Individual Claim Development Models and Detailed Actuarial Reserves in Property-Casualty Insurance

Chris G. Gross, ACAS, MAAA

April 6, 2021

Abstract[1]

Actuarial reserving techniques using aggregated triangle data are ubiquitous in the property casualty insurance industry. By instead starting with the modeling of individual claim behavior using predictive modeling techniques and a modeling framework that describes the full life cycle of a claim, there are numerous benefits including greater reliability of reserve estimates, faster recognition of underlying mix changes, and avoidance of problems in pricing due to differences in development. Component development and emergence models used in conjunction with simulation of currently outstanding claims and simulation of claims still yet to be reported form an alternative framework for generating estimates of reserve need. Algorithmic case reserves at the claim level and algorithmic IBNR estimates at the policy level, actuarially determined and designed to be unbiased, provide valuable information for downstream analyses, a bridge to the generally accepted triangle reserving paradigm, and a means for demonstration of reliability for actuarial purposes.

[1] The author would like to thank Tim Davis and Kevin Madigan for their valuable contributions, comments, questions, and suggested edits to this monograph, with which it was improved significantly.

Contents

Note regarding the use of 'IBNR' in this monograph:

Common actuarial practice, financial reporting, and the existing actuarial literature often refer to Incurred But Not Reported reserves (IBNR) as the difference between the estimated total reserve for future payment and the current case reserve, i.e. inclusive not only of those claims that have not been reported — referred to variably as "emergence", "pure IBNR", "true IBNR", "Incurred But Not Yet Reported (IBNYR)", etc. — but also inclusive of additional development on reported claims — referred to variably as "case development", "Incurred But Not Enough Reported" (IBNER), etc.

When triangle methods are used to develop losses, this distinction is not always important, but since this monograph is focusing on actuarial reserve analysis at the *individual claim level*, the traditional, generalized, definition of IBNR is confusing, and repeated clarification is cumbersome.

Unless specifically noted to the contrary, **"IBNR" will refer to only the provision for those claims that have truly not been reported** as of the date of the analysis, and "case development" will be used to refer to the provision for additional development for known claims.

1.0 Introduction

This monograph describes approaches to incorporate detailed claim and exposure data into the actuarial process of estimating liabilities for property-casualty insurance companies. Approaching loss development with detailed data provides additional insight into needed reserves. Since loss development considerations are critical to questions of pricing, significant insight can be gained in actuarial pricing as well. Internal management reporting also benefits from detailed modeling of reserve development, allowing for more reliable reporting of results at various levels of detail.

The monograph first describes the use of detailed claim development in a predictive modeling framework for performing diagnostic analysis, which is auxiliary to a standard triangle-development approach. Examples of measuring potential changes in case reserve adequacy, claim closure rates, the identification of potential "jumper" claims, and a framework to measure price change to use in the reserving process will be presented.

Then the monograph goes on to describe a simplified approach to incorporating detailed data into the actuarial reserving process through the development of an actuarial case reserve algorithm. We will illustrate the use of this algorithm as an alternative to using the actual booked case reserves and discuss the benefits of considering this technique within the general reserve estimation framework. Testing of the algorithm using report-period triangles will be discussed.

Next, the monograph describes a process for building policy-level reserves for unreported claims (unreported claim value algorithm) to generate an estimate of the value of claims not yet reported. We will then discuss using this unreported claim value algorithm, applied both currently and retrospectively, to build a new kind of triangle in order to test, support and further develop the estimated reserve.

The monograph then goes on to describe a more comprehensive "development component" modeling and simulation approach to building estimates of future claim payments at the claim and policy level. Using predictive models of components of claim development (such as closure activity, claim revaluation, payment rates, etc.) across various policy and claim characteristics can shed significant light on the reserve analysis. Simulation becomes the bridge to take such component models and the current inventory of open claims and written policies to an actuarial estimate of the total reserve. The advantages of using this approach are discussed (greater accuracy, better understanding, etc.), as well as the incorporation of the approach into the case-algorithm and unreported claim value algorithm approaches previously discussed.

The many potential implications of using detailed actuarial claim and unreported claim reserves to support actuarial pricing efforts as well as internal management reporting will be discussed.

There are many dimensions to this topic, particularly when, in many ways, it is a departure from standard loss development techniques. This monograph is, as a result, somewhat circular. We start with the concept of building a case reserve algorithm to actuarially determine case reserves as this is one of the quickest ways to inject detailed information about claims directly into the standard actuarial triangle analysis process. But in order to use all of the available claim information in building such an algorithm, we need an unbiased estimate of the ultimate value of the currently open claims. An initial high-level adjustment to modify case reserves to remove bias is suggested, with the case actuarial reserving algorithm being a refinement to it that reflects additional differentiation across claims. The more comprehensive approach of building development-component models and applying simulation seeks to describe the behavior of claims as an initial step towards developing the estimate of ultimate payments for open claims. In this more complicated case, the actuarial case reserve algorithm becomes a convenient “wrapper” to the complex model to simplify it for purposes of understanding, validation, and practical business use.

2.0 Why Individual Claim-based Actuarial Reserving Approaches Are Needed

The use of analysis based on aggregated triangle data is well establish and engrained within the actuarial profession. The output from triangle analyses is well understood and embraced by management and professionals across the property casualty industry.

Computing power and increasing use of multi-variate predictive analytics in actuarial science are making it possible to substantially improve on this historical framework by systematically considering detailed claim and policy information. It is well known and widely discussed in the actuarial literature that underlying changes in claim mix and/or case reserve adequacy, if left undetected and unadjusted for, can lead to erroneous results. Most approaches within the profession have focused on identifying such distortions by inspection of the aggregated triangle data.

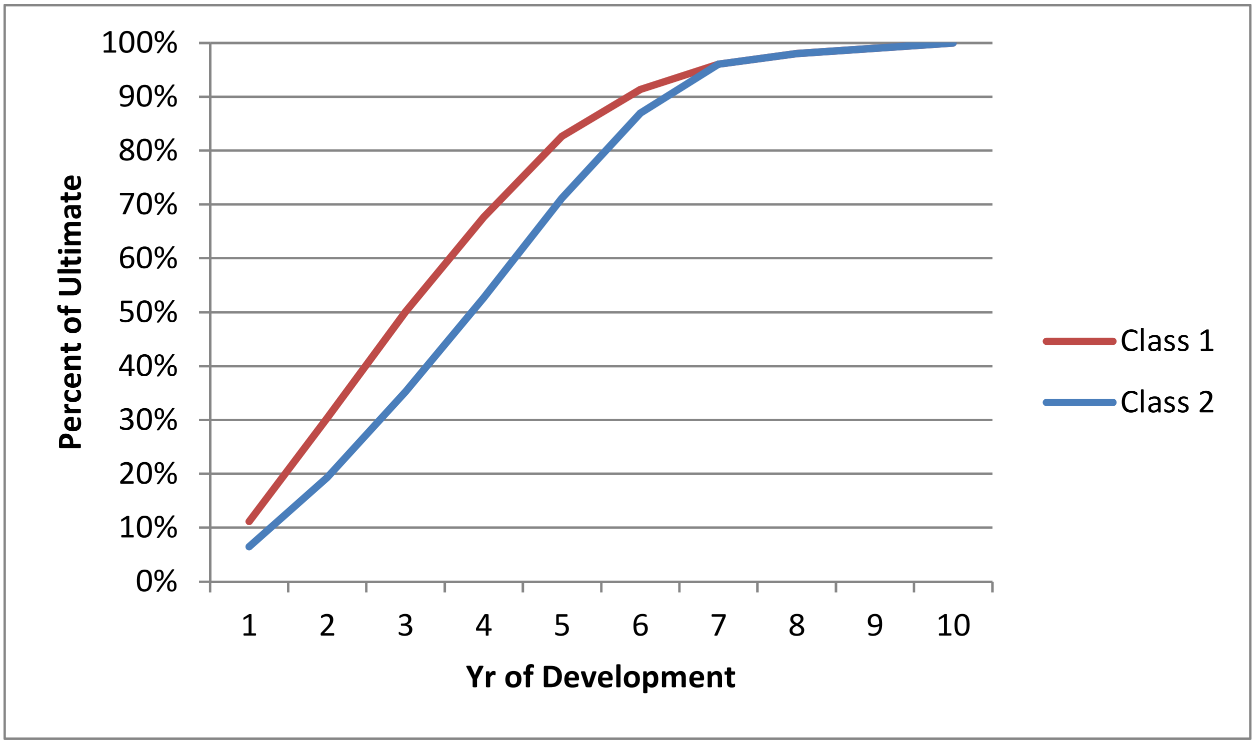
To illustrate why moving to individual claim-based reserving techniques is beneficial, several topics can be identified in which the reliance on aggregated data can be improved by the use of detailed data. We will discuss each of the following topics below:

- Mix shifts
- Changes in case reserving/timing
- Problems with downstream application (allocations)
- Changes in environment
- Cohesive framework across reserving and pricing
- Layer results/Reinsurance

2.1 Mix Shifts

Any significant differences in aggregate loss development patterns that exist across different types of claims or differences in expected loss ratios that exist across different policies have the potential to cause problems for triangle-based reserving processes, unless the mix of claims and exposures is held reasonably constant over the period of time represented by the triangle. This problem is well-known, but due to the wide variety of exposures (different deductibles, locations, policy forms, customer characteristics, etc.) problems with changes in mix can remain hidden for years without detection. Often such problems are not identified by the reserving actuary until years after the shift occurs when changes in the observed triangle development have been shown to be consistently different than in the past.

As an example, consider a hypothetical underwriting unit that writes two classes of business: Class 1 and Class 2. Until recently there has been a stable mix of business between these two classes with Class 1 making up the vast majority of the business. Further, the book has performed well at an acceptable loss ratio a little over 60%. However, due to this acceptable loss ratio, the limited resources of the actuarial department, and the relatively insignificant amount of Class 2 business, difference in performance of the two classes has remained undetected. Class 1 is faster developing and has a lower expected loss ratio of 60%. Class 2 develops slower and has a higher expected loss ratio of 90%. The graph below shows the different expected development patterns of the two classes. The difference may appear slight, but its impact is significant. One will note that at year 3, Class 1 is about 50% developed whereas Class 2 is only about 35% developed. The biggest differences in development happen around years 3, 4, and 5.



Differing development patterns of Class 1 and Class 2 business.

As mentioned above, the mix of business between the two classes had been very stable, but starting in 2013, much more of Class 2 business started to be written. The triangle shown below was generated by assuming the ultimate losses for each year are equal to the expected losses from each line, but that the development of losses happens with random deviations applied to each expected development pattern. The resulting triangle of age-to-age development factors is shown in the next table below.

	Premium		Loss as of:									
Year	Class 1	Class 2	Age 1	Age 2	Age 3	Age 4	Age 5	Age 6	Age 7	Age 8	Age 9	Age 10
2006	100	5	7.53	20.40	32.67	43.49	52.72	58.08	61.20	62.36	63.28	64.50
2007	100	5	8.06	20.72	32.65	43.52	54.68	60.16	63.87	64.15	63.71	
2008	100	5	6.48	19.23	30.80	42.47	52.70	58.32	60.99	62.91		
2009	100	5	7.21	19.21	30.81	42.44	52.93	59.64	61.78			
2010	100	5	7.43	21.88	34.36	43.89	53.76	59.81				
2011	100	5	6.76	19.19	33.07	43.90	54.42					
2012	100	5	7.11	18.49	30.01	40.40						
2013	100	20	8.44	22.18	37.25							
2014	100	40	8.65	25.87								
2015	100	60	9.81									

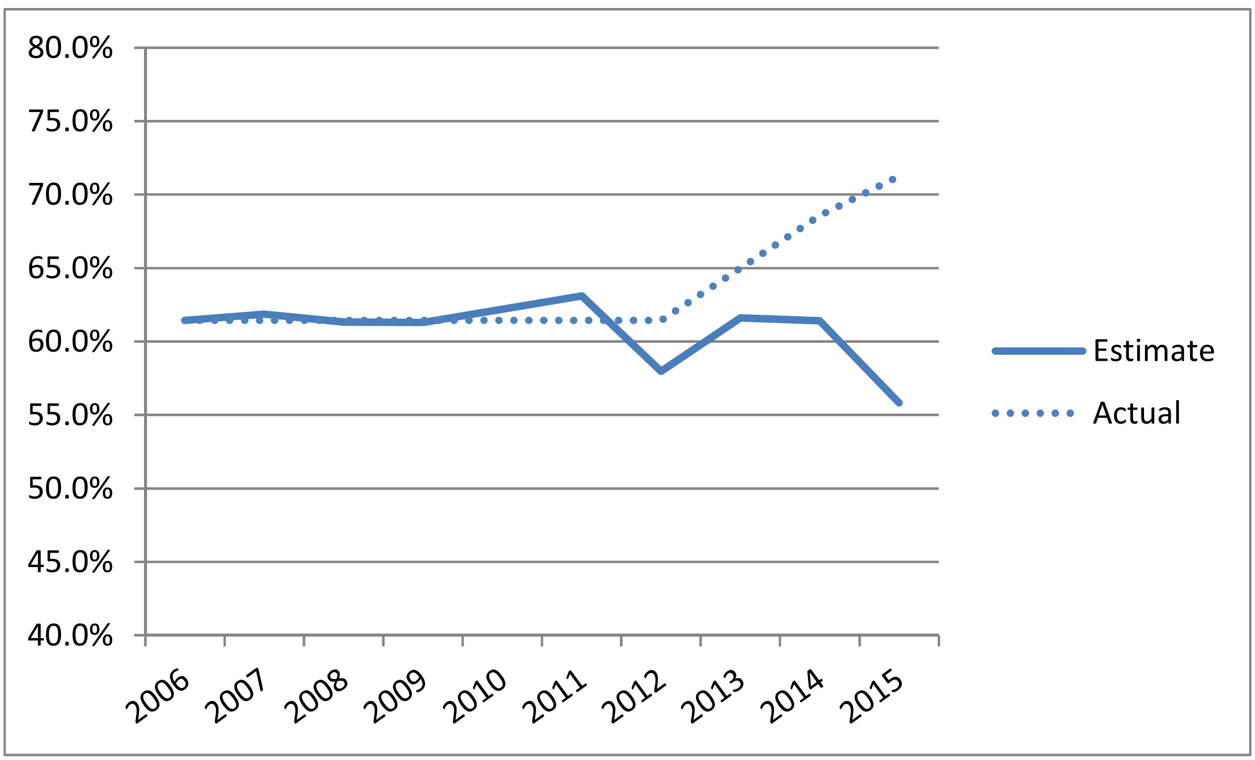
Loss Triangle

2006	2.709	1.602	1.331	1.212	1.102	1.054	1.019	1.015	1.019
2007	2.571	1.576	1.333	1.256	1.100	1.062	1.005	0.993	
2008	2.967	1.602	1.379	1.241	1.107	1.046	1.031		
2009	2.666	1.604	1.378	1.247	1.127	1.036			
2010	2.944	1.570	1.277	1.225	1.113				
2011	2.840	1.724	1.327	1.239					
2012	2.602	1.622	1.346						
2013	2.630	1.679							
2014	2.990								
Last 3	2.740	1.675	1.317	1.237	1.115	1.048	1.018	1.004	1.019
Cumulative	9.108	3.324	1.984	1.506	1.218	1.092	1.042	1.023	1.019

Development Factors

Examination of the triangle would reveal nothing of any particular noteworthiness. The 2014 age 1 to age 2 factor is the highest observed in the triangle, but it's not significantly different from other observed factors. The 2013 age 2 to age 3 factor isn't even the highest observed in the data for this development step. In any case, since the beginning of the change in Class 2 premium in 2013, there are only three data points in the triangle above with which to observe any change in loss development. Since nothing significant is observed, it would be reasonable for the actuary to conclude that no deviations from the indicated development factors should be used. Note that even if a change in development *were* detected within the first two development periods, there is no information in the triangle about how this development will continue beyond age 3.

Applying the derived loss development factors produces the estimated ultimate loss ratios shown in the graph below (via the solid line). These estimated ultimate loss ratios for the last three years, however, are very different from the actual loss ratios of the book (shown via the dotted line).



Estimated Ultimate Loss Ratio vs. True Loss Ratio

The estimated ultimate loss ratio for 2015 is particularly distorted. This is due to lower age 1 losses than what would be historically expected because of the slower reporting of the Class 2 business and the application of a loss development factor which is not appropriate for the current mix of business. Without a mechanism to capture differences in expected loss ratios and development patterns at a level below the triangle, the actuary will be late in identifying the deteriorating loss performance, and will, in fact, predict an improving loss ratio.

This example has demonstrated how the impact of a shift in the mix of business can mislead the actuary when predicting ultimate loss ratios for the three loss years (2013 – 2015) as of the end of 2015. But how much damage can actually be done from this mix shift over the life of the problem?

The first step in thinking about this question is to ask how long it will take for the indications of a higher loss ratio to show up in the triangle. Based on the graph depicting development patterns, age-to-age development on Class 2 business doesn't start to significantly outpace Class 1 business until the year 3 to year 4 step. Given this, there might be a noticeable blip in development in the actuarial analysis that takes place a year later than the one in this example, but even if there is, it is questionable as to whether one blip will capture the attention of the actuary. After several more blips in the following years, it is likely that development patterns will start to be adjusted, but it is unlikely that they will adjust enough based on the data itself. Class 2 business is still a relatively small part of the book through 2015. It moved from roughly 5% of the book in 2012 to about 17% to about 29% to about 38% over the next three years. Further, the changing development pattern first appears only in the early development stages. Class 2 business has a distinctly different development pattern through age 7, at which point the two development patterns become similar. The first evidence of this stronger development at age 7 won't hit the triangle until 2020 when the 2013 loss year reaches 7 years of maturity. Again, it will take several years of higher development to appreciate the impact of Class 2 business at this late development age. Eventually, the ultimate loss ratios for years 2013 – 2015 will come into better focus, but it might not be until four or five years down the road. Understanding the loss ratios for years 2016 and later would probably take even longer.

In the interim, this misdiagnosis of loss ratio can have devastating effects as it will likely influence underwriting decisions. If the increase in Class 2 writings is noticed, it will coincide with a lower initial forecast of ultimate loss ratio due to the slower developing nature of the business. When presented with more new business at a lower ratio, the management decision is easy: write more of it! Class 2 business could very well become a strategic focus of the underwriting unit due to the errant signal the actuarial analysis is providing, and the proportion of Class 2 business written in the book could continue to increase with enthusiasm.

Add to that the growing distortion in the financial statements. New business is being added and booked at a low loss ratio while ultimate liabilities are outpacing their estimates at an accelerated rate. Because of the nature of this shift in business and its misdiagnosed impact by the actuary, the gap between booked reserves and actual needed reserves would grow to a large amount very quickly. When the dust finally settles and the problem is understood, the situation is likely to grab much attention due to the large amount of adverse development which would have occurred. Explaining to management and other stakeholders how this sort of adverse development could have happened is a difficult conversation to have.

It might be easy to dismiss the risks of this situation. In hindsight, the problem is easy to identify – there was a shift in the mix of business by class. It could be that there are mechanisms in place to report on shifts of mix in class so that these changes can be evaluated early, and, as such, this example isn't perceived as posing a realistic threat. The danger comes in not acknowledging that this sort of problem can exist in so many different places and the avoidance of this problem is dependent upon identifying three, interrelated things: 1) that a mix-shift has occurred, 2) a difference in development exists, and 3) there's a difference in loss ratio. The first step is relatively easy. Mechanisms can be set up to monitor for shifts by class, by geography, by deductible, and by any number of other measures. However, the next two steps are very difficult to achieve with common actuarial loss development techniques currently in use. Differences in development will not necessarily be obvious unless triangles are segmented along the dimension in question. It is not feasible to develop and analyze triangles by every slice potentially reflective of differences in development. Further, if differences in development are not isolated, differences in loss ratios cannot be adequately identified until losses are mature. In the example given above, assume that the actuary was monitoring the mix of classes and that they were looking at the loss ratios of the individual classes, assuming that the development patterns were the same. In this case, it would look like Class 2 had seen a dramatic improvement in loss ratio (a mistaken view caused by an insufficient development pattern). Class 2 might be viewed as having a high loss ratio in the past when less of it was written, but now that it is a focus of the business, the loss ratio has improved dramatically. This, of course, is an illusion created by the slower development while in reality, the loss ratio is still high.

The unavoidable conclusion is that it is not at all feasible to believe that one can monitor a book of business for this type of problem across all possible dimensions, without a systematic, multivariate approach to modeling the loss development process and identifying problematic mix-shifts. Changing actuarial loss reserving from the current, aggregated approach is necessary in order to have any reasonable chance of detecting these types of problems before they cause significant financial damage. This monograph outlines considerations regarding how this systematic approach might be implemented and why such an approach would catch potential problems where other approaches would not.

2.2 Changes in case reserving/timing

Changes in case reserving practices within a company can cause significant difficulty in triangle-based reserving approaches. As with mix shifts, the problem is often one of lack of detection. Diagnostics such as triangles of average case outstanding and triangles of closure rates are commonly used as a method to detect changes, but the aggregation of data obscures the view. Changes in case reserve adequacy may occur in such a way as to not be detectable in a triangle until evidenced by changes in loss development factors. For example, consider a scenario in which a company's claim department is under pressure to set case reserves lower while the underwriting department is under pressure to write higher severity accounts. This could result in average case reserve amounts that are similar to the past, despite the drop in case reserve adequacy. The aggregated triangle data and analyses derived from it are ill-equipped to reliably alert the actuary to the changes that can create these reserve and income estimate distortions. Further, the natural variability in loss development clouds the picture and makes it even harder to detect the change. It is financially harmful to wait until the evidence from the aggregated data becomes conclusive. If the changes can be detected before they are reflected in development patterns through systematic investigation of the detailed data, significant damage can be avoided.

2.3 Problems with downstream application (allocations)

It is common to allocate aggregate reserves to a finer level of detail for the purpose of monitoring and managing various segments of the business (profit center, region/office, agency, etc.)

This can sometimes be a clumsy and overlooked final step to a reserve analysis. Typically, the allocation is fairly simple and based on earned premium, outstanding case reserves, payments to date, etc. The algorithms used can create distortions due to inadequacies and over-simplifications of the potential for additional development. When the allocation of these bulk reserves impacts bonus and other incentive payments it is likely to also impact business decisions. If the bulk reserves are allocated in a way that is not appropriate, misguided decisions to expand or contract business can be made. At finer levels of allocation, the allocated bulk reserve can be mistakenly and cynically viewed as simply a tax intended to keep people from receiving a bonus, because those affected do not see any linkage between what happens with actual claims and the allocated bulk reserve.

When the company-wide reserve analysis is performed based on individual claims (and policies in the case of unreported claims), the reserve estimate already exists at the finest level possible. There still will likely be some difference between management's booked reserve and the analysis, but the detailed analysis provides a natural allocation basis for this difference. Not only does this lead to more appropriate business decisions for the dimensions of the business that are already being considered, but also enables powerful management reporting that allows thoughtful drill-downs to other segmentations of the business without additional reserve development analysis. A detailed allocation that is tied directly to open claims, based on their perceived potential to develop differently from each other that ties an IBNR reserve directly to policies reflecting loss ratio and reporting differences across them, will be much less abstract at finer levels of allocation.

2.4 Detect changes in environment

There are often changes within a triangle over time that have nothing to do with mix of business, claims handling practices, or any other action of the insurance company. Examples of these types of changes are inflation, changes in litigiousness, or changes in nature of awards arrived at through litigation.

Companies are certainly aware of and concerned about these changes and are often thinking about them. However, it is likely the case that very little is done to actually detect and measure these changes systematically, especially if the only data being relied on to do so is triangle data. This is because there are several dimensions of time and product-claim mix combined into the triangles, so it is difficult to isolate the impacts of any one of these dimensions.

A predictive analysis in which transaction date is, itself, a predictive variable can be very helpful in detecting these kinds of changes. Such an analysis will help in identifying and measuring impacts observed in the past, including giving context for the relative level of stability or instability in the observed environment.

2.5 Cohesive framework across reserving and pricing

Actuarial reserving functions and pricing functions are often seen as being distinct within a company. Since pricing usually begins with a reserving analysis (explicitly or implicitly), and since reserve estimates are improved by a thorough understanding of changes in pricing and product strategy, the two disciplines are importantly linked. However, when these functions operate separately, important information from each other may not always be communicated. Reserving actuaries aren't always aware of all the changes that are going on across the products being written (what types of mix changes are happening, what pricing decisions are being made, etc.). The pricing actuaries often have too narrow a view which limits their ability to incorporate information more easily seen by looking across business units. This can include changes in reserve adequacy, environmental changes, or even simply the impact of rare but significant losses. They also may make erroneous assumptions about loss development in their pricing analyses by assuming that all policies will see the same proportional development of their losses within a line. This is rarely true and refining this assumption can have a dramatic impact on pricing indications.

Using individual claim-based actuarial reserving approaches forces the reserving actuary to systematically consider things such as changes in the mix of business or changes in price level. If the estimated reserve exists organically at a finer level of detail, it becomes a natural starting point for pricing calculations, providing policy level estimates of reserves. The pricing actuaries can then concern themselves more with the other considerations for pricing as opposed to issues related to development of historical losses.

2.6 Layer results/Reinsurance

When considering expected loss within a particular size of loss layer, either for pricing (deductible, limit) or reinsurance concerns (excess of loss), there are challenges using traditional triangle analysis. Some excess layers may have very limited experience in the observed history. The use of subjectively selected development patterns in different layers can lead to inconsistent results across them. An example is an analysis done gross of reinsurance and another done net of reinsurance, with a relatively thin ceded layer leading to accident periods with negative ceded reserves being implied.

Approaching loss reserving using a detailed approach can help with this problem. By building a model of how individual losses develop over time, including the natural variability around that development, the potential for claims to pierce into individual layers can be considered as part of one cohesive development model.

For a company that is on the ceding side of an excess-of-loss treaty, organizing the necessary information is straightforward. For companies that are on the assuming side or that are writing direct excess coverage, this can be more problematic. Often, however, there is a requirement to report claim activity to the excess carrier/reinsurer once a claim has exceeded a lower threshold, such as when 50% of the retention has been exceeded on a case-incurred basis. Modeling the detailed claim behavior of these sub-layer claims can still add significant amounts of information about exposure to the layer, despite the lack of true first-dollar claim history.

3.0 Predictive Modeling as Reserve Diagnostics and Auxiliary Inputs

The shift from using aggregated triangle data to detailed claim development data in the reserving process is a significant change in mindset. It can be useful to start with some more basic predictive models based on the detailed data instead of attempting to jump directly into development of a completely specified, detailed actuarial reserving model. This can help the actuary gain additional familiarity with the detailed data, identifying potential data problems and needs for additional coding and capturing of data elements.

Some specific predictive models that can be useful

- Case Adequacy Model
- Closure Rate Model
- Identification of “Jumper” Claims
- Price Adequacy

These will each now be discussed in greater detail.

3.1 Case Adequacy Model

The goal of this model is to quantify changes in relative case adequacy (as set by the claims department) over time. Traditionally, the approach used to do this has been to measure the average case reserve (total case reserves divided by open claims) at various evaluation points. These average case reserves can be arranged into a triangle. If a diagonal along this triangle contains cells that are consistently greater than or lower than the values of the cells in the same column, then it is concluded that case reserve adequacy has changed.

The obvious problem with this approach is that the average case reserve may have changed because of shifts in the underlying business from changes in things like class, deductible, limit, geography, etc. A predictive model can be a solution to remove the effects of these distortions.

To set up the predictive model, the target of the prediction is the booked case reserve itself. Include as predictive variables every consistently available attribute of the claims at the policy and claim level of detail. Also include the valuation dates of the case reserves and the development ages at every point at which a case reserve exists (i.e. not just the current valuation) as predictive variables. If the results of this model indicate that valuation date is a statistically significant predictive variable, this is an indication that the reserve adequacy is changing over time. If this were not the case, then the valuation date of the case reserve has no significant impact on the amount of the reserve itself, meaning that reserve adequacy is staying consistent over time.

Since the model adjusts for and quantifies changes in expected reserve amounts which arise from differing mixes of business, the potential distortions mentioned above are less likely to produce misleading results. It could not only correct for what could mistakenly be viewed as a change in adequacy that was actually a change in mix, but it also could reveal an actual change in adequacy that is being masked by a change in mix. As discussed in the previous section, forces which create pressure on the claims department to lower average case reserves that are coincident with pressure on the underwriting department which lead to writing classes of business with higher average severities could offset each other. Without the benefit of the kind of predictive model described above, these two opposing forces could mask the reduction in average case adequacy when looking only at average case reserves over time.

3.2 Closure Rate Model

As the name of this model suggests, the goal of the analysis is to predict the closure rate of claims for future development periods (for simplicity's sake, let's assume the development period is a quarter) relative to the number of open claims at the beginning of the quarter. Another goal can be to measure if the closure rate of claims is changing over time.

Again, the traditional method has been to establish a triangle of closure rates (the number of claims closed during the quarter divided by the number of open claims at the beginning of the quarter for each cell within the triangle). Selections can be made for the closure rate of a given column of the triangle, and analysis can be done to evaluate if the closure rate is changing over some time dimension (column or diagonal).

The problems of mix shifts in business discussed in other sections have a similar distorting effect on the historical data here due to different types of business having different expected closure rates because of the nature of the claims they inherently generate. Similarly, the goal of the predictive model is to neutralize the impacts of historical mix shifts and to project a prospective closure rate that is appropriate for a block of business or a probability of closure in the upcoming quarter for a particular open claim.

To build the predictive model, the data must first be organized. For each historical evaluation point, the list of open claims will be those claims which had a positive case reserve at the beginning of the quarter. Each of these claims will be given a closure value of 0 or 1, depending on if it closed during that quarter. If the open claim closed during the quarter, that claim gets a closure value of 1, and it gets a value of 0 if it did not close during the quarter. The predictive variables are all those consistently available attributes of the claims at the policy and claim level of detail. Also included as predictive variables are valuation date and development age at which the claim is open. The target to be predicted is closure value. Note that the predicted value of this target variable will be a number between 0 and 1, which represents the claim's probability that it will close in the coming quarter. This probability is driven by the development age (which is considered in the traditional method described above) *as well as* all of the attributes of the claim and the policy from which that claim arose.

The indicator of whether the closure rate is changing over time is again the significance or non-significance of the valuation date as a predictive variable. If the valuation date is significantly predictive, then this is an indication that the closure rate of the book of claims is changing, and if so, how much (after adjusting for all other variables).

3.3 Identification of "Jumper" Claims

There is currently a lot of work being done by actuaries and others to predict which claims have the likelihood to "jump," or to grow greatly in value from the level at which it is currently reserved to a much higher level at a later date (Robbin, 2016). What is important in this analysis is whether or not a claim grows by a certain magnitude over the course of time.

To set up this model, each historical open claim would be given a "jumper" value of 0 or 1, much as described above. With this type of model there is an important preliminary question to be answered, namely the time horizon over which the jump occurs. Since the jumper value needs to be either 0 or 1, the entire lifetime of a claim cannot be considered. This is because at the time of the analysis, there will be many claims which are still open, and it would be inappropriate to assign them a value of either 1 or 0 because we don't know what will eventually happen with the ultimate value of those claims. By looking at just closed claims, we would likely get a biased answer. To solve this problem, jumper claims must be defined by two points in time (age) in the claim's life. The set of claims might be defined as all claims with positive case reserves as of the first selected point in time (or age) for the claim. Jumper claims will be identified as those claims which have grown by a certain magnitude by the second selected point in time (or age) in the claim's life. This can be thought of as focusing on two columns of a triangle. For example, the analysis can consider all historical claims which had a positive case reserve as of age 18 months and define jumper claims as those which have grown by a certain magnitude as of the evaluation of the claim at 54 months (or three years later). Note that this specific analysis can be done using any data which is 4 ½ years old or older. This highlights the need for a defined, and relatively short, period to measure jumper claims. The longer your period to measure jumper claims, the smaller and more outdated that data becomes.

With this data convention, there is now a set of claims with a jumper value of 0 or 1, and the predictive analysis can be performed. The prediction will be based on the age of the claims (defined by the two selected points in time or columns of the triangle) as well as the attributes of the claims and the policies from which they come.

This is, however, a limited analysis from an actuarial perspective due to the constraint that it is limited to the two selected points in time. One could certainly run multiple analyses based on different selected points in time, but there will always be the competing forces of desiring a predictive model which identifies a claim which can jump at any point during its lifetime and that of having data which is relevant and not outdated.

Section 6 of this monograph expands on this topic by discussing an approach of using simulation to generate a distribution of ultimate values for open claims at any age. This distribution is based on a predictive time-step model of claim behavior in which all available claim and policy data fields can be used as potential predictive variables. The resulting distribution of ultimate values for an open claim provides the likelihood that it would increase dramatically in size from its current evaluation to ultimate. This approach bypasses the need to identify time periods over which to measure how much a claim can increase in size.

3.4 Price Adequacy

The three preceding examples dealt with estimating or predicting values related to the claims themselves or the claims closing process. In other cases, it may be of value to build a model to create additional fields of interest to use as inputs into the reserving process. One such example is to build a model of relative rate level to generate current rate level premium.

There are many Bornhuetter-Ferguson type models in which premium can be used in the loss reserving analysis, but it is important to ask *what* premium should be used. Collected premium can introduce many inconsistencies due to differing levels of rate adequacy. These inconsistencies will distort the results of any model which is based on premium. What is ideal is a measure of premium at some consistent rate level across all policies. Neutralizing changes in premium rates being charged can play an important role in generating that consistent rate level of premium across all policies. (Bodoff, 2009)

While companies often have processes to measure changes in rates over time, these measures often suffer from a number of problems. One way of measuring rate change is to rerate historical business at current rates and measure the amount of premium generated with current rates relative to the historical premium. This method breaks down when discretionary pricing factors, such as schedule mods, are present. To overcome this challenge, often the change in rate for the renewal book is measured by looking at the relative premium charged for each expiring and renewing policy, which takes into consideration the change in things like schedule mods. However, with this fix, new and non-renewing business is now ignored. If the new business is being written at a higher or lower relative rate than the renewing book, or if the non-renewing business was written at a higher or lower relative rate than the renewing book, the impacts of these changes to the book on relative rate level are not captured.

A predictive model can be used as an alternative approach to simultaneously incorporate all these factors to measure changes in premium rate level from year to year. In this predictive model, the target of our prediction is the collected premium (which is not typically the target of a predictive model), and the exposure base for the policy can be used as a weighting. The predictive variables to be used in this model are all rating factors which can be used to generate the policy premium. This includes class, geography, schedule mods, deductible, limit, etc. In addition to these variables, the policy effective date is the key predictive variable when we are concerned with premium changes over time. The model will assign to each predictive variable the marginal predictive impact of that variable on the policy premium across all policy dates. The policy effective date will, in turn, have its own marginal predictive impact on the policy premium. Since the impacts of all potential rating variables (class, geography, schedule mods, deductible, limit, etc.) have been quantified, the difference in rate level between policy effective dates, adjusting for all other variables, is reflected in the parameters for the policy effective date variable. Interaction effects should be considered between the policy effective date variable and other variables as it may indicate specific rate or other targeted pricing actions taken at different points in time.

As mentioned above, the resulting parameters for policy effective date from the predictive model provide a relative rate level change from one policy effective date to a different policy effective date, adjusting for other variables. As such, the resulting policy effective date curve is a single vector of rate adjustment factors by policy effective date (i.e., a rate level adjustment factor vector) to restate all historical premium to a common level. For aggregated Bornhuetter-Ferguson analyses this may be sufficient.

For making rate-level adjustments to policy-level premium data for use in unknown claim modeling, such as what we will describe in Section 7, a more detailed adjustment is appropriate, but can be made using this same predictive model. The appropriate premium base to use would be the predicted premium for an individual policy based on the predictive factors of all variables included in the predictive analysis *except for* policy date. What is left is a modeled premium that is free of changes in rate over time and reflective of statistically significant impact of key rating variables. It is important to note that this premium is *not* necessarily an estimate of actuarially sound premium, but rather one that is stated at a *consistent* (or benchmark) level of adequacy for all policies within the book being analyzed. As such, this benchmark premium is an appropriate base from which to perform the loss reserve analysis because it is free from distortions in amount which can arise from different rating variables (class, geography, deductible, schedule mod, etc.) or from policy date.

Once this benchmark premium is generated, care needs to be taken to use it appropriately. There are two options on how to use the benchmark premium in a predictive model for loss reserving. One option is to replace the collected premium with benchmark premium outright. However, since the collected premium is also an important benchmark for loss ratio calculations, it may be helpful to instead use both the collected premium and benchmark premium in the predictive analysis. To reduce issues of co-linearity, including the ratio of actual premium to benchmark premium together with the actual premium is a practical strategy.

4.0 Building an Actuarial Case Reserve Algorithm

The use of case reserves by actuaries in the reserve estimation process has been generally beneficial to actuaries, but often detrimental to the insurance companies that they support. They can provide useful information to the actuary about a large portion of the total reserve need. However, changes in how case reserves are established and revised cause significant problems for traditional triangle analysis. This leads actuaries to often put implicit or explicit pressure on claims departments to not change the way that the case reserves are set (or even to deny that changes have occurred). Since the amount of the case reserve can be an important consideration in the handling of the claim, this pressure can lead to sub-optimal decision making and results at the claim level. Because case reserves are rarely established on a true expected value basis, changes in claim settlement rates are also problematic for triangle analysis. This also leads to pressure from actuarial departments to the claim department to maintain the status-quo, potentially leading to sub-optimal economic results as well.

Changes in case adequacy and claim settlement rates are commonly dealt with by actuaries by using Berquist-Sherman techniques to adjust for these changes (Friedland, 2010, ch13). Unfortunately, these adjustments may be inappropriate when applied injudiciously. Consider the hypothetical scenario of a company that, in a difficult market, and with struggling financial results, begins to write higher severity classes of business. Lacking experience in those classes, and desperate for premium, the company underprices the new policies. When loss ratios begin to develop upward, the actuaries, under pressure to reduce reserve estimates, note that the average case reserve amount is higher than in the past. Adjusting historical case reserves to the current level using a Berquist-Sherman adjustment, observed development is flattened, the reserve estimate comes down, and apparent loss ratios are reduced. The company continues to write the policies at an unprofitable level and a large reserve deficiency continues to grow. The problem was not with the Berquist-Sherman technique itself, but rather that it was inappropriate to use it in this situation because the increase in case reserves was not an increase in case adequacy, but rather due to a changing mix of business. Detection of such changes are difficult when only aggregated triangle data is considered, particularly given the wide range of variables that could be shifting (industry classification, geography, deductible, limit, etc.)

The potential for unreliability of subjectively determined case reserves has led some to conclude that case-incurred loss development should be relied on less than paid loss development when estimating total reserves (Zehnwirth, 1994). Taken to an extreme, an actuary may decide that the case-incurred triangle is completely inappropriate to the task based on the unreliability of case reserves over time. This is unfortunate when it occurs because there is some information that is lost by excluding the case reserves. Large organizations may be able to rely only on paid losses without considering the impact of open claims, but in most cases ignoring the information contained in case reserves by the actuary would be imprudent. Consider a small insurer that has seen an abnormally large number of full limits losses. Estimating a total reserve need that is based only on paid losses observed to date and ignoring the case reserves would likely be inappropriate. Even for a large insurer that may be able to rely on paid information only to establish a reserve estimate, the need for information at more granular levels for internal and external reporting purposes, suggests that case reserves are not easily ignored.

Rather than ignoring or discounting the information contained in booked case reserves due to their subjective unreliability, an approach that uses the *objective information* about open claims in a reserve analysis would avoid many of the problems with traditional claim-department case reserves, while still seeking to use the information available about the loss potential of the currently open claims that the actuary was seeking from the case reserves in the first place.

By developing an objective case reserving algorithm, developed by actuaries for actuarial purposes, ***specifically for the purpose of estimating total reserves***, these problems can be avoided, and significant improvement to the information contained in the actuarial analysis can result.

The approach at a high level involves the following steps:

- Organize detailed claim data into a table that lends itself to the specific predictive model to be built (i.e. what would the ideal case reserve have been at various points in time for each claim)
- Adjust open claims to an estimated ultimate value to remove bias. [This can be rough or sophisticated; we will discuss a sophisticated model in Section 6
- Build the Case Algorithm Predictive Model
- Apply the Case Algorithm retrospectively to every open claim for each triangle cell that the claim was open
- Replace the actual case reserves in the case-incurred triangle with the actuarial case reserve
- Develop the triangle as usual

4.1 Organizing the data

The amount that we are trying to predict in the case algorithm model is the total amount of future payments, at any given point in time, for any open claim. It is useful to concentrate on those specific points in time that would be of interest for building a triangle, such as at quarterly development ages. For closed claims, such as the one below, future payments as of various historical points in time are known quantities, and the history of these claims can be restated in these terms.

Claim	Quarter of Development	Case Reserve Balance	Paid to date
A	1	0	0
A	2	0	0
A	3	1000	0
A	4	2000	100
A	5	1500	100
A	6	0	1200
A	7	0	1200
A	8	0	1200

This claim was open at three points of evaluation so we would have three records in our data table:

Claim	Quarter of Development	Paid Remaining
A	3	1200
A	4	1100
A	5	1100

Each claim that is open at any of the triangle evaluation points would have records such as these that would be included in the table. In addition, fields to be used as predictive variables would be included. Some of the fields that are likely to be of predictive value are:

- The age of development
- Prior paid amounts (often useful to separate these into recent payments and older payments as well as indemnity, expense, etc.)
- The claim limit remaining
- Cause of loss
- Injury type
- Geographical area
- Business/Industry classification
- Information about the claimant
- Claim Severity Classification
- The accident period (trend)
- Etc.

Open claims that ultimately close without payment should be included in the table as well, since that is an important potential outcome that should be reflected in the actuarial case reserve estimate.

If only claims that are currently closed are used, with currently open claims excluded from the experience period, there will likely be a downward bias in the estimate due to more severe claims often closing later (in most cases, with a notable exception of total loss on property claims). A possible solution to this problem is to only include claims from accident periods that are essentially closed, but then much information about more recent claims is potentially lost, making this solution unworkable for long-developing lines.

Instead, it is useful to include the open claims, usually with adjustment, removing known biases from the current case reserve. For example, with the following claim:

Claim	Quarter of Development	Case Reserve Balance	Paid to date
B	1	8000	300
B	2	8000	300
B	3	10000	300
B	4	8000	2300

Assume some model has been built to adjust open claims to estimated ultimate value (the next section will discuss ways in which this can be done in more detail). It is now necessary to further organize the data to show amounts yet to be paid. For the example claim above, assume that the case reserve at the current point in time (age 4) is estimated to be 50% adequate based on a study of overall levels of development from that point forward. An estimated ultimate value for the claim is then $8000*2 + 2300 = $18300 and the records that would be added to the table would be:

Claim	Quarter of Development	Paid Remaining
B	1	18000
B	2	18000
B	3	18000
B	4	16000

Note that it is acceptable to use dynamic variables (those that change over time) in a case algorithm. For example, a claim department coded claim severity classification may exist. Care should be taken to make sure that the values of these dynamic variables are the values *as of the time of the prediction*, not as of the most recent valuation. Otherwise, significant distortions can result, as in any predictive model when future-valued predictive variables are inadvertently used.

4.2 Adjust open claims to an estimated ultimate value

In order to develop currently open case reserves to an unbiased (ultimate) level such as in the example above (what we will term here an "un-biasing procedure"), some initial analysis is necessary. In the example above, a very crude technique of using overall case adequacy was applied as a simple illustration. In practice, additional analysis to perform this adjustment is warranted. A useful starting point is to analyze the development in a report period triangle.

Friedland (2010) in chapter 12 describes the case outstanding development technique, wherein losses are developed to ultimate based on developing the case reserve balances along the diagonal. Friedland explains that the use of this approach is limited because it is based on the case reserve and ignores the potential for IBNR claims. However, in the case of modeling the behavior of reported claims only, this is not a problem.

The following formula is a convenient and often used formulation of the approach because it uses development factors that are already commonly used in triangle development.

Formula 1

$$\frac{(\text{Reported CDF to Ultimate} - 1.00) \times (\text{Paid CDF to Ultimate})}{(\text{Paid CDF to Ultimate} - \text{Reported CDF to Ultimate})} + 1.00$$

where CDF is cumulative development factor[2].

Marker and Mohl (1980, Appendix A) refer to the same general approach to basing future development on current case reserves as the "backwards recursive algorithm", named for how ultimate reserve factors are developed from single time-step factors:

P_k = (Paid as of age k+1 - Paid as of age k)/(Age k reserve)

R_k = (Reserve as of age k+1)/(Reserve as of age k)

$D_{N-1,N} = P_{N-1} + R_{N-1}$

$D_{N-2,N} = P_{N-2} + R_{N-2} \times D_{N-1,N}$

$D_{N-3,N} = P_{N-3} + R_{N-3} \times D_{N-2,N}$

…

$D_{0,N} = P_0 + R_0 \times D_{1,N}$

where $D_{n,N}$ is the case reserve development factor which brings reserves at age n to their valuation at age N

This backwards recursive algorithm approach, which is more cumbersome than formula 1, is conceptually consistent with it, but *not technically identical*, as demonstrated by the example below.

[2] The term "CDF" is used to be consistent with the nomenclature of Friedland. Later in this monograph, CDF will be used to mean "cumulative distribution function."

Payments

	Age 1	Age 2	Age 3	Age 4	Age 5	Age 6	Age 7	Age 8	Age 9	Age 10
Yr 1	18,112	4,864	24,726	1,545	352	61	210	3	3	0
Yr 2	51,038	15,806	550	1,463	173	342	329	601	0	
Yr 3	8,973	10,730	4,034	724	180	78	200	10		
Yr 4	42,872	7,837	6,328	1,807	224	468	108			
Yr 5	20,240	8,181	13,019	532	520	189				
Yr 6	18,040	6,731	4,694	2,068	1,023					
Yr 7	13,715	14,594	3,033	2,647						
Yr 8	19,568	11,837	1,067							
Yr 9	19,503	6,808								
Yr 10	17,481									

Case Reserves

	Age 1	Age 2	Age 3	Age 4	Age 5	Age 6	Age 7	Age 8	Age 9	Age 10
Yr 1	8,423	4,258	879	695	441	127	115	103	102	100
Yr 2	6,274	2,287	960	421	1,303	1,016	1,015	4	1	
Yr 3	8,288	4,917	740	341	107	128	18	8		
Yr 4	8,903	4,217	316	1,504	304	17	39			
Yr 5	7,071	3,020	1,270	185	192	15				
Yr 6	5,300	2,049	1,013	234	206					
Yr 7	5,935	4,211	1,210	373						
Yr 8	9,061	3,192	1,588							
Yr 9	5,333	4,102								
Yr 10	6,041									

Case Outstanding Development Technique Using Calculated Averages for Incurred and Paid Development

yr	Cumulative Payments	Cumulative Case Incurred	Case Reserve Balance	Incurred Dev	Pd Dev	Case Outstanding Development Factor	Case Outstanding Development Reserve
1	49,877	49,977	100	1.000	1.000		100
2	70,303	70,304	1	1.000	1.000	0.116	0
3	24,929	24,937	8	1.000	1.000	0.412	3
4	59,643	59,681	39	0.997	1.004	0.593	23
5	42,681	42,697	15	1.001	1.008	1.093	17
6	32,556	32,763	206	1.001	1.013	1.093	225
7	33,989	34,362	373	1.007	1.022	1.471	548
8	32,473	34,061	1,588	1.034	1.059	2.424	3,850
9	26,311	30,413	4,102	1.162	1.282	2.727	11,187
10	17,481	23,522	6,041	1.393	1.810	2.706	16,345
Total							32,300

Backwards Recursive Technique Using Calculated Averages for Paid/Case and Case/Case Factors (each age distinct)

yr	Paid/Case	Case/Case	Backwards Recursive Factor	Backwards Recursive Reserve
1			1	100
2	0.002	0.980	0.983	1
3	0.030	0.959	0.973	8
4	0.535	0.101	0.633	24
5	0.657	0.921	1.240	19
6	0.485	0.555	1.174	242
7	0.731	0.755	1.618	603
8	1.689	0.588	2.639	4,192
9	2.041	0.283	2.789	11,438
10	1.353	0.499	2.745	16,586
Total				33,214

In both approaches, only weighted average factors were used in order to highlight the computational non-equivalence between the approaches. Tail factors of 1.0 in both cases were used as well. The difference is one of weighting. In the case of the backwards recursive approach, the individual factors are weighted by case reserves. For the case outstanding development technique, the paid loss to date is counted heavily in the weighting of factors. Since we are using case reserves as the basis for this projection technique, it is logical to weight on the case reserves themselves to reflect how they behave (i.e. the behavior on years with much case reserves still outstanding is considered more than the behavior in years with little case reserve amounts still outstanding).

Other problems with the case outstanding development technique are introduced when judgment is applied in the selection of paid and incurred development patterns. The actuary is likely to focus on development patterns that are reasonable in themselves and may end up selecting patterns that are *unreasonable* in their implied relationship to *each other,* which is critical for developing the case reserves to ultimate.

The table below adjusts the case outstanding development example by selecting development patterns that are smoothed out in the later periods and that include modest incurred and paid development tail factors, both reasonable, subjective adjustments.

Case Outstanding Development Technique Using Selected Factors for Incurred and Paid Development

yr	Cumulative Payments	Cumulative Case Incurred	Case Reserve Balance	Incurred Dev	Pd Dev	Case Outstanding Development Factor	Case Outstanding Development Reserve
1	49,877	49,977	100	1.010	1.020	2.020	202
2	70,303	70,304	1	1.010	1.020	2.013	2
3	24,929	24,937	8	1.010	1.020	2.008	17
4	59,643	59,681	39	1.010	1.024	1.713	66
5	42,681	42,697	15	1.010	1.029	1.552	24
6	32,556	32,763	206	1.010	1.033	1.441	297
7	33,989	34,362	373	1.017	1.043	1.668	622
8	32,473	34,061	1,588	1.044	1.080	2.298	3,650
9	26,311	30,413	4,102	1.173	1.308	2.681	10,995
10	17,481	23,522	6,041	1.406	1.847	2.706	16,345
Total							32,221

Notice that despite *raising* every paid and incurred development factor, the total reserve estimate using the case outstanding development method *decreased* due to the specific relationship between the selected paid and incurred development factors. It is the relationship between the two patterns that is important for this approach. This can lead to strange results, particularly in the tail. (Note the example given above belies how sensitive results are to the selected tail factors. If the paid tail factor was much different than 1.02, the resulting indicated Case Outstanding Development Reserve would quickly diverge from $32,221 by a significant amount.) The backwards recursive technique, by focusing attention on the relationship between payments and case reserves and the relationship between beginning and ending case reserve balances, is not as sensitive to seemingly small changes in judgment.

When using the backwards recursive technique, it is natural to begin to consider the behavior of open claims as one of a general process rather than as a series of distinct, specific factors. When a typical reserve analysis is performed, selections for both the paid and incurred LDFs must be made at each age of the triangle. These patterns are generally declining in scale, but do not generally describe a consistent process (i.e. each individual factor selection describes a part of the curve). In contrast, selecting paid to case factors and case to case factors, leave open at least the *possibility* that these factors could be the same over broad sections of development ages (e.g., all development ages after a certain age have essentially the same behavior with respect to the next time-step). In the backwards recursive table above, it is difficult to say with any certainty that the relationship of incremental payments to beginning case reserve balances and the relationship of ending case reserves to beginning case reserves is meaningfully different after 4 years of development, but instead that these may actually constitute a stable process and can be considered together. This can be very useful when claim settlement activity starts to become less frequent.

Backwards Recursive Technique Using Calculated Averages for Paid/Case and Case/Case Factors (later ages grouped and tail implied)

yr	Paid/Case	Case/Case	Backwards Recursive Factor	Backwards Recursive Reserve
1			1.686	169
2	0.606	0.640	1.686	2
3	0.606	0.640	1.686	14
4	0.606	0.640	1.686	65
5	0.606	0.640	1.686	26
6	0.606	0.640	1.686	348
7	0.606	0.640	1.686	629
8	1.689	0.588	2.679	4,255
9	2.041	0.283	2.800	11,484
10	1.353	0.499	2.751	16,620
Total				33,611

When paid to case ratios P and case to case ratios R are assumed to be constant after a particular age, the tail factor for the backwards recursive approach can be calculated as P/(1-R).

4.3 Build the Case Algorithm Model

It is natural to ask, "If we already have a way to bring the current individual case reserves along the diagonal to an estimated ultimate value through the un-biasing procedure, then why would we need to build a case algorithm? Don't we already have one?"

In some cases, such as the approach described above, the estimated ultimate value for individual case reserves might be very rough. While the aggregate estimate may be unbiased, there may be many biases that exist claim to claim. The case algorithm will create estimated ultimate values which are more appropriate for each, individual claim. In other cases, the estimated ultimate value for individual case reserves may have been rigorously generated through a complex simulation process (such as the one described in Section 6), in which the estimates are already appropriate for each, individual claim. In this situation, the case algorithm should be considered a "wrapper" to the complex analysis that simplifies and condenses the results for easier use in other contexts, such as audit or analysis of pricing adequacy.

Also, one of the very problems that we are trying to avoid is the impact of changing case reserve adequacy over time. Case reserves that are subjectively determined may include valuable information that is not available in coded fields, but they open the door to these changes over time. By constructing an algorithm that does not depend on actual case reserves when it is applied, we avoid this problem. While the current case reserves themselves are important in giving us information about future payments in the un-biasing procedures discussed above, we seek to use that information, and then condense and assign that knowledge back to objective, consistent, predictive fields. That means eliminating the actual case reserves from the algorithm.

Another benefit of building an actuarial case reserving algorithm is that although the un-biasing procedure is a reasonable first step, it is likely to contain a flaw in that there is an implicit assumption of independence across development periods. In the process of chaining through the various time periods to create ultimate factors in the un-biasing approach discussed above, we are essentially assuming that $E(XY) = E(X)E(Y)$, which is not true when X and Y are correlated. Note that this is not only a problem with the un-biasing procedure, but also with the basic chain-ladder approach itself. In loss development generally, if there is a shock loss (large factor), it often is accompanied by lower development relative to the loss to date (smaller factors) going forward. That is an example of negative correlation. If there is a change in the mix of claims or in the overall environment, there is likely to be positive correlation. When there is positive correlation, multiplied factors will be biased low as an estimate of total development, and when there is negative correlation, multiplied factors will be biased high as an estimate.

In building an actuarial case algorithm, we are jumping from each particular valuation of the claim value to the *ultimate* value, incorporating all the contributors to piecewise correlation, at least for the now closed claims. For the open claims, however, the report period adjustment still has this potential problem. It will generally be considerably less serious than other sources of bias from using unadjusted case reserves, and since the case algorithm relies on both closed and open claims, the remaining impact of bias from unrecognized correlation is reduced.

With bias removed for current case reserves, at least generally, we can now proceed to provide a more general actuarial case reserving algorithm to provide an alternative estimate of case reserve for any claim at any point in time (most notably as of triangle evaluation points).

Consider the table of claims below in which there are three claims: A, B, and C. A is the claim used in the example at the beginning of the "Organizing the Data" section of this monograph. It is a closed claim that closed for $1,200 in its 6th quarter of development. B and C are both open claims as of the date of the analysis (12/31/2015), which have respective unbiased, estimated ultimate values of $1,750 ($1,250 of which has already been paid) and $1,500 (of which none has been paid to date). Also included for each claim are one predictive variable based on policy information (risk state) and one predictive variable based on claim information (cause of loss). This data is prepared in such a way as to build a predictive case algorithm.

Claim	Evaluation Date	Quarter of Development	Paid to Date	Remaining Paid	Risk State	Cause of Loss Code
A	9/30/2014	3	0	1200	19	K
A	12/31/2014	4	100	1100	19	K
A	3/31/2015	5	100	1100	19	K
B	12/31/2014	4	0	1750	27	R
B	3/31/2015	5	0	1750	27	R
B	6/30/2015	6	1250	500	27	R
B	9/30/2015	7	1250	500	27	R
B	12/31/2015	8	1250	500	27	R
C	6/30/2015	2	0	1500	31	K
C	9/30/2015	3	0	1500	31	K
C	12/31/2015	4	0	1500	31	K

With the data organized in this way for all the claims, the goal is to build a predictive model which will give estimates for the remaining paid amount based on policy and claim characteristics as of various points of evaluation. Development age should be considered as one of the predictive variables. Other examples of potential predictors are risk state, cause of loss, and evaluation date (see the important notes about trend and evaluation date a couple of paragraphs below). The model would seek to identify and isolate the impact each predictive variable has on the expected remaining payments to be made. By isolating these impacts, if there are shifts in the book over time across the predictive variables, the distortions to development of those shifts can be quantified.

Note that there are many more predictive variables which can be used compared to what is shown in this example. The amount of paid losses to date for the claim is likely to be a predictive variable. Additionally, the paid losses to date can be categorized in ways that provide additional predictive power. One example is to distinguish between payments made in the most recent period and those paid in previous periods as two separate variables. Also, it can be very helpful to break claim payments into their different components, such as indemnity payments, medical payments, expense payments, etc. Once claim payments are broken out into their different components, there are different choices to be made about how the model is structured. For example, a model might seek to forecast all claim components simultaneously by developing estimates of future payments for indemnity components at the same time estimates of future medical payments are made, using all the subcategories of historical payments as predictive variables. Alternatively, the model can be structured to forecast only future indemnity payments, but still using all payment categories as predictive variables. Section 8 goes into more detail about how these models can be structured.

In an earlier paragraph, evaluation date was highlighted as a predictive variable along with a note about the use of this variable as regards the impacts of trend. Care needs to be taken when using evaluation date in this predictive model. It can be a powerful way to incorporate the impacts of trend, but it can also defeat the very purpose of the model if not used carefully.

The application of trend can be an important consideration when developing an actuarial case algorithm model. There is predictive value in the evaluation date simply because of the force of a trend over time. This can be thought of as a generalization of the Berquist-Sherman technique in that case reserve (unbiased, actuarially based, expected future payments on claims in this case) are systematically adjusted to reflect systematic differences over time. However, the indiscriminate use of evaluation date in the predictive model can be very counter-productive to the task that it sets out to accomplish.

If the evaluation date is used as a predictive variable with no constraints put upon it, it is quite likely that the parameter for evaluation date will fluctuate, perhaps in an apparently random way, from early evaluation dates to later evaluation dates. The evaluation date factor is attempting to isolate the impact of evaluation date after adjusting for mix shifts in the other variables being used, but additional non-included variables may be implicitly captured. In addition, changes in case reserving philosophy over time will get captured in this variable. Also, development age and evaluation date are highly related, and although a multivariate model attempts to isolate their relative importance, it is easy for development age impacts to bleed into the evaluation date variable if it is unconstrained (evaluation date becoming a proxy for development age). If care is not taken to constrain this parameter, and then it is applied to generate the algorithmic case reserves, the effect would be to reintroduce some of the biases that the model is working to eliminate.

To keep the evaluation date from having this effect, simple constraints can be put on the calculation of the evaluation date parameter. For instance, it might be reasonable to assume that trend will impact claim severity in a linear fashion or an exponential fashion (or according to some other rule). With the imposition of a rule on the evaluation date parameter (e.g., it must increase year to year according to an exponential curve with a constant exponent), then the model will derive evaluation date parameters that fit this rule while providing the best prediction of future payments on claims. An added benefit of this approach is that it is objective. Certainly, there is judgement required by the actuary in deciding what rule will best describe the impact of trend (e.g., linear, exponential, etc.), but it is not nearly as subjective as selecting an exogenous trend rate.

4.4 Applying the Case Algorithm

Once the Case Algorithm has been constructed, it is straightforward to apply it to all of the open claims at each point in the history, and then to proceed as normal with triangle analysis. There are a few items that deserve comment.

The un-biasing procedure for currently open claims is most appropriately performed by looking at report period triangles (since emergence of IBNR claims does not cloud the picture). However, once the case algorithm is developed, based on all the claims (closed and adjusted open), it can be applied retrospectively to each claim at each point in time, with the available information for that claim. This can be used to build an accident period triangle in which the primary source of development is claim emergence. By separately estimating the case development, late claim emergence becomes isolated and better understood.

While it is tempting to build the triangle elements – paid and (alternative) case by accident period and development age – directly from the detailed development data, there are often individual discrepancies in the data due to mismatches between the loss and exposure data or claims that do not lend themselves, for whatever reason, to the process (coded through an alternative system, etc.). Rather than deal with such discrepancies when reconciling back to the source data, it is more straightforward from an audit perspective to simply provide the adjustments to the claims that can be modified (i.e. are in both sets), summarize the modifications by accident period and development age, and modify the triangle accordingly. This will ensure that all claim payments are still captured and that if an alternative case reserve was not possible for some claims, they will at least be included in unadjusted form.

Once developed, an actuarial case reserving algorithm can be easily applied for future analyses as long as all of the predictive variables are available. Triangles can be quickly updated as a new step in the regular reserve analysis process, without re-determination of parameters. Re-parameterization of the case algorithm parameters can be done less frequently, for example every year instead of every quarter.

4.5 Exploring Results of the Case Algorithm Model

4.5.1 Validation of Results

By having a single case algorithm that is openly and objectively applied to historical points in time in a triangle, and then calculating development factors in the usual way, the efficacy of the case algorithm itself can be justified. Regardless of how the case algorithm was determined, its ability to generate consistent, unbiased aggregate development of reported losses as claims transition from early stages of reporting, through interim payments and into the final settlement of the claims is evidence of the algorithm's strength as an actuarial tool.

The report period triangle is very useful to test the case reserve algorithm and to compare results to those of the unadjusted case reserves. While the accident period triangle that includes the actuarial case reserves will generally exhibit considerably faster convergence to ultimate than one using subjective case reserves, the report period triangle specifically highlights the effectiveness of the implemented case algorithm to provide for future payments on a specific cohort of claims without the impact of late reported claims to complicate the picture.

In a perfect world, the case reserve would be an unbiased estimator of future payments for the claim.[3] If there is a bias, it is beneficial to at least know that it is consistent over time. Both of these can be tested.

For these tests, it is convenient to make the simplifying assumption that age to age factors in the report year triangles (both traditional and based on the actuarial case algorithm) are lognormally distributed[4] for each age of development.

- Organize the cumulative case-incurred triangle by report period and development age
- Calculate the age-to-age factors in the triangle
- Take the logarithm of each age-to-age factor
- Calculate a linear regression of the age-to-age factor logarithms to time, for each age of development, including the standard error for the two regression coefficients.

An unbiased estimate with no change over time is represented by a null hypothesis of zero for each coefficient of each regression. This null hypothesis will fail regularly for typical subjective case reserves. Using a well-constructed actuarial case reserving algorithm will result in fewer and less extreme rejections of this null hypothesis.

While the actuarial case algorithm will still often result in some failures of the hypothesis, the test can be useful to compare alternative formulations of the case algorithm. For example, the importance of

[3] This statement is being made specifically from the perspective of the actuary. From the perspective of the claim department, this may not be true. Section 10 discusses this concept in more detail.

[4] This assumption is being made for calculation simplicity and based on the general characteristics of non-negativity and skewness typically observed in loss development factor triangles. Other distributions could be used if more appropriate.

including a trend parameter across reporting periods can be tested. The level of consistency with the null hypothesis, with and without the inclusion of the trend parameter, can be tallied.

4.5.2 Illustrating Changes in Case Reserve Adequacy

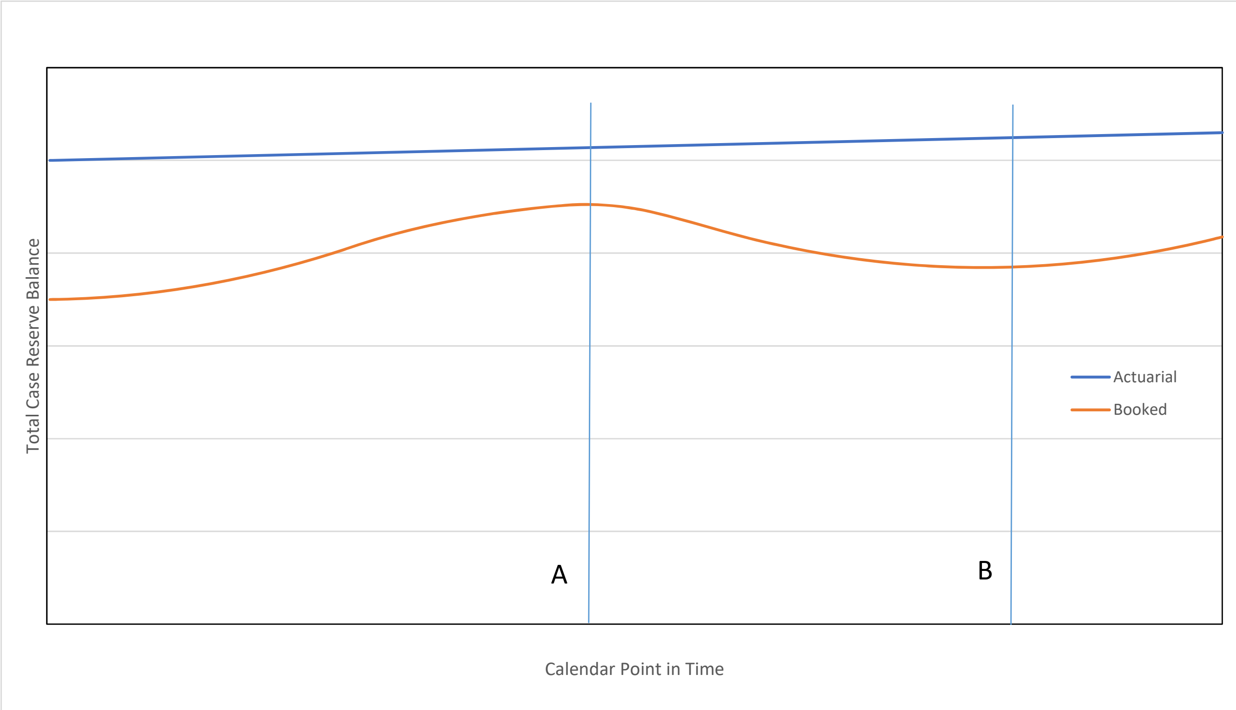
With the case algorithm established, it is illustrative to look at the relationship of the booked case reserves and the actuarial case reserves. It is likely that there will be a significant difference between the two because 1) the booked case reserves are subjectively set and 2) it is not necessarily the intention of the claim department for the booked case reserve to be an unbiased, actuarial estimate of the future payments for the claim. In fact, from a claim management perspective, an actuarially appropriate estimate may very well be less than ideal, because it would reflect relatively rare, extreme events, pushing the mean into higher percentiles, and potentially encouraging higher settlements generally. As stated in many actuarial reserving contexts (e.g. Berquist 1977), inadequacy or redundancy in booked case reserves is not necessarily problematic in a triangle as long as it remains consistent. Comparing the booked case reserves to the objectively determined actuarial case reserves is a useful approach to identifying and illustrating changes in case adequacy and therefore potential distortion to aggregated triangle data. If the relationship between the booked case reserves and the actuarial case reserves was consistent across development age, industry class, geographic area, cause of loss, injury type, etc. (unlikely), or if the mix of business across all of these dimensions is constant (also unlikely), then there is little chance that aggregate triangles would be distorted. However, if these conditions don't exist, then such distortions might be present in the aggregate triangles. The comparison helps identify the extent to which this distortion may be important.

If the relationship between booked case reserves and actuarial case reserves is changing over time for a given development age, then this is an indication that there is a change in case reserve adequacy. This can come about because of many reasons, and, as mentioned earlier, can give misleading indications of ultimate reserve needs.

Even if an explicit change in a claim department process or guideline is not made, reserve adequacy may change simply because of the nature of the claim or the policy from which the claim has come. For example, the claims department might consistently set GL case reserves for a particular profit center at 40% of what is truly needed for claims/policies that originate out of Florida, but also consistently set case reserves at 60% of the need for claims/policies that originate out of Michigan. This consistent difference in case adequacy is important because it could lead to unadjusted actuarial techniques understating ultimate reserve needs if there is a shift to relatively more Florida business or overstate ultimate reserve needs if there is a shift to relatively more Michigan business. Identifying differences in reserve need across all relevant variables is the first step. The second is to measure if there is shift in mix between segments with differing levels of reserve adequacy.

In addition to its use by the actuary in developing an overall reserve estimate, the case reserving algorithm is useful for illustrating changes in case reserve adequacy over time. By comparing the aggregate booked case reserve amount at different points in time to the aggregate case algorithm amount for open claims at the same points in time, a picture of changing case reserves can develop. Looking at the total dollar amount is valuable as is looking at averages per open claim.

Consider the graph below:



The blue line at the top represents the total actuarial case reserves while the orange line represents the total booked case reserves for a book of business over time. The orange line, however, fluctuates up and down relative to the blue line. This displays a changing level of average case adequacy over time.

There can be definite adverse consequences from situations such as the one presented above. The point in time labeled "A" is a time in which case adequacy is relatively high. It is likely that if standard triangle methodologies are used at this point in time that the total reserve need will be overstated. Recently observed LDFs would be high because of the overall case strengthening, and these would be applied to relatively high case reserves. If underwriting decisions were made based on these projections of needed reserves, they would likely be less aggressive than appropriate. Conversely, at the point in time labeled "B", standard triangle-based methodologies would likely understate needed reserves. Recent LDFs are likely small due to falling case adequacy, and they are applied to case reserves which are less strong. Underwriting decisions made based on this information would likely be more aggressive than appropriate.

A chart of this type helps in discussion with management or others with regard to differences between estimates that are aided by the detailed analysis and estimates derived from unadjusted triangles. Because the actuarial case reserves and booked case reserves both exist at the claim level, similar graphs can also be generated for any subset of the data.

5.0 Unreported Claim Value Algorithm

Similar to the use of an actuarially generated case reserving algorithm, actuarial triangle analysis can be further refined by the creation of an unreported claim value algorithm. This algorithm provides the expected value of claims which have not yet been reported for a particular policy as of a particular date.

The Bornhuetter-Ferguson technique (Bornheutter, 1972) can be thought of as a simple case of an unreported claim value algorithm where the loss ratio and development is the same across all exposures of a particular age.

Like the Bornhuetter-Ferguson simple case, we will start with premium (collected, or adjusted to be at constant rate level – see discussion in section 3.4) but will allow loss ratio and reporting lag to vary across policies. Also, we will use policy written premium rather than earned premium as a starting point. As long as we are concerning ourselves with claims that have not yet been reported, and since we are looking at a policy level, there are advantages to also being able to estimate not only the IBNR claims (Incurred But Not Reported), but also the claims that would normally be associated with the unearned portion. For convenience we will term these WBNI claims (Written But Not Incurred). In addition to potentially providing additional information about the losses not yet incurred, using written premium has the advantage that it is more directly comparable across different policies. For example, if two policies are exactly the same, but one has premium that is fully earned and one is only half earned, this is different from two policies that are both fully earned, but where one is half the premium as the other. By considering the policy as a whole and then looking at the pattern of emergence of new claims from the effective date to the valuation date and beyond, we can differentiate between the IBNR and the WBNI claims.

As with the development of the actuarial case algorithm, the organization of the data is a necessary starting point and will help focus understanding. Below is an example (with a few policies shown). Note that many evaluation dates are included for each of these policies.

Policy	Effective Date	Evaluation Date	Days from Effective Date	Paid to Date	Case Reserves	Case Incurred	Hindsight Development	Total from Reported	Hindsight Unreported	Estimated Ultimate
1	1/23/2014	3/31/2014	67	0	0	0	0	0	35,000	35,000
1	1/23/2014	6/30/2014	158	0	0	0	0	0	35,000	35,000
1	1/23/2014	9/30/2014	250	0	1,500	1,500	7,500	9,000	26,000	35,000
1	1/23/2014	12/31/2014	342	0	1,500	1,500	7,500	9,000	26,000	35,000
1	1/23/2014	3/31/2015	432	9,000	0	9,000	0	9,000	26,000	35,000
1	1/23/2014	6/30/2015	523	9,000	0	9,000	0	9,000	26,000	35,000
1	1/23/2014	9/30/2015	615	9,000	18,500	27,500	4,500	32,000	3,000	35,000
1	1/23/2014	12/31/2015	707	12,000	16,500	28,500	3,500	32,000	3,000	35,000
2	8/14/2014	9/30/2014	47	0	10,000	10,000	-1,500	8,500	6,000	14,500
2	8/14/2014	12/31/2014	139	0	10,000	10,000	-1,500	8,500	6,000	14,500
2	8/14/2014	3/31/2015	229	8,500	0	8,500	0	8,500	6,000	14,500
2	8/14/2014	6/30/2015	320	8,500	3,000	11,500	2,000	13,500	1,000	14,500
2	8/14/2014	9/30/2015	412	9,250	2,750	12,000	1,500	13,500	1,000	14,500
2	8/14/2014	12/31/2015	504	13,500	0	13,500	0	13,500	1,000	14,500
3	3/1/2015	3/31/2015	30	0	0	0	0	0	2,500	2,500
3	3/1/2015	6/30/2015	121	0	0	0	0	0	2,500	2,500
3	3/1/2015	9/30/2015	213	0	0	0	0	0	2,500	2,500
3	3/1/2015	12/31/2015	305	0	0	0	0	0	2,500	2,500

From the table above, we can see that Policy 1 has had two claims reported to date (12/31/2015). The first was reported in the third quarter of 2014, was originally under-reserved, and eventually closed for $9,000. The second claim was reported in the second quarter of 2015, is still open, had a partial payment of $3,000 in the fourth quarter of 2015, and has an actuarial estimate of ultimate of $23,000. As of this evaluation, the expected amount of future emergence of as yet unreported claims for Policy 1 is $3,000. Policy 2 has had two claims reported, both of which are closed. The expected amount of future emergence (at 12/31/2015) for Policy 2 is $1,000. Policy 3 currently has no claims reported and an expected amount of future emergence of $2,500, part of which is IBNR and part of which is WBNI since the policy is not yet fully earned. The estimate of ultimate for these three policies as of this analysis (done as of 12/31/2015) is $52,000 ($35,000 + $14,500 + $2,500).

Where did the $2,500, $1,000, and $3,000 come from in the “Hindsight Unreported” column? These are prospective emergence estimations, and yet, the policy unreported claim value algorithm being discussed hasn’t been completed. It’s natural to wonder how these values were determined. A deeper discussion of what is labeled as “Hindsight Unreported” is needed at this point.

The reason “Hindsight” is included in the name of this data field is that for earlier evaluation dates it includes amounts related to claims which have been reported as of the date of the analysis, but had not been reported as of the historical evaluation dates. The amount shown in this column as of the current evaluation date is truly a prospective estimate of claims not yet reported. This estimate, however, is not the final output of the model, and it can come from several potential sources. It could have come from a very rough estimate, such as a Bornheutter-Ferguson based model (assuming that case reserves are developed to ultimate when estimating the emergence pattern). It could have come from the emergence patterns derived from data in which the emergence is deemed to be complete. It could have come from a very sophisticated predictive analysis of expected emergence, such as the component emergence models to be discussed in Section 7. Wherever the estimates came from, the policy emergence algorithm will create a predictive model of expected emergence (IBNR and WBNI) at any point in time for a policy based on the policy (or sub-policy) characteristics. If the initial estimates are rough estimates, the resulting unreported claim algorithm estimates will represent a refinement, informed by the more developed policies. If the initial estimates are from more complicated component model estimates, the unreported claim algorithm will produce a simplified process from which to derive similar estimates. In either case, the algorithm will condense the input information into something that is useful and will be easy to use going forward.

Additionally, in looking at the table above, one might ask why the opinion of ultimate doesn’t change over time as one moves from evaluation date to evaluation date. It is important to understand that all information is stated for historical points in time using the indications from the current reserve analysis. The table for these three policies might have looked like the one below for the reserve analysis done a quarter earlier. This shows that the estimate of ultimate for these three policies was $52,250, so there has been a slight improvement in the forecast given new information.

Policy	Effective Date	Evaluation Date	Days from Effective Date	Paid to Date	Case Reserves	Case Incurred	Hindsight Development	Total from Reported	Hindsight Unreported	Estimated Ultimate
1	1/23/2014	3/31/2014	67	0	0	0	0	0	34,900	34,900
1	1/23/2014	6/30/2014	158	0	0	0	0	0	34,900	34,900
1	1/23/2014	9/30/2014	250	0	1,500	1,500	7,500	9,000	25,900	34,900
1	1/23/2014	12/31/2014	342	0	1,500	1,500	7,500	9,000	25,900	34,900
1	1/23/2014	3/31/2015	432	9,000	0	9,000	0	9,000	25,900	34,900
1	1/23/2014	6/30/2015	523	9,000	0	9,000	0	9,000	25,900	34,900
1	1/23/2014	9/30/2015	615	9,000	18,500	27,500	4,250	31,750	3,150	34,900
2	8/14/2014	9/30/2014	47	0	10,000	10,000	-1,500	8,500	6,100	14,600
2	8/14/2014	12/31/2014	139	0	10,000	10,000	-1,500	8,500	6,100	14,600
2	8/14/2014	3/31/2015	229	8,500	0	8,500	0	8,500	6,100	14,600
2	8/14/2014	6/30/2015	320	8,500	3,000	11,500	2,000	13,500	1,100	14,600
2	8/14/2014	9/30/2015	412	9,250	2,750	12,000	1,500	13,500	1,100	14,600
3	3/1/2015	3/31/2015	30	0	0	0	0	0	2,750	2,750
3	3/1/2015	6/30/2015	121	0	0	0	0	0	2,750	2,750
3	3/1/2015	9/30/2015	213	0	0	0	0	0	2,750	2,750

In moving from the analysis done as of 9/30/2015 to the one done as of 12/31/2015, one can see that the opinion of ultimate improved a little bit from $52,250 to $52,000. The estimate of emergence decreased for all three policies as they aged. Offsetting this a little, the opinion of the ultimate value of the one open claim under Policy 1 deteriorated slightly.

It might also be noticed that instead of “Development Quarter,” the tables above use “Days from Effective Date.” For modeling purposes, the Days from Effective Date field would be used as a predictive variable and binned as appropriate. At any point in time individual policies will be anywhere along this curve, and interpolation can be used in conjunction with the predictive model. This becomes especially important in the early ages of a policy when emergence is still a large part of expected reserves and changes more quickly with the passage of time.

With the data arranged in this format, a predictive model can be built. The goal of this model is to predict the total amount unreported for a policy at any point in time. Predictive variables that could be included in the analysis include various rating variables or anything else that is associated with the policy, as has been discussed earlier in this monograph, as well as the policy age at the point of evaluation. The level of detail of the predictive variables could be finer than policy, as long as claims can be assigned directly to that level of detail. For example, a workers compensation policy may encompass many offices and industry classes. As long as claims can be assigned to this lower level of detail, the approach can be used at that lower level to get additional predictive variables. Sometimes however other items of interest, such as schedule and experience modifications may only be associated at the policy level and may need to be pushed down or allocated.

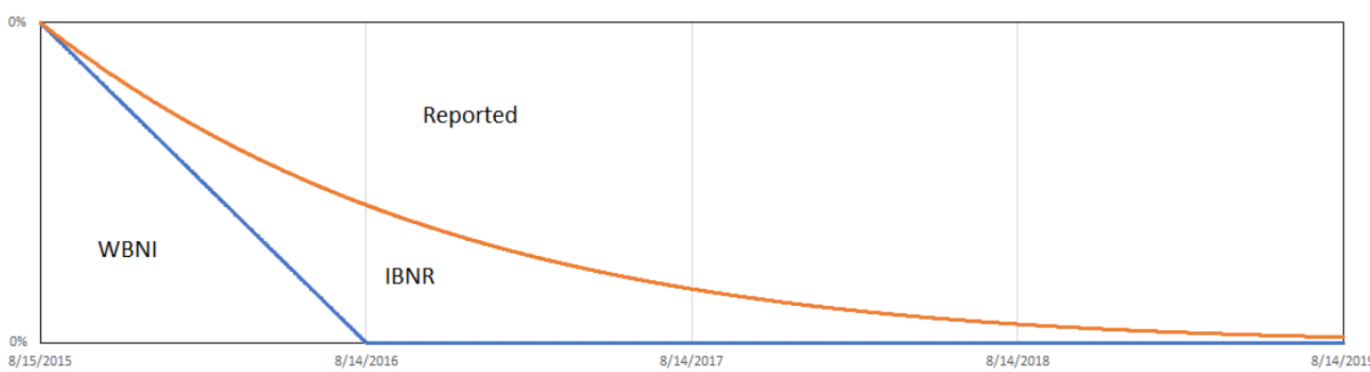
Since we are focused on only unreported losses in this algorithm, we take steps to avoid a potential problem with Bornhuetter-Ferguson of missing a changing mix of exposures that impacts the reporting pattern or expected loss ratio. Such mix changes are explicitly reflected in these algorithmic policy-level IBNR reserves.

Separate from mix changes, the Bornhuetter-Ferguson technique can lead to odd results when significant case savings are reflected in the development pattern by determining future savings by premium volume rather than by case balances themselves. For example, assume that for a given line, case-incurred losses are typically at 120% of ultimate at 24 months and that the a priori loss ratio is 60%. In this example. the Bornhuetter-Ferguson technique projects case savings of 12% of *earned premium* [0.6*(1.0-1.2)] at 24 months regardless of the *case reserves*. Even if there are no case reserves from which to *have* savings, 12% of premium is the projected savings amount. By using an actuarial case algorithm to project future payments on known claims and a separate policy-level IBNR algorithm to project as yet unreported claims, we avoid this problem.

It is noted above that policy characteristics are used as the predictive variables, but it is also possible that current paid and case (booked case reserves or actuarial estimates generated from the case algorithm) could also be used as predictive variables for the algorithm. For some lines or classes of business, it might be that current incurred information would have no impact on the prediction. For this type of business, emergence is a stable force which decays independently of other claims experience over time. In other cases, higher current incurred amounts for a policy to date could indicate a higher estimate of amounts yet to emerge. This is analogous to lines of business in which experience rating is a beneficial determinate of expected loss costs and hence premium. Higher current incurred information could also be a predictor of lower expected emergence in some cases. An example could be property policies in which property that is already damaged is not exposed to any or as much damage in the future.

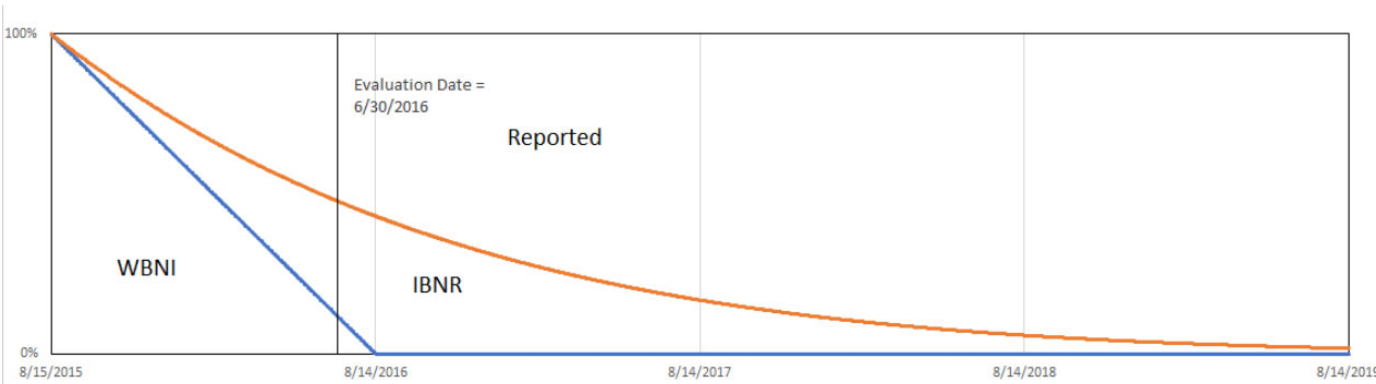
While it is not common for actuaries to think about estimating emergence at a policy level, this procedure, combined with the actuarial case algorithm discussed above, can be very powerful to help one examine a book of business. Since estimates of ultimate now exist at a very granular level (policy or even sub-policy), the profitability of any number of slices of the business can be evaluated. With actuarial reserving techniques that are currently in common use, triangles would have to be developed and factors selected along these slices of the business. With indications of profitability at a finer level of detail, the company is able to be much more proactive (defensive or opportunistic) in their underwriting and marketing decisions.

The following graph illustrates the expected emergence of ultimate losses for a policy over time.



The top of the graph (100%) represents the ultimate value of expected claims to emerge from this policy. This can be thought of as a reserve estimate or projection of ultimate loss for the policy at time zero (effective date of 8/15/2015). The blue line represents the value of WBNI claims (claims associated with unearned premium) at any point in time. Notice that this is a straight line from the effective date to the expiration date (this policy has a coverage term of one year and an assumed constant level of exposure over the year). The orange line represents the value of all unreported losses (IBNR and WBNI) at any point in time. Thus, the vertical difference between the orange and blue lines represents the value of IBNR claims at any point in time. After the expiration date, there is no WBNI, and all unreported claims are IBNR. Note that all claims are assumed to be at an estimated ultimate value when reported (via an actuarial case algorithm).

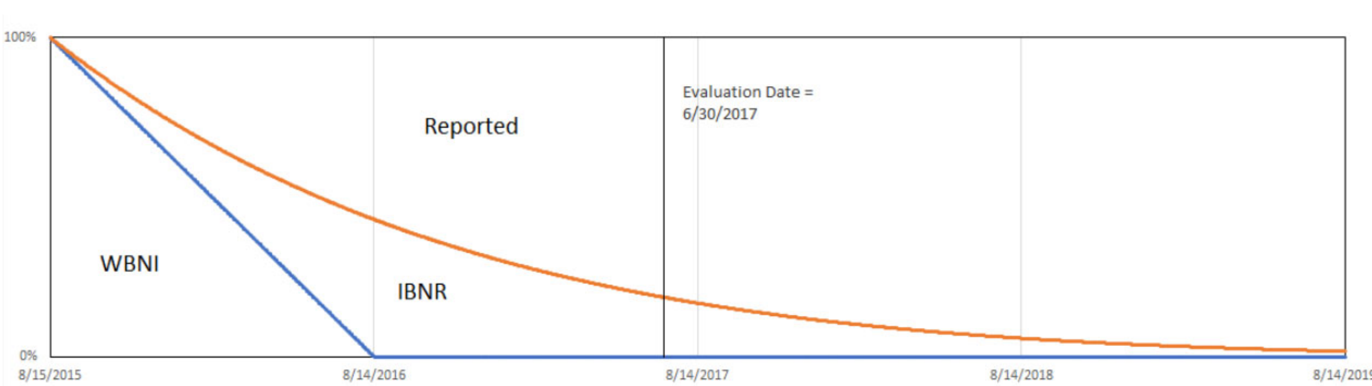
Now, consider the following graph of the policy at evaluation date of 6/30/2016.



At this point in time, 48% of ultimate loss is expected to emerge sometime in the future. The expected amount of WBNI is 12% of ultimate loss. Thus, the expected amount of IBNR at this point is 36% of ultimate loss.

A natural question at this point is to ask what is meant by *ultimate loss* in this context. It is important to note that it is not the projected ultimate for the policy as of the 6/30/2016 evaluation date. There is likely claims activity which has been greater or less than the expected amount to emerge from the policy inception date of 8/15/2015 to the evaluation date of 6/30/2016. The projected ultimate for the policy is the current incurred (using the actuarial estimates of case reserves generated from the case algorithm) plus the estimated value of unreported claims generated by this algorithm. This makes the algorithm behave very much like a Bornheutter-Ferguson model in which the unreported claim value algorithm determines the initial estimate of ultimate loss (or expected loss ratio) and the reporting lag pattern.

The following graph illustrates the expected emergence at a later evaluation date that occurs after the policy expiration date.



The estimate of WBNI at this point is zero (all premium has been earned), and the estimate of IBNR is 19% of the *ultimate loss* as described above.

The algorithmic policy level IBNR reserves at specific points in time can be added to the case algorithm adjusted accident period triangle to create a new triangle. The policy level IBNR and WBNI could be added to a policy year triangle that includes case algorithm amounts as well. Both of these triangles, accident year and policy year, can be used to test and illustrate the combination of the case reserve algorithm and the unreported claim value algorithm. Remaining development in both of these triangles should be minimal, illustrating the algorithm's effectiveness.

6.0 Known Claims: Component Development Models and Simulation

This section will focus on developing a more detailed modeling of the claim development process, by considering what happens within an individual time-step. Rather than focus on the ultimate value of claims, we will drill into the behavior of a claim from one period to the next. What is the likelihood that a claim will close in the next quarter? What is the likelihood that it will change in value? If it does change in value, how much? What is the probability of a payment? If the claim does have a payment, how much? By considering these components of development within a single time period, we can develop an understanding of the process that can ultimately be extended out to ultimate, but in the process take advantage of information on claims that are exhibiting behavior but are still immature.

To get started with this more detailed approach, we will first revisit the broad adjustments to current case reserves that were discussed in Section 4. Up until this point, we have focused on predicting all future payments of a claim directly as a function of claim characteristics as of a particular point of time. For claims that have already closed, this is an observable amount, but for claims that are still open, if we wish to use the information already observed about those open claims, we have to make some assumptions about how they will develop from the current point in time to ultimate before we can use them. The backwards recursive approach described in Section 4.2 allows us to make this assumption based on the single variable of the age of the claim. While this is a construct that is useful to get started with this type of analysis, it is overly simplistic. The very fact that we are differentiating across claims and across other variables than just the age of the claim suggests that we are likely introducing bias into the calculation of the case algorithm by using this assumption. The level of bias that is likely to be introduced is directly related to the proportion of the ultimate loss in the observed data that is still unpaid.

In Section 4, we discussed the use of the backwards-recursive algorithm to develop case reserves to an ultimate value for the individual open claims. It is illustrative to generalize on this concept here. An alternative formulation is presented below:

$$\widehat{C_t} = C_{t-1} \times \prod_{j=0}^{m} \gamma_j$$

$$\widehat{P_t} = C_{t-1} \times \prod_{j=0}^{n} \rho_j$$

where C_t denotes case reserve at a particular point in time t, P_t denotes incremental payment made in the period ending at time t, and γ_j and ρ_j denote factors for predictive variables. The first formula above represents a multiplicative model of m predictive factors, corresponding to m different predictive variables and a base factor γ_0, that estimates changes in case reserves during an evaluation period for a given claim. The second formula represents a model with n predictive factors (some predictive variables potentially the same as used for the change in case reserve, but not necessarily) and a base factor ρ_0 that estimates future payment amounts in the upcoming period for a given claim.

The basic backwards recursive algorithm in the Marker-Mohl technique is a special case of this broader framework, where there is only one predictive variable – the time since the claim was reported. The formulas above could be restated for this case as follows:

$$\widehat{C_t} = C_{t-1} \times \gamma_1 \times \gamma_0$$

$$\widehat{P_t} = C_{t-1} \times \rho_1 \times \rho_0$$

In this case, both γ_0 and ρ_0 both equal 1 and $\gamma_1 = R_{t-1}$ and $\rho_1 = P_{t-1}$, where the terms R_{t-1} and P_{t-1} come from the backward recursive formula described in Section 4.2.

This one-factor model is an aggregate model, based only on the development age, but more complex predictive models can be built in which m and n are greater than 1. In these cases, the predictive factors γ_j and ρ_j can be claim and/or policy characteristics which can be tested for predictive reliability.

Approaching a projection model with covariates included introduces specific modeling challenges when we extend the model to ultimate. The size of the case reserve itself typically will be a very important predictive variable, with small case reserves usually growing by a much larger factor than large case reserves. Once the claim level case reserve is considered as a predictive variable, the single period time-step model is not so easily combined together across development ages to provide a factor to develop the case reserve to ultimate. An individual claim may develop to a number of potential values in the next time-step. Using the mean predicted case reserve and using it as a predictive variable for the next time-step would be inappropriate. This is not an issue with more commonly used actuarial reserving methods because there is an implicit assumption of independence of the development factors and the paid and/or incurred amounts to which they are applied[5]. Using this assumption of independence at an individual claim level is extremely problematic and unrealistic. The following example illustrates this concept.

Consider a simple claim development process in which an open claim has three possibilities in the next time-step:

- the claim will close for nothing (with 1/3 probability)
- the claim will close, paying out the current case reserve (with 1/3 probability)
- the reserve will increase by 1 with no payment in the time-step

In this example, a claim currently open, with a case reserve of 1 has an expected value after one time-step also equal to 1 (0*1/3 + 1*1/3 + 2*1/3). If we take this expected value, treat it as a case reserve in the next time-step, and move it forward, it will also have an expected value of 1. Carried forward infinitely, the value is always 1.

But when we consider each of the possible paths that this open claim could take, we see that this approach is incorrect. The expected value one time-step out is indeed 1, but two time-steps out it is (0*4/9 + 1*1/3 + 2*1/9 + 3*1/9) = 8/9. After three time-steps the expected value is 22/27. As the number of time-steps approaches infinity, the expected value of the claim approaches 3/4. This

[5] This assumption is not always valid for aggregated data either and can cause problems for triangle analysis, but the assumption is not nearly as problematic as it is when considering claim-level development.

illustrates the problem of using the mean of a probabilistic model and using it in a subsequent model (either a later time-step or another component model).

In order to develop an estimate of ultimate loss from these time-step models with covariates, we need to describe not only the mean result in a time-step for a given claim with its given characteristics, but also the distribution of potential results in that time-step. With the introduction of various component models, as discussed below, this becomes even more important.

Possible approaches to projecting results for individual claims over multiple time-steps (and combining together the component models discussed below) include formulaic or numerical integration and stochastic simulation. With the level of complexity involved, and with flexibility of model choices regarding characterization of the distributions of the component models (and to a certain extent the component models themselves), this monograph will concentrate on the simulation approach.

Simulation practicalities suggest a need for modeling specific facets of claim development. When considering incremental payment amounts and changes in case reserves, there are large probability masses at zero (i.e. no incremental payment and/or no change in the case reserve). Instead of trying to incorporate these probability masses in the distribution of results, it is helpful to break the development process down into several components which are modeled at each time-step for each open claim within the simulation. Examples of these components are whether or not a claim closes, whether or not the value of a claim changes, whether or not there is an incremental payment, how much is the change in the value of a claim given that one happens, and how much is the payment given that one happens. The process is to model these behaviors individually and sequentially for each open claim at each time-step and simulate them accordingly.

Thus, the initial task on the path to building a comprehensive model of claim-level development is to build a number of time-step models of these facets, or components, of development. This approach seeks to understand the entire process of the claim's life cycle. Various aspects of the claim reporting, development, and payment processes are modeled. In addition to eliminating potential distortions from currently open claims, this can yield insight into changes that may be occurring within the process and assist with understanding the impact of potential changes in the process on the needed reserve. We observe how individual claims develop from month to month, quarter to quarter, or year to year, and then use that information to project the open claims to ultimate by chaining through multiple time-steps.

When first approaching a time-step model for individual claims, especially from a starting perspective of triangle development, it is natural to start simple with a single model such as:

Prediction: Value of the Claim at the End of the Time-step
Candidate Predictive Variables:

- Open/closed status
- Payments to date as of the beginning of the time-step
- Case reserves at the end of the time-step
- Age of the claim

Various claim characteristics
Various exposure characteristics

Note that this model has a number of dynamic variables as candidate predictors, i.e. values that will change over time. This is problematic when trying to project claims forward to an ultimate value, because we would also need information about these dynamic variables over time. Some changes are known in advance (the age of the claim will increment by 1 with each time-step), but others are not (the amount of the case reserve will change, the open/closed status at future dates is uncertain, etc.). The behavior of these dynamic variables must also themselves be modeled if we are to move from one time-step to another. The use of these multiple component models gives us additional information that is useful for understanding the claim process, but more importantly, is critical for being able to project not just for one time-step, but out to an ultimate valuation.

A specific model framework is provided here as an example. This is by no means the only such approach that could be taken to building a component time-step model. Other choices could be made, but by showing a specific example here, we illustrate how to overcome some particular challenges.

Claim Development Component Models:

- Closure Probability
- Change Probability
- Payment Probability
- Change Amount (Large)
- Change Amount (Small)
- Reopen Probability
- Reopen Amount

Before discussing each of these models individually, we first define some terms that will be used across the different models.

6.1 Definitions (for a given claim at a given time-step):

Beginning Case Reserve –case reserve at the beginning of the time-step

Ending Case Reserve – case reserve at the end of the time-step

Paid Loss – incremental paid loss amount within the time-step

Previous Paid to Date – total of all paid Loss for previous time-steps

Ending Value – Ending Case Reserve + Paid Loss. This represents an amount that can be compared to the Beginning Case Reserve to measure the change in value of the claim during the time-step.

“Loss” here is used generically to mean indemnity, expense, medical payment, or any combination of these.

<u>General Variables included in each model</u>
Beginning Case Reserve
Development Age
Transaction Date
Accident Period
Claim Characteristics
Exposure Characteristics

Development Age, Transaction Date, and Accident Period are redundant within two time dimensions. It is useful to consider each of them when constructing a component model because they each represent different things, but in a final version of any component model, there is likely to be no more than two of these. In section 4.3 we discussed the complications with regard to modeling the trend via the evaluation date. These same issues are considerations here when looking at transaction date. It can be a useful indicator and measurement of underlying trend, or in systematic changes that have occurred in the claim development process, but care will need to be taken to consider the prospective outlook, which cannot be measured directly. Accident period can also provide predictive value, but often this is an indicator of mix shifts in variables that have not been properly identified (variables that have not been included). Where possible, it is optimal to include such variables directly. Care should also be taken to avoid using accident period as a proxy for development age (since the more recent periods will contain only immature development ages). In such cases development age should be used in place of accident period to avoid projecting development for immature accident periods that is characteristic of early development as they progress into later development periods.

By claim characteristics we mean any claim specific variable that is captured, such as age of claimant, nature of injury, cause of loss, etc. By policy characteristic we mean any variable that is captured specific to a policy record to which a claim can be related. Note that this may be at a level of detail finer than at the policy level. For example, a workers compensation policy will likely have information at the location and payroll classification level. As long as the claims are also coded to location and payroll classification, the policy information at this finer level of detail can be attached to the claim.

6.2 Potential Models to be Employed in a Time-Step Model

The following are examples of component models which can be used in the development of a time-step modeling process. What is presented here is by no means the only way to structure a component based time-step development model, but it outlines an approach which could be used.

6.2.1 Closure Probability Model

This model estimates the probability that a given claim, open at the beginning of a time-step, will be closed at the end of the time-step. A claim is obviously much more likely to change in value if it is still open (reopened claims are considered here as a separate model). Therefore, being able to predict whether a claim will still be open in the future is important.

Definition: P(Ending Case Reserve = 0 | Beginning Case Reserve > 0 or Reopened = True)

Note that in this model we are defining a claim as being open or closed by looking at the case reserve. When the case reserve is larger than zero the claim is considered open for modeling perspective. The actual claim status the company recorded at the point in time could be used, but using the case reserve itself as the indicator avoids potential issues with inconsistent coding of claim status over time or timing discrepancies between status changes and case reserve changes. There are sometimes notices of claims that are given, particularly in claims made lines, that may be then considered open but have no case reserves. Consider using a notional case reserve of some small amount to identify such claims for actuarial modeling purposes if this approach to defining claim status is used.

6.2.2 Change Probability Model (for claims remaining open)

This model estimates the probability that a claim will change in value from the beginning of the time-step to the end of the time-step. This is equivalent to saying that the case reserve at the beginning of the time-step is different from the sum of any incremental payments in the time-step and the case reserve at the end of the time-step.

The reason for including a change probability model is that when considering the changes in value of open claims over time, there is a distinct possibility that in a given time-step there may be no change in value. When projecting losses forward, reflecting this probability mass at no change is more realistic than what would result from simply modeling the change in values broadly and including zero within the range of the potential change.

The probability of claim changing in value is typically very high for a claim that is in the process of closing. Including a variable that indicates whether or not the claim closes in the quarter would capture this, but it is likely that there would be numerous interaction effects between this variable and the others. For that reason, we have separated this model into one that considers only those claims that are remaining open vs. those that are closing.

Definition: P(Ending Value ≠Beginning Case Reserve | Beginning Case Reserve > 0 and Ending Case Reserve > 0)

Note that we are excluding claims that have zero case reserve at the beginning of the time-step. Changes in values of these claims is contemplated by the reopen probability and reopen amount models.

Also note that the very definition of this model depends on what is considered the result of one of the other component models (the closure probability model). This dependency will be important to consider when it comes time to project the losses forward.

6.2.3 Change Probability Model (for closing claims)

This models the probability that a claim which is closing will change in value.

Definition: P(Ending Value ≠Beginning Case Reserve | Beginning Case Reserve > 0 and Ending Case Reserve = 0)

Often this probability is very close to 1. For many cases quantifying this probability across all claims may be sufficient, with no additional differentiation provided by predictive variables.

6.2.4 Reopen Probability Model

This model is for the possibility that a given claim, closed at the beginning of the timestep, will have additional (positive) payments during or case reserve at the end of the time-step. As formulated here, this would also include payments on claims technically not reopened, but where additional payments occur after the case reserve goes to zero.

Definition: P(Ending Value > 0 | Beginning Case Reserve = 0)

The number of time periods that the claim has been closed is an important variable for this model, with additional payments often occurring in the period immediately following claim closure.

6.2.5 Reopen Amount Model

Given that a claim which was closed at the beginning of the time-step has additional (positive) payments or case reserves in the time-step, what is the amount?

Definition: E(Ending Value | Beginning Case Reserve = 0 and Ending Value > 0)

In this approach, the ending value, i.e. the ending case reserve amount plus the incremental payment, is used as the target variable of the model. The portion of this value that is paid out vs. that that remains as case reserves at the end of the time-step will be covered in the partial payment model.

6.2.6 Change Amount Model(s)

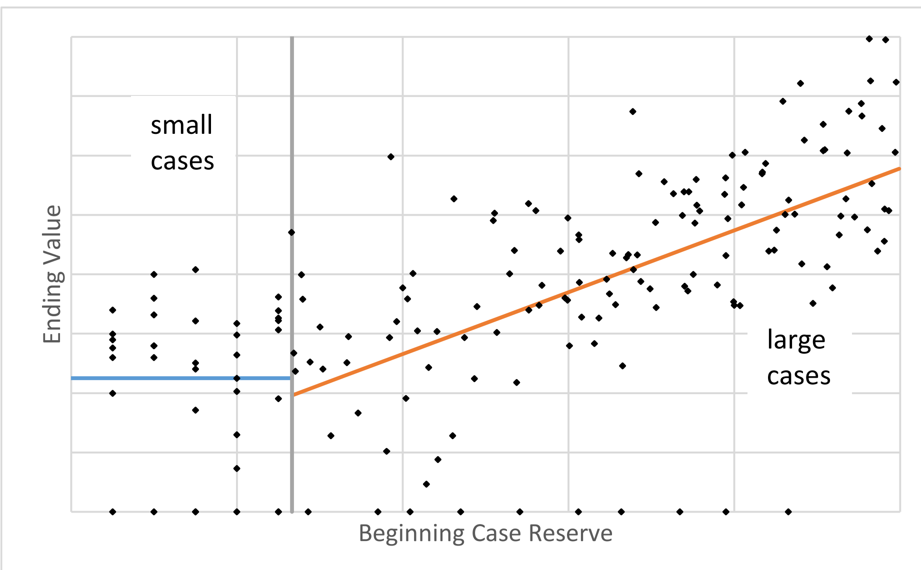
This model defines the changes in value of a claim that is open at the beginning of a time-step, given that it changes. By concentrating only on the cases where the claim changes in value, difficulty in modeling the distribution of ending values caused by the probability mass at "no change" is avoided.

The ending value is defined as being the ending case reserve plus any incremental payments, to be comparable with the beginning case reserve.

The ending value of the claim in the time-step is expected to be strongly related to the value at the beginning of the time-step (i.e. a strong positive correlation between the two), albeit with significant variation. This relationship is far from proportional, however, with small case reserve amounts growing by much larger factors on average than large case reserve amounts. For example, it is relatively easy for a $1000 reserve to grow by a factor of 20, but it is harder for a $1,000,000 reserve to grow by a factor of 20. With a multiplicative factor model framework, the case reserve amount itself will often become the most important predictive variable, and the model can become very sensitive to slight binning changes at the low end of the curve.

Another issue is that often for very small case reserve amounts, the monotone increasing relationship between the beginning case reserve and the ending value breaks down. Often this is due to the existence of "signal" reserves or other place holders that do not necessarily have a monotone relationship with the ending value. For example, a value of "1" for the beginning reserve may specify a particular type of claim, "2" indicate a different type of claim and there be no expectation that a "2" claim would be double the severity of a "1" claim or even that it would have a higher severity. For this reason, it is often beneficial to use separate models for small case reserve claims and for large case reserve claims.

It is helpful to transform the case reserve itself into a variable more closely related to the ending value, before even considering the impact of other predictive variables, which then serves as the exposure variable for the change amount model, replacing the case reserve. One such "change amount exposure" variable is shown graphically below:



Below the small case cutoff value, there is not an increasing modeled relationship between the beginning case reserve and the expected ending value. Above that value there is a linear, increasing relationship between beginning case reserve and the expected ending value. The inclusion of an intercept in the relationship for large cases provides the more significant multiplicative differential between smaller beginning case reserve amounts and the ending value. This avoids the problem of sensitivity to bin determination for the case reserve variable. The parameters defining this transformation are the two linear parameters above the cutoff (two for claims that are closing and two for claims that will remain open), the single parameter below the cutoff, and the cutoff itself (six total parameters). Given a particular value for the cutoff, the linear parameters can be calculated by least squares regression, and the parameter below can be calculated as an average of the ending value below the cutoff. The total least squares across the entire set of observations can be tabulated, and the optimum cutoff value can be determined using numerical minimization of the least squares amount. In some cases, the minimum least-squares amount will be at a cutoff of zero. It is appropriate to limit all of the other parameters to be non-negative as well, so boundary solutions should be considered.

Preliminary Calculation

Where Ending Value ≠ Beginning Case Reserve, define **Change Amount Exposure** =

$C_{small,0}$ where Beginning Case Reserve ≤ Small Case Cutoff

$C_{close,0} + C_{close,1}$ * Beginning Case Reserve where Beginning Case Reserve > Small Case Cutoff and Ending Case Reserve = 0

$C_{open,0} + C_{open,1}$ * Beginning Case Reserve where Beginning Case Reserve > Small Case Cutoff and Ending Case Reserve > 0

with $C_{small,0}$,$C_{close,0}$, $C_{close,1}$, $C_{open,0}$, $C_{open,1}$ and **Small Case Cutoff** estimated by minimizing least squares on the training data with Ending Value as the target variable.

Change Amount Model (small case reserve)

This model covers the cases where there is not a generally increasing relationship between beginning case reserves and ending value in the time-step.

Definition: E(Ending Value | Ending Value ≠ Beginning Case Reserve and Beginning Case Reserve > 0 and Beginning Case Reserve ≤ Small Case Cutoff)

The binned case reserve can be used as a categorical variable in this model (together with the other variables being considered). If there are specific signal reserve values, they can be set as distinct bins. A more sophisticated model would be to build models of changes of state from one claim type to another, within the different types of signal reserves, if it is common for claims to transition from one type to another before transitioning to an actual case reserve reflective of the expected loss payment amount.

Change Amount (large case reserve)

This model reflects changes in value for claims with beginning case reserve larger than the cutoff value.

Definition: E(Ending Value | Ending Value ≠ Beginning Case Reserve and Beginning Case Reserve > Small Case Cutoff)

Using the Change Amount Exposure variable discussed above, a multiplicative model can be used by setting the expected ending value of the claim equal to a base factor multiplied by the change amount exposure variable, multiplied by modifiers for each of the other variables being considered. The Change Amount Exposure variable may also be binned and treated as a categorical variable to capture possible imperfections in the crude linear relationship that was used as a first approximation or to reflect changes in the relationship that emerge when considering the other variables.

6.2.7 Payment Probability Model

This model describes the probability that there will be a payment on a claim that is either open at the beginning of the time-step or has “reopened” within the time-step.

Definition: P(Paid Loss > 0 | Ending Value > 0)

Notice that the way we have defined the "Ending Value" variable handles all the possibilities for payment since payment itself is included within the ending value. It may seem a little circuitous to construct the model in this way, but it is helpful to have a single model (the change value model) that governs both the incremental payments and ending case reserve generally, and then we consider a potential payment that is bounded between 0 and the ending value.

An important variable to include in this model is whether the claim closes in the period. As such, this model will be dependent on the closure probability model when projecting forward.

6.2.8 Partial Payment Amount Model

In the case where a payment does occur in the time-step and the claim closes, the payment amount is already determined by the ending value variable. In cases where there is a payment made but the claim is not closing (i.e., ending case reserve > 0) we need to know how much of the ending value is made up of payments and how much remains as case reserves. This partial payment model describes that relationship.

Definition: E(Paid Loss | Ending Case Reserve > 0 and Paid Loss > 0)

The Ending Value variable is important for this model, giving an upper bound for the payment amount.

6.2.9 Recovery Probability Model

What is the probability that a claim with previous payments will receive a recovery (i.e. negative payment) of some amount within a given time-step?

6.2.10 Recovery Amount Model

Given that there is a recovery in a time-step, what is the amount?
Note: Full payment in the time-step is implied automatically by whether a claim is closed and there is a payment.

6.2.11 Dynamic Variable Model(s)

Variables that change over time (dynamic variables) pose additional challenges, just as they do when used for segmentation of triangles in a traditional analysis.

Consider a "pension indicator" variable that is 1 to indicate that a workers compensation claimant is determined to be receiving permanent indemnity payments. That determination may change from 0 to 1 several development periods/time-steps after the initial determination is made. If segmentation of reserving triangles uses this indicator, then when a claim flips to a 1 either the history changes, or it is seen as an incremental increase in one of the triangles and a decrease in another, depending on whether only the current value of the indicator is used in the segmentation or if the historical values of the indicator are used for each diagonal separately. In the case where history changes within the triangle, care must be taken to estimate the impact of future changes from 0 to 1 within the current inventory of claims.

A similar issue exists in the claim life cycle model approach. Training a model using the current value of a dynamic variable for observations before it took its current value is not appropriate. Only the values of that variable that existed prior to the observation should be used. This is directly analogous to the triangle segmentation problem described above. If historical values for the variable are available, then it also becomes necessary to include simulated changes in the value of that variable in the simulation of future development, so that the new values can be used as inputs in the models that use that dynamic variable.

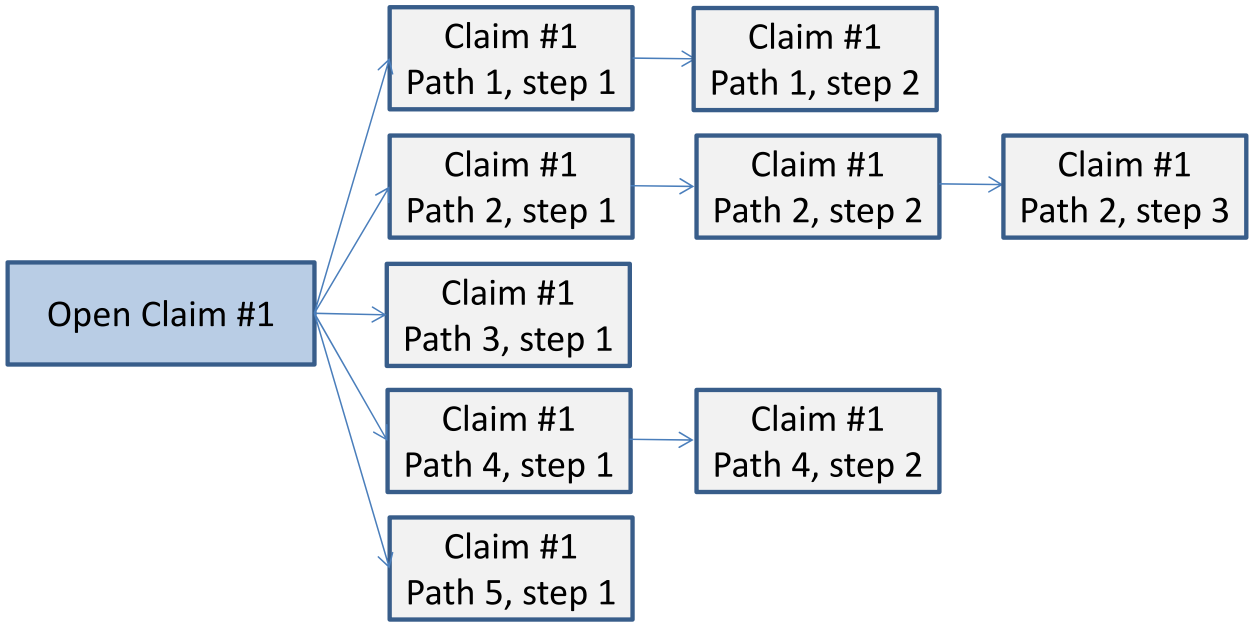
An example of one of these state change models is for a dynamic variable that takes values A, B, C, and D. Four separate predictive models could be created:

- What is the probability that the value would be A at the end of a time-step?
- What is the probability that the value would be B at the end of a time-step?
- What is the probability that the value would be C at the end of a time-step?
- What is the probability that the value would be D at the end of a time-step?

The value of the variable at the beginning of the time-step would obviously be an important variable, particularly given the potential for a variable to be unchanged. The sum of these four probabilities must equal unity, so adjustment of the predictions across the four models must be made to ensure this would be appropriate when applying the models. Each of the other predictive variables should be considered as possible predictors. Alternatively, these could be modeled with a series of conditional models (e.g., Probability of B given not A, Probability of C given not A and not B, etc.)

6.3 Claim Development Simulation

Each currently open claim is simulated forward one time-step using each of the Claim Development Predictive Models over a specified number of paths. Those paths still open are simulated forward another time-step. This process is continued until all claim paths are closed.



Additionally, claim re-openings are simulated and projected forward until they are re-closed, both for currently closed claims, as well as for currently open claims that eventually close.

Before time-step 1:

1. Generate a number of paths for each open claim (e.g. 1000)
2. Simulate claim re-opening from current inventory of closed claims, schedule them for reopening in later time-steps, and assign a path number
3. Simulate Ending Value for each of these reopened claims in the time-step in which reopened

In each time-step, for each claim-path combination:

1. Increment development age
2. Simulate changes in dynamic variables (other than the case reserve)
3. Select which of the claim-paths close
4. Select which of the claim-paths change in value
5. Select which of the claim-paths have a payment
6. Simulate Change Amount Exposure for each claim-path
7. Simulate Ending Value for claim-paths with Beginning Case Reserve > Small Case Cutoff
8. Simulate Ending Value for claim-paths with Beginning Case Reserve <= Small Case Cutoff
9. Set Ending Value = Beginning Case Reserve for claim-paths that do not change in value
10. Simulate Paid Loss on (0, Ending Value] for those claims having a payment
11. Set Ending Case Reserve = Ending Value – Paid Loss
12. Select which closed paths will reopen later, and schedule them
13. Simulate Ending Value for each of the claims to be reopened
14. Increment Development Age by 1
15. Repeat the process until Ending Case Reserve is zero for all claim-paths

Notes on Simulation

Random selections for binary models (such as Closure) are based on calculating the probability from the appropriate predictive model (using predictive characteristics) and simulating a Bernoulli.

The simulations for continuous variables (such as Ending Value, Paid Loss, etc.) are more challenging. Distributional forms can be used, but they are likely to be naïve with regard to distributional differences across variables, and this problem becomes more problematic due to the chained nature of the simulations across time. The simulated outputs from the first time-step are the inputs into the second time-step, the second into the third, and so on. Simplifying assumptions that may be reasonable in a single time-step may distort into unrealistic projections when compounded. The case reserve itself is typically one of the more important variables predictive of changes, with small reserves able to grow by a large factor, and large reserves unable to grow by large factors. Imposition of limits, actual or practical can also keep simulations from developing out of control.

One approach to reflecting nuances not necessarily reflected in a distributional form is to use bootstrapping techniques to simulate. In this way, differences in variability and higher moments across different categories of claims can be reflected. With many variables, it is unlikely that there will be sufficient numbers of observations of each combination of variables to represent the potential variability for any given risk, but by dissembling error terms across variables, randomly sorting, and then recombining them, a more nuanced reflection of variability can be achieved. By sampling residuals from the test data instead of training data, model and parameter risk are contemplated. The approach is highlighted in the following steps:

- Apply the predictive model to the records in the test data
- Calculate the residual for each test data record
- Allocate/disaggregate the residuals to each of the various predictive characteristics for each record (we will discuss this step in greater detail below)
- For a given claim-path-timestep to be simulated, randomly select one disaggregated residual for each predictive characteristic, from among the set of matching characteristics from the test data
- Combine the residuals for the claim-path-timestep to a single residual
- Combine the modeled residual and the expected result to give a simulated value
- Apply limits or other constraints
- Rescale the mean and variance across paths if necessary

The disaggregation of test data residuals across predictive characteristics is what allows this approach to generate variability patterns that are similar to what has been observed for similar claims in the past, while still allowing for combinations of characteristics that have not been observed. With skewed, positive distributions, it is helpful to use multiplicative residuals rather than additive residuals.

The example below illustrates the concept of this type of bootstrapping with two predictive characteristics, State and Class. The concept is generalizable for more variables.

Test Record	Prediction (A)	Actual (B)	Residual (C) = A/B	D = ln(C)	State	Class
1	96	40	0.4167	-0.8755	X	A
2	132	200	1.5152	0.4155	X	B
3	126	37	0.2937	-1.2254	X	C
4	96	90	0.9375	-0.0645	X	A
5	88	100	1.1364	0.1278	Y	B
6	84	110	1.3095	0.2697	Y	C
7	64	73	1.1406	0.1316	Y	A
8	88	120	1.3636	0.3102	Y	B

Test Record	State Factor (E)	Class Factor (F)	G = ln(E)	H = ln(F)	I = G/(G + H)	J = H/(G + H)	Disaggregated residuals State D * sqrt(I)	Class D* sqrt(J)
1	1.20	0.80	0.182	0.223	0.450	0.332	-0.587	-0.504
2	1.20	1.10	0.182	0.095	0.657	0.127	0.337	0.148
3	1.20	1.05	0.182	0.049	0.789	0.058	-1.088	-0.296
4	1.20	0.80	0.182	0.223	0.450	0.332	-0.043	-0.037
5	0.80	1.10	0.223	0.095	0.701	0.120	0.107	0.044
6	0.80	1.05	0.223	0.049	0.821	0.056	0.244	0.064
7	0.80	0.80	0.223	0.223	0.500	0.309	0.093	0.073
8	0.80	1.10	0.223	0.095	0.701	0.120	0.260	0.107

The square roots in the last two columns of the above table are in recognition of the reshuffling of residuals across characteristics that will occur in the simulation. If columns I and J were used directly without the square root, resulting variability would be too low, because the correlation between the disaggregated residuals at the record that exists with the observed residuals is eliminated when the simulation draws are performed independently across the characteristics.

Note that in the above approach, more of each observed residual is assigned to the variable with the stronger predictive effect. The predictive factors in this example are normalized to 1, so a factor close to 1 is an indicator that the characteristic value does not describe much of the difference in the target of prediction. The embedded assumption in assigning the residuals in this way is that the stronger the factor (further from 1) the more of the residual is assigned to that variable.

The table below shows a couple of simulated results using the approach.

Simulated Result	State	Class	Residual Draw - State	Residual Draw - Class	Combined Residual Factor	Predicted Result	Simulated Result
1	X	C	-0.043	0.064	1.021	126	128.62
2	Y	A	0.244	-0.504	0.771	96	74.03

6.4 Time-step models vs. case-algorithm model and other practical considerations

Both the time-step component model approach, with simulation, and the case algorithm approach provide estimates of the value of an open claim given its characteristics, but it is worth highlighting the differences.

Benefits of using the case algorithm approach are that it is considerably simpler, easier to audit and understand in its application, and is fast to apply to previous valuations of open claims for purposes of adjusting a triangle. On the negative side, potential biases could still exist from the use of a broadly determined case adjustment factor, particularly if the mix of claims or exposures has changed recently.

One benefit of using the time-step component models and simulating individual claims to ultimate are that additional insight into the claim life cycle is gained. Differences in settlement rates of claims across predictive variables, volatility differences, etc. are revealed. Changes in claim management practices over time are more easily characterized and identified by the actuary. Another benefit is that through simulation, an unbiased estimate of currently open claims is provided with far more specificity than in the simple un-biasing procedure such as the backwards recursive approach, reflecting the measured bias differentials from different types of claims. On the negative side, the simulation process to bring all the component models together and project over multiple time-steps can be challenging to audit due to its inherent complexity, and can be time-consuming, especially if all previous evaluation points for claims need to be projected.

One way to get the benefits of both approaches is to combine them together into a single framework using an approach such as the following:

1. Develop an understanding of the claim development process through building time-step component models, developing the currently open claims to an estimated ultimate value.
2. Simulate from the current state to ultimate, using the time-step models that are developed.
3. Use the simulated results as the starting point for building the case algorithm model, given the projected ultimate values for each known claim. In many cases this may simply be the final value of a closed claim, but in the case of open claims it is likely to be different depending on the characteristics of the claim.
4. Apply the case algorithm model to historical valuations of open claims.
5. Summarize the adjustments, by period and age, and adjust triangles accordingly.

The consolidation of the results from the more comprehensive investigation into a case reserve algorithm provides a simplified approach to applying the knowledge gained by the individual claim development modeling process and the ability to apply easily to future open claims. The resulting algorithm can be tested in report period triangles to illustrate its effectiveness.

Practical considerations should be made with regard to using the modeling approaches outlined in this monograph. This is particularly obvious in the case where the analysis is supporting a Statement of Actuarial Opinion (SAO) and would likely be audited, but it is also the case more generally. It is useful to be able to consolidate the analysis into a fairly straightforward adjustment to the generally used triangle inputs. Subjecting the complete environment of component development models, simulations, etc. to an audit would likely add great cost and time to the process, when the resulting analysis can actually be supported with considerably less effort. Using a case algorithm to summarize the learnings gained from a more detailed review, provides a straightforward path to showing its efficacy. Regardless of the derivation of the specific parameters in the algorithm, if the underlying variables are reliably determined and available contemporaneously with the case reserve, and the resulting case reserves result in stable development patterns of case-incurred loss, support for the algorithm's use is provided.

In some cases, it may be that the analysis leads to very similar conclusions to those indicated by simply applying standard triangle techniques without adjustment to the basic triangle data. In those cases, it may be appropriate to simply state that additional work was done to stress test the triangle development approach, providing summary information about the detailed analysis, and that the conclusions did not indicate that adjustment to the underlying data or techniques were required.

In cases where the actuary does feel that the detailed analysis provided truly valuable additional information, resulting in a significantly different reserve indication, providing adjustments to the triangle data on a claim-by-claim basis provides an auditable path for comparing to the source data. The objective nature of the fields used in the case algorithm itself can be verified, and the mathematical application of the algorithm to the individual claims can be verified as well. Without going into details about how the algorithm was developed, its efficacy can be verified by the resulting incurred development itself (or more to the point, the lack of development). In most situations the objective case algorithm will be considerably more auditable than the subjective case reserves they are replacing and also show considerably less aggregate development.

Even in situations where an SAO is not being supported, the simplicity and verifiability of using a case algorithm is beneficial in that it allows others to be able to get comfortable more easily with the analysis, and it provides a very straightforward way to update future analyses.

7.0 Unknown Claims - Component Emergence Models and Simulation

To this point we have focused on the development of known claims. While it is useful to have a model of claim development for understanding differences in the development potential of current open claims, it is an incomplete model, because it does not include an estimate of the future cost associated with unreported claims.

Not having been reported, the potential for these claims is driven not by claim characteristics (which do not yet exist) but only by exposure characteristics. The nature of the predictive models necessary to describe the unreported claim potential will depend on the nature of the simulations to be performed. If it is desired to simulate claims with specific claim characteristics and then to simulate development for each of these claims based on those characteristics, then detailed predictive models to predict all of these detailed characteristics would have to be built. It is considerably easier to simulate the ultimate values of the IBNR claims, in which case the claim characteristics are unnecessary. In addition to losing the detailed information about the nature of the claims, which may be useful, the timing of payments is lost. The simplification is dramatic, though, so this is a very important consideration.

7.1 Component Emergence Models

Describing the simpler approach, there are three basic component models to predicting the unreported claims – a report lag model, a frequency model, and a severity model.

7.1.1 Report Lag Model

We define the report lag as the time difference between the incurred date and the date reported. If in the case development models, we defined the date reported as being the first date a case reserve was established or a payment was made, we should use the same definition here.

One significant problem with modeling differences in the reporting lag across different policy characteristics is that, like the claim development models, we have incomplete data that we would like to incorporate. Using a triangle analogy, the relationship between the reported claim count at age 2 may be a different factor to the age 1 claim count for one type of policy than for another. This says something about the report lag differences, even though the information may be from an incomplete incurred period. One way of incorporating the immature incurred periods into the analysis is through the use of conditional reporting patterns. For example, given that an incurred quarter is aged two quarters, we can hypothesize a distribution of time reported within those two quarters, given that a claim is reported within two quarters. Any claim then that is reported within those two quarters can be identified as being at a particular point of the conditional CDF[6] of that reporting time distribution, i.e. a value on [0,1]. Claims from each incurred period can similarly be scored to a CDF value on report lag distributions conditional on reporting date less than the age of each incurred period. Differences in these CDF values across policy characteristic can be modeled across all periods at the same time, with

[6] Unlike in Section 4.2.1, the meaning of CDF here is "cumulative distribution function."

higher CDF values corresponding to a lengthier report lag and lower CDF values corresponding to a shorter report lag.

As an example, assume that in general report lag distribution is as follows, based on performing chain-ladder calculations on a reported count triangle:

Qtr Age	Reported
1	40%
2	50%
3	58%
4	65%
5	71%
6	76%
7	81%
8	87%
9	91%
10	95%
11	98%
12	100%

Suppose we have a claim that is in an incurred quarter that is now 6 quarters old, and that the report date was 100 days after the loss date. The conditional CDF for the claim is
Conditional CDF = CDF(100 days)/0.76.

CDF(90 days)=0.4 and CDF(120 days)=0.5. The skewness of the overall distribution usually results in reporting more claims early in a period than late in the same period, so straight interpolation would be somewhat crude. Instead, if we assume that the development within a given period is represented by a segment of an exponential distribution, we can calculate parameters for each section:

Qtr Age	Reported	Lambda
1	40%	0.5108
2	50%	0.1823
3	58%	0.1744
4	65%	0.1823
5	71%	0.1881
6	76%	0.1892
7	81%	0.2336
8	87%	0.3795
9	91%	0.3677
10	95%	0.5878
11	98%	0.9163
12	99%	0.6931

Where lambda is calculated as $\ln[(1-CDF_t)/(1-CDF_{t-1})]$

CDF(100 days) is then 0.4+ [1-exp(-0.1823*10/90)]*(1-0.4) = 0.412. Our conditional CDF for the claim described above is then 0.412/0.76 = 0.542.

The approach then for modeling differences in reporting across policy characteristics is to assign each reported claim a report date CDF, conditional on the age of the claim as shown above, and then use this CDF as the target variable of prediction.

7.1.2 Frequency Model

Observed claim frequency is dependent on the maturity of the policy, due to the lag between incurred date and report date. Therefore, reporting lag should be modeled first, and the measures of claim frequency can be relative to a premium that has been adjusted to reflect the maturity of the policy due to this reporting lag as well as for any unearned portion of the policy.

In addition to adjusting for earning and report maturity, adjusting the premium for rate level changes may be appropriate as was discussed in Section 3.4. Alternatively, including a predictive variable that represents the relationship between a constant rate level premium and the collected premium can be useful.

The effective date of the policy can be included as a predictive variable to measure frequency trend. In the case where premium is not adjusted to a constant rate level, the effective date variable may reflect differences in rate.

7.1.3 Severity Model

For modeling claim severity differences across policy characteristics, we are interested in *ultimate* claim severity. Case-incurred losses include whatever distortions exist in the case reserves. Often a solution to this problem is to use *closed claims only* when building a severity model. Unfortunately, this tends to introduce bias due to differences between open and closed claims. For example, in liability claims, it is likely that the less severe claims are settling faster than the more sever claims. In this case, a closed-claim severity model, if not carefully constructed, will tend to result in severity estimates that are too low. A closed-claim property model might be biased high, with faster settlement of accidents that resulted in a total loss. One approach to removing open/closed bias could be to include closed claims only from periods that are essentially fully developed, but this will likely exclude valuable information about more recent claims. When there is a trend present or where the underlying claim severity environment is otherwise changing, this loss of recent information may be very problematic. Instead, by first developing the known claims to an ultimate level, before modeling differences in claim severity, we can include the open claims – eliminating the open/closed bias, but hopefully without the potential distortions caused by traditional claim-department case reserves developing differently across different types of claims.

Using the mean projection of ultimate payments for a claim on those claims that are currently open will tend to underrepresent the variability of the ultimate claim value. This can be problematic when being used in predictive models, particularly when testing alternative models against each other, but also simply for characterization of the variability of severity generally. For this reason, it may be useful to select a particular simulated path for each open claim, rather than the projected mean across paths.

The relationship between report lag and loss severity is typically very strong, and generally the difference between the incurred date and the reported date should be included as a potential predictor.

As with claim frequency, the policy effective date could be considered a potential predictor, representing severity trend. Alternatively, the incurred date could be used.

7.2 Unknown Claim Simulation

As mentioned in the Section 5, the simulation of unknown claims may involve simulating individual claim characteristics and resulting claim development behavior or may simply simulate their ultimate values. The reason why this option exists for unknown claims, while the Case Development simulation was performed in time-steps, is that when we have open claims at a varying level of maturity, we need a bridge from each point in time to ultimate. For IBNR claims the only maturity component that is different across the claims is the report lag. With the ultimate severity of claims now defined by the claim development modeling and simulation (as discussed in Section 6), we can simplify the unknown claim simulation into Frequency, Severity and Report Lag. As discussed in Section 7.1 we will focus in this section on the simulation of unknown claims under this simplified model.

The simulation process for the unknown claims is as follows:

1. Each policy that is seen as still having potential for claim reporting is assigned a policy maturity factor based on its modeled report lag and the portion of the policy period that has been earned. Expected unknown claims are calculated as the premium, multiplied by unity minus this maturity factor, and then applying the frequency model to this amount based on account characteristics.
2. Individual claims occurrences are then simulated for each policy with a mean equal to this number of expected unknown claims. Paths are assigned randomly.
3. Date of Loss and Report lag are simulated for each of the emergence claims according to the report lag model.
4. Ultimate severity is simulated for each of the emergence claims according to the severity model.

Similar to a case reserve algorithm being a useful simplification of the claim development model and simulation approach, use of an unknown claims algorithm to summarize and simplify the component driven unknown claims simulation approach is useful for continued downstream applications.

8.0 Other Modeling Considerations

In actuarial reserve analysis it is common to break out specific types of payments into separate analysis, such as to separate indemnity payments from expense payments and/or medical payments. In addition to reporting requirements that may require separate estimates, the types of payments typically develop differently, and it often is beneficial to analyze them separately in order to be able to provide additional insight (Friedland, 2010, Ch3).

When building an analysis based on detailed data, this is still the case. In addition to illustrating different development, the application of limits can be reflected with greater sophistication if simulation of the future payments is being performed. One of the strategies for using the payment type is simply to identify the payments with a distinct claim ID (i.e. treat the expense or medical as being a separate claim from the indemnity) with the payment type as just another variable. Scanning for possible interaction effects may indicate whether a complete separate analysis is warranted. For example, in a workers compensation analysis that treats indemnity payments as separate claims from medical, it is likely that several of the variables are likely to have interaction effects with the payment type. The more of these interaction effects there are, the more straightforward it is to simply break out the payments into separate, distinct analyses, rather than deal with multiple interaction variables.

One of the additional benefits of considering these detailed payment types is in their ability to be included as predictive variables themselves. As mentioned earlier, payments to date and recent payments can both be very predictive of future payments for an open claim. Staying with the workers compensation example, it is common that the amount of indemnity payments to-date on a claim is predictive of the medical payments and vice versa. In fact, drilling down even further to provide additional details about recent claim payments can add significant information for predicting the future behavior of open claims. Contrast this with the common technique in reserving for workers compensation of performing separate analyses for Medical Only claims and Medical with Indemnity Payment claims. Instead, a single Medical claims model may emerge with the amount of indemnity payment to date as a variable of interest. Zero indemnity payment is an important value, but it may be that it is not very different from indemnity payment of less than $1000. Also, while the status of "no indemnity payment" is fairly stable for a claim, it does have the potential to change, and that can create issues for a triangle. Reflecting that changing status as a paid indemnity variable in a medical claim actuarial case reserving algorithm avoids that problem.

When adding such payment-type fields to a component development model, as mentioned earlier, it is incumbent to include the future simulated payments of various types as inputs into the simulation of the other payment types, adding significant complexity to the simulation process. Including such cross payment-type relationships into the case algorithm is considerably more straightforward since it only requires payment types to be captured at each point in time in the history.

9.0 Explanatory Models

While stochastic simulation is very useful to project reserve development from a complex combination of predictive models over a number of time periods, it can also make it difficult to attribute the specific differences in development across claims and exposures to specific fields. For example, suppose the claim closure rate is significantly slower for a particular geographic area. How much impact did the difference in closure rate make for the total estimated change from the current case reserve to ultimate for a claim in that area, given that there are more opportunities for changes in development the longer a claim remains open? For this reason, it is helpful to create explanatory predictive models at the end of the simulation process. Using the result of the simulation process as the target of prediction, and all the variables that were found to be of importance in any of the component predictive models, a simplified explanation of which variables are the most important with regard to overall impact is provided, as is the nature of the impact of those variables.

In addition to being a simplified explanation of which variables drove the simulation results, it also can be used at future points in time to mimic what the simulation process *would* generate for claims open at *those* points in time. It may not be necessary to go through the complete simulation process each time there is a refresh to the open claim list, if instead the explanatory model can do a good job of mimicking the results. This can be useful for ongoing review of the open claim inventory for purposes of claim management.

This post-simulation explanatory model (development prediction) is different from the actuarial case algorithm discussed in previous sections. Like the actuarial case algorithm, the prediction is of future payments for an open claim, as of a particular point in time, but there are a number of differences:

	Actuarial Case Reserve Algorithm	Explanatory Analysis (Development Prediction)
Appropriateness for use in actuarial pricing and reserving models	Higher	Lower
Appropriateness for use when managing specific claims	Lower	Higher
Uses Claim Dept Case Reserve as a predictor	False	True
Closed Claims Included when parameterizing	True	False

The actuarial case algorithm is designed for use in actuarial work of reserving and pricing. Individual claim results are less important in this context than aggregated results and changing practices regarding case reserve levels are problematic.

The explanatory analysis (development prediction) is focused on individual claim level predictions and is focused on the current body of open claims and the potential of those claims to develop adversely. There is less concern about shifts in general case reserve adequacy at this level.

While including the claim department case reserve as a predictor defeats the purpose of an actuarial case reserve due its subjectivity, that same subjective information is useful when considering the potential future development of an *individual* claim. Therefore, including the claim department case reserve in the explanatory analysis as a predictor is appropriate.

Because the explanatory model is focused on the current portfolio of open claims and how the simulation process is projecting them forward, only those claims that are being simulated are included in the parameterization. For the actuarial case reserve algorithm, all claims are considered valuable for parameterization, with cradle-to-grave information included in the portfolio of closed claims, and a combination of historical information and simulation included in the open claims.

10.0 Case Reserves - Actuarial vs. Claim Department

As the actuaries develop alternative case reserving models, it is tempting to suggest that such models should replace existing case reserves in other contexts, such as by use in the claim department. This temptation should be considered carefully.

The case reserves that would be ideal for the actuarial department are likely different from the case reserves that would be ideal for the claim department. The actuaries are best served in reserving and pricing by case reserves that are unbiased for cohorts of claims and policies along significant rating and other characteristics. Referring to mean-valued expectation when *settling* claims is likely *not* optimal. The mean value includes the often-significant impact from the small probability that the claim will blow up. If the mean is then used as a reference point (compared to a case reserve that ignores the skewed distribution such as the median expectation), it may lead to higher settlements generally, and an increase to the mean payment itself. The median outcome, with its easily understood evaluation of "just as likely to develop up as to develop down" may be a more appropriate value for claim department use than the mean outcome. At the claim level, the median outcome and the mean outcome can be significantly different.

The claim department also has information about the claims that is likely not coded in a way that is available to the actuaries, including significant subjective information. This subjective information is extremely valuable at the individual claim level with regard to settlement, while being problematic for the actuary due to the inability to access it and due to its potential to change over time. Therefore, this information should be included in the claim department's case reserve estimate but generally not in the actuary's case reserve estimate.

This bifurcation into two separate estimates is good from the perspective of allowing the claim department to operate more freely to make appropriate claim settlement decisions without concern that the actuary's triangles will be impacted in unexpected ways. A common request from the actuary to the claim department is to make no changes. This is unrealistic and suboptimal, as there are often good practical and economic reasons for making changes. Freeing up the claim department to make changes to case reserving that lead to better decision-making down the road is a distinct economic benefit. This also extends to the speed of claim settlement, which also historically has had the potential to distort the actuary's triangles. If the actuary instead uses an algorithmic case reserve, under her/his control that is developed with the goal of being unbiased, a speed-up or slow-down in the settlement of claims, will impact case-incurred development only if it translates into a change in the amount of ultimate payment for the claim. If it does not, payments are exactly offset by a drop in the algorithmic case reserve and development is the same as it would have been if the speed-up or slow-down had not occurred.

11.0 Use in Pricing and Internal Management Reporting

In building a detailed model of reserving that includes IBNR, a pricing model is a natural by-product. The Frequency and Severity models described in Section 7.1 together form a model of pricing. The IBNR Algorithm of Section 5.0 at time zero is also a pricing model.

It is important in actuarial pricing to consider the extent to which observed losses used for the pricing analysis need to be developed to an ultimate level, and how to develop them. Often very broad-brush approaches are used to address this question. Examples include:

- Using reserves that have been allocated to the level of detail being priced
- Applying development factors or premium development to policy level loss or premium
- Comparing differentials in case-incurred loss across different types of policies as an estimate of differentials in ultimate loss

Each of these approaches are fine *if and only if* differences in loss development across each of the pricing variables are being properly reflected. Too often they are not. While most actuaries will recognize that loss development differences exist between different deductibles and that loss experience should be adjusted accordingly when using observed results to price deductible credits, they may not always consider that different industry classes, or geography, or policy form, or any other variable of interest may have differing loss development exhibited across its values. These differences are significant and indicated profitability and pricing indications can be very different between making blanket assumptions and properly reflecting difference in potential for additional loss development. Every pricing variable should be checked for potential distortions from loss development whenever immature data is being used to develop indications.

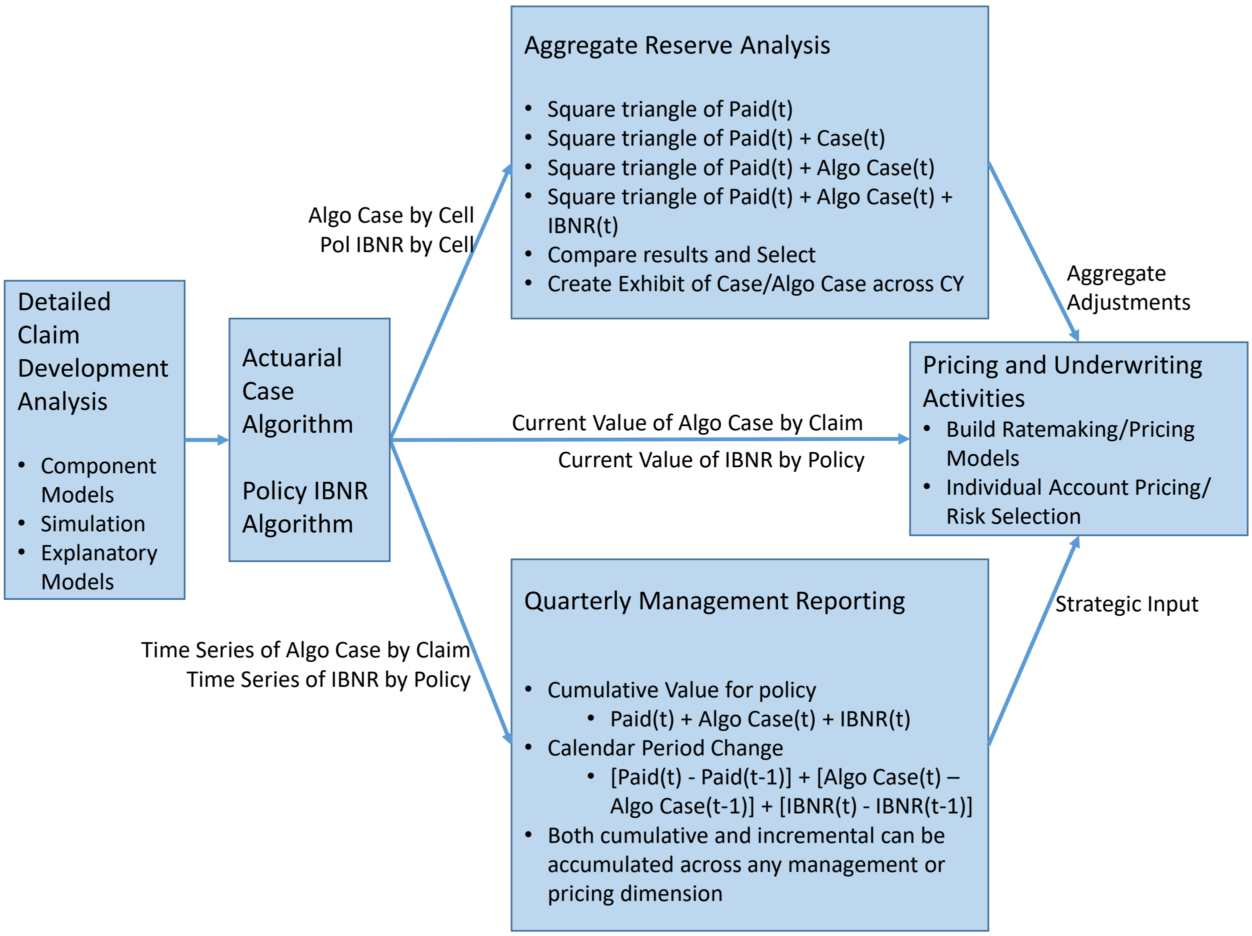
Sophisticated predictive modeling techniques are being used now by actuaries to price insurance, but when the true variable of interest is ultimate loss and the item being measured is case-incurred loss, there are likely to be significant distortions.

Possible choices for incorporating loss development differences in pricing include:

- Use Frequency and Severity models from a claim life cycle model directly.
- Use IBNR(0) directly from an Unreported Claim Algorithm
- Use the results of case algorithm and IBNR algorithm as a starting point for independent pricing predictive models, in conjunction with payments to date.

In addition to providing more appropriate input for actuarial pricing calculations, using reserve estimates calculated at the claim and exposure level are valuable for internal management reporting. Rather than relying on crude allocations of reserve estimates to the various levels that may be reported (office, region, business unit, agency, etc.), the sum of claim and policy level reserves can be easily provided. If there is a final adjustment needed to be allocated to be in line with a reserve total that differs from the summed detail reserves, using the detailed reserves as the allocation basis will provide a much more thoughtful result. This has great potential for providing much more timely and reliable information about the results of underwriting efforts than that provided by crude allocation.

The graphic below illustrates the way in which the detailed claim development analysis can impact actuarial, underwriting, and strategic decision-making.



12.0 References

Berquist, James R., and Richard E. Sherman, “Loss Reserve Adequacy Testing: A Comprehensive, Systematic Approach,” Proceedings of the Casualty Actuarial Society 64, pp. 123–184, 1977.

Bodoff, Neil, “Measuring Rate Change,” CAS RATEMAKING CALL PAPERS, 2009

Bornhuetter, Ronald L., and Ronald E. Ferguson, “The Actuary and IBNR,” Proceedings of the Casualty Actuarial Society 59, pp. 181–195, 1972.

Friedland, J.F., “Estimating Unpaid Claims Using Basic Techniques,” Casualty Actuarial Society, Third Version, July 2010.

Marker, J.O., and J.J. Mohl, “Rating Claims-Made Insurance Policies,” Casualty Actuarial Society Discussion Paper Program on Pricing Property and Casualty Insurance Products, pp. 265-304, 1980.

Robbin, Ira, “Pitfalls of Predictive Modeling,” CAS E-Forum, Spring 2016

Zehnwirth, Ben, “Probabilistic Development Factor Models with Applications to Loss Reserve Variability, Prediction Intervals and Risk Based Capital,” Casualty Actuarial Society Forum (Spring), 2, pp. 447–606, 1994.