correlated.

MODELING CORRELATION

Looking deeper at the previous examples, the residuals for segment C and segment D are identical to the residuals for segment A and segment B, respectively. This was done to illustrate the concept of “inducing” correlation. Even though the values are the same, they are not in the same order, but comparing the relative orders to each other is how we measure correlation – i.e., how likely is one variable to move in the same direction as the other. Thus, the correlation between two simulated variables can be changed by re-sorting one compared to the other – i.e., the desired level of correlation can be induced by re-sorting.

A standard method of simulating correlated variables is to simulate them from a multivariate distribution, after specifying the parameters and the correlations for each variable in the distribution. Unfortunately, this type of simulation is most easily applied when the distributions are all the same and they are known in advance (e.g., if they are all from a multivariate normal distribution). Since we are estimating the distributions with bootstrapping, we don’t know them in advance and they might have different shapes, a re-sorting algorithm proves to be an excellent tool for correlating the model distributions.

In order to induce correlation between different segments in the bootstrap model, we first calculate the correlation matrix using the Spearman Rank Order correlation as illustrated above for each pair of segments. The model then simulates the distributions for each segment separately. Using either the estimated correlation matrix, or one supplied by the user, the model calculates the correlated aggregate distribution by re-sorting the simulations for each segment based on the ranks of the total unpaid for all accident years combined.[78]

Another example will help illustrate this process. As noted above, the first step is to run the bootstrap model separately for each segment. The sample output for three segments is shown below (based on 250 iterations). For this example, we have included results by accident year as well as by future calendar year, which sum to the same total. Note that other parts of the simulation output (e.g., loss ratios) could also be included as part of the correlation process.

78 Note that the resorting is based on total values and the coefficients are based on residuals. Correlating the residuals reflects a far more complex algorithm, but our research has indicated that these different approaches are reasonably consistent. Thus, resorting is an easier, although quite robust, algorithm.

Step 1: Simulate Individual segment Distributions

SEGMENT A

	TOTAL UNPAID							TOTAL CASH FLOW						
	Accident Year							Calendar Year						
ITERATION	1	2	3	4	5	...	TOTAL	1	2	3	4	5	...	TOTAL
1	4	18	219	366	818	...	14,513	3,669	3,268	2,160	2,393	1,469	...	14,513
2	1	75	413	667	1,100	...	16,200	4,280	4,173	2,271	1,887	1,363	...	16,200
3	11	-6	861	773	1,379	...	16,826	4,769	4,120	2,400	1,670	1,347	...	16,826
4	0	126	335	1,299	543	...	17,504	5,233	3,707	3,237	1,505	1,486	...	17,504
:	:							:						
250	38	122	470	575	1,191	...	20,210	4,370	3,767	2,807	2,145	1,450	...	20,210
AVG	19	118	512	782	1,015	...	19,256	5,330	4,260	3,227	2,180	1,613	...	19,256

SEGMENT B

	TOTAL UNPAID							TOTAL CASH FLOW						
	Accident Year							Calendar Year						
ITERATION	1	2	3	4	5	...	TOTAL	1	2	3	4	5	...	TOTAL
1	420	35	446	592	1,212	...	45,151	11,058	12,762	8,898	5,921	3,024	...	45,151
2	233	802	302	1,484	1,621	...	23,077	10,107	6,151	2,458	2,107	941	...	23,077
3	330	177	737	344	2,548	...	37,989	10,990	12,038	7,029	3,847	2,144	...	37,989
4	738	68	589	540	803	...	18,430	5,291	4,377	4,267	1,776	1,833	...	18,430
:	:							:						
250	0	15	440	1,113	2,453	...	30,816	12,148	8,186	7,066	2,008	1,178	...	30,816
AVG	207	500	658	954	2,213	...	31,930	10,072	8,363	5,681	3,395	2,114	...	31,930

SEGMENT C

	TOTAL UNPAID							TOTAL CASH FLOW						
	Accident Year							Calendar Year						
ITERATION	1	2	3	4	5	...	TOTAL	1	2	3	4	5	...	TOTAL
1	0	0	3	3	184	...	4,045	1,445	941	815	239	288	...	4,045
2	0	0	8	2	15	...	3,022	1,432	925	305	235	29	...	3,022
3	8	0	0	39	82	...	3,233	1,181	817	561	324	175	...	3,233
4	0	0	5	40	86	...	3,972	1,475	1,017	748	342	327	...	3,972
:	:							:						
250	0	0	0	121	156	...	2,599	1,113	767	365	142	22	...	2,599
AVG	3	3	5	45	74	...	3,495	1,228	985	584	335	201	...	3,495

Step 2: Rank the Simulation Results

The second step is to sort the simulation results based on the values of the total unpaid for all years combined. Sorting in ascending order gives us the rank of each simulation.[79]

SEGMENT A

	TOTAL UNPAID							TOTAL CASH FLOW						
	Accident Year							Calendar Year						
RANK	**1**	**2**	**3**	**4**	**5**	**…**	**TOTAL**	**1**	**2**	**3**	**4**	**5**	**…**	**TOTAL**
1	27	84	545	425	581	…	12,123	3,787	3,056	1,868	1,312	1,118	…	12,123
2	27	73	250	750	491	…	12,792	3,800	3,166	2,267	936	1,081	…	12,792
3	1	-209	23	377	692	…	12,883	3,240	3,266	2,337	1,841	1,009	…	12,883
4	-25	200	470	488	812	…	13,240	4,100	2,611	2,392	1,440	969	…	13,240
:	:							:						
250	2	307	653	1,366	1,015	…	30,200	4,522	3,472	2,506	1,874	1,519	…	30,200
AVG	19	118	512	782	1,015	…	19,256	5,330	4,260	3,227	2,180	1,613	…	19,256

SEGMENT B

	TOTAL UNPAID							TOTAL CASH FLOW						
	Accident Year							Calendar Year						
RANK	**1**	**2**	**3**	**4**	**5**	**…**	**TOTAL**	**1**	**2**	**3**	**4**	**5**	**…**	**TOTAL**
1	136	216	263	584	697	…	15,042	5,271	4,843	2,268	1,697	334	…	15,042
2	30	434	222	142	1,822	…	15,658	6,157	4,088	2,122	1,274	1,623	…	15,658
3	463	101	597	104	1,031	…	17,012	6,930	6,794	1,244	954	283	…	17,012
4	182	148	618	835	1,411	…	17,236	6,957	5,052	2,522	1,942	265	…	17,236
:	:							:						
250	382	1,167	1,333	2,264	3,256	…	64,376	14,018	16,007	12,046	9,279	3,238	…	64,376
AVG	207	500	658	954	2,213	…	31,930	10,072	8,363	5,681	3,395	2,114	…	31,930

SEGMENT C

	TOTAL UNPAID							TOTAL CASH FLOW						
	Accident Year							Calendar Year						
	1	**2**	**3**	**4**	**5**	**…**	**TOTAL**	**1**	**2**	**3**	**4**	**5**	**…**	**TOTAL**
1	0	27	18	200	59	…	2,285	789	747	349	211	95	…	2,285
2	92	5	1	27	15	…	2,367	807	496	652	143	116	…	2,367
3	0	0	0	4	2	…	2,451	673	818	374	160	351	…	2,451
4	1	0	0	5	90	…	2,524	1,068	658	145	343	213	…	2,524
:	:							:						
250	0	0	0	215	211	…	5,351	1,318	1,159	1,009	629	442	…	5,351
AVG	3	3	5	45	74	…	3,495	1,228	985	584	335	201	…	3,495

79 This step is actually for illustration purposes only. During actual simulations, only the ranks are needed for each iteration (separately by segment), which are then used in Step 4.

Step 3: Generate Correlation Matrix

The third step is to calculate the rank orders that will give us the desired correlation between each segment pair. The method illustrated here is to use a multivariate normal distribution with the desired correlation matrix and simulate random values from that distribution. Using the simulated values, the ranks of those simulated values will give us the rank orders for the desired correlation.[80]

	MULTIVARIATE NORMAL			RANKS		
ITERATION	A	B	C	A	B	C
1	-2.652	-0.978	0.030	1	41	128
2	-0.070	0.445	1.136	118	168	218
3	0.030	0.643	-0.536	128	185	74
4	0.915	1.491	0.274	205	233	152
:	:			:		
250	-0.885	1.080	0.412	47	215	165

Step 4: Re-Sort the Simulation Results Based on Correlation Ranks

The fourth step is to re-sort the individual simulation results for each segment based on the rank orders from step three. After this re-sorting, the correlation coefficients for each pair should match the desired correlations specified in the correlation matrix from step three.

SEGMENT A

	TOTAL UNPAID							TOTAL CASH FLOW						
	Accident Year							Calendar Year						
RANK	1	2	3	4	5	…	TOTAL	1	2	3	4	5	…	TOTAL
1	27	84	545	425	581	…	12,123	3,787	3,056	1,868	1,312	1,118	…	12,123
118	0	293	525	805	661	…	18,819	5,095	3,899	2,812	1,826	1,648	…	18,819
128	1	439	388	1,038	1,215	…	19,079	4,838	3,637	3,255	1,995	2,609	…	19,079
205	10	386	362	1,110	1,080	…	21,851	4,398	4,341	4,008	2,644	2,665	…	21,851
:	:							:						
47	1	78	516	697	740	…	16,779	5,551	4,405	2,638	1,708	1,129	…	16,779
AVG	19	118	512	782	1,015	…	19,256	5,330	4,260	3,227	2,180	1,613	…	19,256

SEGMENT B

	TOTAL UNPAID							TOTAL CASH FLOW						
	Accident Year							Calendar Year						
RANK	1	2	3	4	5	…	TOTAL	1	2	3	4	5	…	TOTAL
41	37	240	135	462	1,170	…	23,945	6,994	6,500	5,222	2,546	1,982	…	23,945
168	134	705	212	1,604	3,135	…	33,764	9,755	9,307	5,790	3,849	1,819	…	33,764
185	329	18	149	1,163	1,579	…	37,006	12,598	9,414	7,085	3,852	2,311	…	37,006
233	181	227	962	897	1,505	…	46,286	14,619	11,230	7,839	4,036	4,029	…	46,286
:	:							:						
215	55	997	763	2,100	2,944	…	41,587	13,804	8,545	8,269	4,718	2,923	…	41,587
AVG	207	500	658	954	2,213	…	31,930	10,072	8,363	5,681	3,395	2,114	…	31,930

80 As a technical note, the multivariate T random number generator used in our model uses the Cholesky decomposition to induce the desired correlation between independent random values.

SEGMENT C

	TOTAL UNPAID							TOTAL CASH FLOW						
	Accident Year							Calendar Year						
RANK	1	2	3	4	5	...	TOTAL	1	2	3	4	5	...	TOTAL
128	0	0	0	3	70	...	3,480	1,456	788	672	167	65	...	3,480
218	0	12	0	55	159	...	4,036	1,565	1,461	314	274	195	...	4,036
74	0	0	2	24	29	...	3,163	1,373	824	388	410	157	...	3,163
152	0	0	0	99	112	...	3,595	1,027	1,061	560	377	452	...	3,595
:	:							:						
165	0	0	0	8	17	...	3,644	1,205	853	786	237	357	...	3,644
AVG	3	3	5	45	74	...	3,495	1,228	985	584	335	201	...	3,495

Step 5: Sum the Correlated Values

The fifth step is to sum the re-sorted values from Step 4 across the segments to get aggregate results for each iteration.

TOTAL ALL SEGMENTS COMBINED

	TOTAL UNPAID							TOTAL CASH FLOW						
	Accident Year							Calendar Year						
ITERATION	1	2	3	4	5	...	TOTAL	1	2	3	4	5	...	TOTAL
1	64	324	680	890	1,822	...	39,548	12,237	10,344	7,762	4,026	3,165	...	39,548
2	134	1,010	737	2,464	3,956	...	56,619	16,415	14,667	8,916	5,950	3,661	...	56,619
3	330	457	540	2,225	2,824	...	59,249	18,809	13,875	10,728	6,256	5,078	...	59,249
4	191	612	1,324	2,106	2,698	...	71,731	20,044	16,632	12,406	7,057	7,146	...	71,731
:	:							:						
250	56	1,075	1,279	2,805	3,701	...	62,010	20,560	13,803	11,693	6,662	4,408	...	62,010
AVG	230	621	1,175	1,781	3,301	...	54,681	16,629	13,609	9,493	5,910	3,928	...	54,681

Step 6: Summarize

The final step is to use the aggregate results for all simulations to describe the distribution of unpaid claims from all the results, including means, percentiles, etc. All of the same summaries that are created for each individual segment can also be created for the aggregate results (e.g., cash flows, loss ratios, or graphs, etc.)

[THIS PAGE INTENTIONALLY LEFT BLANK]

[THIS PAGE INTENTIONALLY LEFT BLANK]