On the Accuracy of Loss Reserving Methodology

Tapio Boles and Andy Staudt

Abstract

We evaluate the performance of various loss reserving methods and their associated parameterizations under a number of environments (e.g., changes in case reserve adequacy). We simulate proxy loss development data for each environment, which enables us to measure the accuracy of various actuarial projection methods. Then, based on our results, we offer a roadmap the reserving actuary may use in order to select appropriate methodologies and parameterizations given the current, past, and expected future environmental conditions affecting the reserving process.

Keywords: suitability testing; loss reserving; reserving methods; loss development; management best estimate; simulation.

1. INTRODUCTION

We evaluate the accuracy of various loss reserving methods and their associated parameterizations for several lines of business under a variety of common environmental conditions.[1] Based on our results, we offer a roadmap to guide the actuary in evaluating the appropriateness of these methodologies under different circumstances, understanding the differences in projections between various methods and supporting the choices an actuary makes given the past, present, and expected future conditions.

1.1 Use of Simulated Data

Most similar research falls into one of three categories: hindsight testing,[2] mathematical proof, or simulation.[3] In theory, all are viable options, but limitations in the former two make them unsuitable for our current purposes. Essentially, we are interested in how various loss reserving methods will perform given our reasonable expectations as to the future in a real-world setting. With hindsight testing, we are only able to evaluate the performance of methods under one set of environmental conditions—namely the past; and even then, we are only able to make these evaluations many years after the fact. Furthermore, as the future presents entirely new, unknown environmental conditions that reasonably can be expected to differ from the past, we cannot extrapolate results from hindsight

---

[1] By environmental conditions we refer to characteristics such as inflation, changes in case reserve adequacy, changes in rate adequacy, changes in claim settlement practices, and changes in the mix-of-business.

[2] Hindsight testing is the process by which past predictions are compared with current results; this occurs some years after the predictions are made and gauges the effectiveness of the methods (and/or actuary) that produced those predictions. See Mahon [12] and Jing, Lebens, and Lowe [10].

[3] Stanard [22], Pentikäinen and Rantala [17], Rollins [19] and Narayan and Warthen [15].

history. And most mathematical proofs, while elegant, can be complex and difficult to apply to a diverse array of methods and environments.

For these reasons, we chose to simulate proxy loss development data. We simulate the historical triangle (what the actuary sees) as well as future periods, thus enabling us to evaluate the accuracy of various methods at ultimate. We are also able to isolate environmental conditions, in order to determine how the accuracy of methods is affected by various environmental changes. Finally, by adding noise to the simulated data, we are able to evaluate how susceptible the accuracy of each method is to random volatility.

1.2 Outline

Section 2 describes the various aspects of our approach. Section 3 discusses possible biases in our approach. Section 4 presents the application of our approach to specific examples.

We have included several appendices to help the reader understand the specifics underlying many of the concepts. In Appendix A, we describe numerous loss reserving methodologies and our implementation of them. In Appendix B, we classify these methods into families based on various common characteristics. In Appendix C, we provide more detailed descriptions of the environmental scenarios evaluated. In Appendix D, we describe how to read the graph that we use to present many of the results. In Appendix E, we rank the methods by their accuracy in various environments. In Appendix F, we show in what direction, if any, the methods were biased in various environments. In Appendix G, we give a complete list of the abbreviations and notations used throughout this paper.

2. BACKGROUND AND METHODS

The following briefly introduces the various dimensions of our work including how we simulated data, how we programmed various loss reserving methodologies and how we evaluated their performance.

2.1 Simulation Method

To create proxy data, we used the following general process. When reviewing this section, it may be helpful to refer to Appendix C, to learn basic properties of the proxy data and how the environments were constructed, and Appendix D, to understand how the tests were applied. Each of the data components that we created are italicized in the section below.

First, we deconstructed the loss process into basic component parts and a system of mathematical functions to describe the relationships between the parts. Essentially, we used *exposure* and *frequency* vectors as a starting point to produce ultimate claim counts. We then applied *incremental-reported-on-unreported claim count patterns* to derive reported claim counts at various evaluations. We then applied *incremental-closed-on-incremental-reported claim count patterns* and *incremental-closed-on-open claim count patterns* to get closed and open claim counts at various evaluations. To estimate incremental loss payments, we applied *incremental-paid-on-closed severities* and *incremental-paid-on-open severities* to the incremental closed and open claim counts, respectively. To estimate case reserves, we applied *case-reserve-per-open severities* to the open claim counts.[4] Both the paid severities and reserve severities differ by evaluation period.

We parameterized each proxy data component mentioned above with real world data to produce a deterministic data set with a stable environment. At this stage, the development patterns for each accident period are identical. To assure that the result was realistic, we evaluated various aggregate diagnostics of the proxy data, such as cumulative loss development patterns and loss severities by age of development and compared the results to diagnostics of the underlying data.

To build each of the environments, we adjusted the basic deterministic components based on the unique characteristics of the environment (as described in Appendix C). For example, in environment 4, the change in case reserve adequacy affects reserved severities but often has no impact on claim counts or paid severities. Each of the environments contains an identical stable history of loss development data (i.e., the upper left portion of the triangle), prior to the first testing period (as defined in Appendix D). The first environmental change (whether applied on an accident-year or calendar-year basis) coincides with the first testing period. We do not apply any tests to accident years prior to the start of the first environmental change (although we can reasonably assume that some of the environmental changes affect the ultimate losses of these older accident years).

The process above results in deterministic data sets for each environment. In our last step, we produced stochastic data sets by applying noise multipliers (by accident year and evaluation period) to each of the basic deterministic components mentioned above (except for *exposure* and *frequency*).

[4] Because simulated paid losses are not directly based on simulated case reserves, a change in case reserve adequacy does not affect paid losses. However, as mentioned later, the simulated noise multipliers are correlated between data components, so that random changes in paid losses are not independent from random changes in case reserves.

Each of the noise multipliers is a normal random variable with a mean of 1.0. We estimated coefficients of variation based on the underlying real world data that we used to parameterize the proxy data. We also correlated the noise multipliers applied to each data components based on correlations that we observed in the real-world data (i.e., higher incremental paid loss on open claims is correlated with higher incremental paid loss on closed claims).

2.2 Loss Reserving Methods

Previous work regarding suitability testing of loss reserving methods focuses on a few methods. To provide a more complete picture, we included techniques beyond the traditional actuarial methods. We focused on methods that can be automated and that do not require optimization routines or knowledge of advanced mathematics or computer science to implement. Where possible, we contacted the authors for exact implementations of their methods to ensure accuracy; we are grateful to those who responded. While our paper is not a complete survey of loss reserving methods, we have included many of the methods that are commonly used by practicing actuaries when developing losses to ultimate. See Schmidt [20] for an excellent bibliography of more recent literature, and Skurnick [21] for descriptions of earlier methods.

2.3 Tests, Criteria, and Statistics

Estimating ultimate loss is vital for major actuarial functions, including loss reserving and ratemaking. While the reserving actuary is interested in projecting ultimate loss for all immature years, the ratemaking actuary may be interested only in projecting ultimate loss for the latest few years. To simplify the presentation of results, we focused on evaluating methods on how well they project loss from earliest evaluation (i.e., 12 months) to ultimate.

There are a variety of criteria that an actuary can use to evaluate the performance of a loss reserving method: accuracy, bias, stability, responsiveness, robustness, consistency, independence, etc. However, we focused on those we believe most important for the practicing actuary—accuracy and bias. We chose accuracy for the obvious reason and bias because it is often helpful to know which methods err in opposite directions, in order to provide upper and lower bounds around the actuary's estimate.

To measure the accuracy of a method, we used the mean absolute percentage error statistic.[5] We

[5] We use the mean absolute percentage error for three reasons. First, accuracy is not dependent on whether a method misses high or low, but rather how close the method is to the true value. Second, we chose the absolute value, rather than the commonly used squared error, as the latter implicitly is a function of the standard deviation and as such does

defined error as the projected ultimate loss minus the actual ultimate loss and the percentage error as the ratio of the error to the actual ultimate loss. We also used the mean percentage error statistic, which represents a combination of accuracy (magnitude of error) and bias (direction of error).

We did not attempt to measure stability or responsiveness, but we use these terms qualitatively. If we describe a method as stable, we mean that its estimates are "sticky" or relatively unaffected by noise or environmental changes. If we describe a method as responsive, we mean that this method corrects itself to produce accurate results shortly after an environmental change. Responsive methods, however, often suffer from a temporary period of inaccuracy during a period of change.

Throughout this paper, we generally refer to results within the context of accuracy. If we say that a method is distorted by or susceptible to a change, we mean that the accuracy is reduced.

3. CAVEATS

Prior to discussing results, we should highlight some of the possible biases in our simulation model and caution the actuary against blindly applying the results, without serious consideration of the differences in situation.

3.1 Specific Books of Business

The most obvious bias in our work is that it necessarily reflects the data we used to parameterize the simulations. The underlying line of business is the medical component of workers compensation. To parameterize the proxy data, we relied on publicly available California industry data, as summarized by the Workers' Compensation Insurance Rating Bureau of California (WCIRB). If we had chosen data from a different region, for example, the errors would be different but the main conclusions would likely be similar. Also, workers compensation is characterized by partial payments on open claims. If we substitute workers compensation with a long-tailed liability line of business, in which there are very few payments prior to claim settlement, there may be differences in some of the conclusions, particularly with methods that separate loss between frequency and severity components. Application of similar testing to other sets of data and other lines of business represents an opportunity for future research.

---

not purely assess accuracy. Finally, we use "percentage" error so as not to give disproportionate weight to tests with large dollar values.

3.2 Number of Evaluation Periods

For presentation purposes, we used 11 evaluation periods for each accident period, where the first 10 evaluation periods represent normal development and the eleventh evaluation period is the tail period (i.e., 120 months to ultimate). However, we also tested results using loss triangles with 30 evaluation periods (i.e., with a tail from 348 months to ultimate) and found that the results were more or less invariant to the number of evaluation periods as long as there was a sufficient volume of data at later evaluation ages. The accuracy of methods based on cumulative data (such as the chain ladder method) was relatively unaffected when switching between 30 evaluation periods and 11 periods. However, a method that is dependent on open claim counts or incremental payments may break down if there are no open claim counts or incremental payments in later evaluations. The practicing actuary should consider the credibility of data in the tail before applying the observations in this paper. Alternatively, an actuary may want to combine different methods for different evaluation periods based on the volume of data available as well as the relevant environmental effects by maturity level.

3.3 Structure of Simulated Data

When designing the building blocks of the proxy data, we chose a structure that we believe is realistic (i.e., it maintains appropriate relationships between data types) without being overly complex. If we had chosen a different underlying structure, it likely would have impacted our testing results somewhat. As mentioned previously, our simulated losses are based on the product of simulated claim counts and simulated severities. If instead we had simulated claim count data independently of loss data, then it is likely that methods that exploit the relationship between claim counts and loss severities (such as the Adler-Kline method) would perform poorly.

3.4 Distribution of Noise

As mentioned previously, in order to produce stochastic data, we simulated noise multipliers based on normal[6] random variables. We then multiplied these random variables by the basic components of our deterministic data, such as paid loss severities. Without knowing the underlying frequency and severity distributions, we used the simplifying assumption that a normal distribution would adequately approximate the shape of aggregate noise affecting the development triangles. To test this assumption, we also considered gamma and lognormal noise multipliers. The results were

[6] Although in theory it is possible for normal random variables to produce negative values, this was not an issue in our simulation, because the standard deviation was small enough so that the probability of a negative value was infinitesimal.

approximately the same.

3.5 Amount of Noise

The results of our testing are based on levels of noise consistent with that observed in the data used to parameterize the proxy data. Our aim was to reproduce the level of noise that would be typical for a large insurance carrier, with a fairly consistent history of homogeneous exposures. In order to assess the sensitivity of our results to various levels of noise, we increased and decreased the coefficient of variation of the noise multipliers. We found that if we had chosen a higher level of noise consistent with a smaller, less credible set of exposures, it is likely that methods based on more granular data (e.g., incremental payments, claim counts, incremental paid severities) would suffer due to the leveraged effect of noise on the basic components of such methods.

3.6 Methods

The results of any loss reserving method are influenced by the way in which the method is parameterized. As noted in a Section 4.2.6, methods whose parameters are based on short-term observations are more responsive and less stable than those based on long-term observations. Similarly, methods that rely on estimates of loss trend, such as the incremental additive method, can be significantly affected by choosing a short-term or long-term trend rate. More complicated methods, such as the Berquist-Sherman adjustments, are dependent on how the methods are constructed.

As much as possible, we have attempted to construct our methods using the same rules as described in the original literature referenced in Appendix A. The actuary should consider how the structure or parameterization of a method may impact its accuracy in various situations.

3.7 Environments

Our results are very much tied to the specific environments we tested and to the parameterizations we chose to describe and define those environments. These are described in detail in Appendix C. For each of the environments, our aim was to model a significant change in the development data, large enough to show a measurable distortion in projection methods, while being reasonably likely to occur. If we had chosen environments with smaller or larger changes, our testing results would have been muted or exaggerated, respectively.

Another element to consider is time. For environment 2, for example, we modeled three years of elevated inflation. If we had increased the number of years of elevated inflation, responsive methods (such as the incremental multiplicative) would have been more accurate and unresponsive methods (such as exposure-based methods) would have been less accurate. The inverse is also true.

Conclusions about which methods perform well or poorly under various conditions (such as "the Berquist-Sherman adjustment for case reserve adequacy is accurate during a period of changing case reserves") would not change if we had chosen different values for the environmental parameters: only the relative difference in accuracy between the methods would change.

3.8 Test Statistics

In some respects, our results directly depend on the tests we used to evaluate the various methods: the mean percentage error and the mean absolute percentage error for the 12-month to ultimate projection. However, we also considered various other statistics such as the mean squared error to measure the volatility and responsiveness with regard to overall accuracy of these methods. We found that using these other statistics did not noticeably change our conclusions regarding the *relative* accuracy of the methods.

3.9 Limitations of Our Recommendations

The recommendations in this paper should not be used in place of the actuary's due diligence and appropriate judgment. Our findings presuppose that the actuary is able to review relevant diagnostics or leading indicators to evaluate the characteristics of the current environment and make assumptions about future conditions. It is outside the scope of this paper to provide a list of diagnostics or to determine how easy or difficult it may be to determine the current environment based on these diagnostics. In some instances, it may be difficult for an actuary to ascertain the precise nature of the underlying environment affecting the loss development data. However, we believe the environments reviewed in this paper are broad enough so that they could be identified with diagnostics or other available information, such as a law change or economic data.

4. RESULTS AND DISCUSSION

The following is our roadmap to help the actuary in evaluating various loss reserving methodologies. In Section 4.1 we present several of our high-level findings and general recommendations. In Section 4.2, we comment on several basic components of loss reserving methods that are not unique to any one method. In Section 4.3, we present findings as they relate to families of projection methods, where the families are defined in Appendix B. In Section 4.4, we present findings by environment.

4.1 General Findings and Recommendations

4.1.1 Stable versus responsive methods in periods of consistent conditions

We found that stable methods outperform methods that are more responsive in environments where, although there is still random noise, conditions are consistent over time (or more generally, when the actuary cannot discern the cause of variability). Stable methods tend to rely on a longer history by design or through selection of parameters, and therefore are better at avoiding distortions due to variability that does not reflect environmental changes. In environments that remain consistent over time, methods using parameters based on longer-term averages outperform those whose parameters reflect shorter histories; exposure-based methods (like the Bornhuetter-Ferguson) beat methods that rely on loss only; and cumulative methods outperform incremental methods. Other methods that incorporate a longer history by construction are the Berquist-Sherman adjusted methods (which perform better than the corresponding unadjusted methods), as well as several more complex methods that rely on the entire triangle to project ultimate loss.

4.1.2 The importance of environmental changes

In practice, it is unlikely that conditions would be consistent over a long period. Loss triangles are constantly subject to forces that can distort loss reserving methodologies. The workers compensation system is subject to forces that cause shifts in loss development data, whether slow and subtle (e.g., a change in the mix of claim types due to a shift away from manufacturing) or sudden and dramatic (such as legislative benefit reform). Therefore, while we found that stable methods would theoretically outperform responsive methods in environments where conditions are consistent, in practice the actuary may not often encounter such environments. More likely, there will be subtle shifts in a manner that is either unknown or at least not yet quantifiable.

4.1.3 Adjusting the data during periods of significant upheaval

During periods of significant upheaval, the mechanical application of any loss reserving method to raw data is unlikely to yield reasonably accurate projections. In fact, we found that all methods tested perform quite poorly under such circumstances. This result highlights the importance of not relying blindly on loss reserving methods when the underlying data has been significantly distorted by environmental changes. In these situations, alternatives include making data adjustments (e.g., restating history to current cost level or current claim mix) or otherwise correcting for environmental changes before applying loss reserving methodologies. If the nature of the distortion is understood, it is also useful for the actuary to identify which method's performance will be most affected.

4.1.4 Responding after periods of severe environmental change

If the assumption is that the system has reached a plateau after major disruptions, our findings would point the actuary toward more responsive methods. Our work confirms that methods with short or no *memory* (such as methods where parameters are selected based on recent observations only or incremental methods) fare better than those that use a longer history (e.g., longer-term averages or cumulative methods) under these circumstances.

4.1.5 Type of change versus direction of change

In general, how a method performs under each environment is defined by elements of the environment that change, rather than by the *direction* of the change. For example, we found that methods that work well when claim settlements slow down prove to also work well when those settlements accelerate.

4.1.6 Accident year vs. calendar year effects

We tested the impact of various environmental changes, including accident year shifts (such as an increase in frequency from one accident year to the next) and calendar year shifts (such as a change in inflation, which affects all accident years simultaneously). Our analysis showed that:

(i) Calendar year changes (e.g., inflation that simultaneously impacts all accident years) always affect the accuracy of loss reserving methodologies, since they always distort development patterns.[7] As a result, consideration should be given to adjusting the data underlying the development projection.

(ii) Accident year changes (e.g., change in frequency) do not affect accuracy of methods based on loss unless the shift also causes a change in loss development patterns (e.g., a change in the mix of claim types).

(iii) While an accident or calendar year shift will distort most methods for many years after

[7] Even methods that do not rely on development patterns are affected, because calendar year changes like those described in this paper affect actual unpaid losses, resulting in prediction errors.

the change, incremental methods, by their nature, are able to respond immediately after the change to calendar year shifts.

(iv) Accident year shifts may distort even incremental methods for many years after the change, as the projection for the latest accident year is dependent upon observations from older accident years.

The following figures compare the incremental multiplicative (IM) method with the chain ladder (CL) method as an illustration. Both methods are parameterized equivalently save that the former is applied to incremental paid loss, and the latter is applied to cumulative paid loss.

Figure 1[8] compares the mean error of the IM and CL methods[9] in an environment affected by a calendar year shift in medical inflation.[10] In the first three testing periods, inflation is higher than normal (15%) and in the fourth and subsequent periods, inflation is consistently at its historical average rate (5%). Both methods are distorted; however, the incremental method immediately corrects itself after the change. The cumulative CL method does not—and it will not produce unbiased estimates until the distortion disappears from the data the actuary is using.

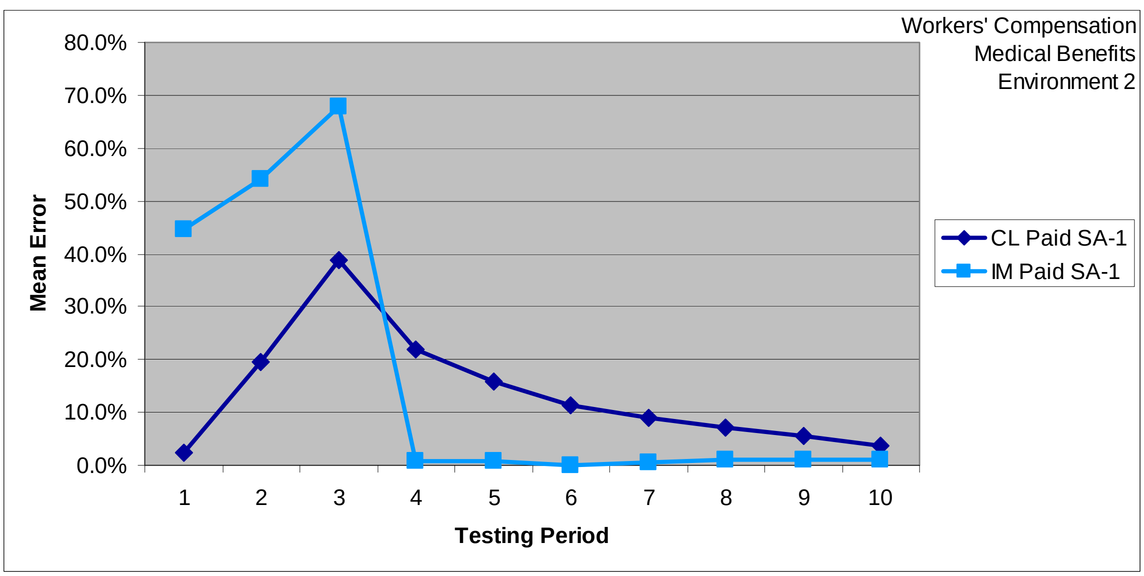


Figure 1: Comparison of incremental and cumulative methods during a calendar year shift.

[8] Please refer to Appendix D, which describes in detail how to interpret the graphs in the paper.
[9] Appendix A provides information about each of the methods and how they were programmed.
[10] Appendix C describes each of the environments in detail and may be useful in understanding logic underlying the conclusions presented.

Figure 2 compares the IM and CL methods during a permanent accident year shift in the frequency of serious injuries, which distorts the development patterns. Both methods are affected, but the IM method corrects itself more quickly after the change.

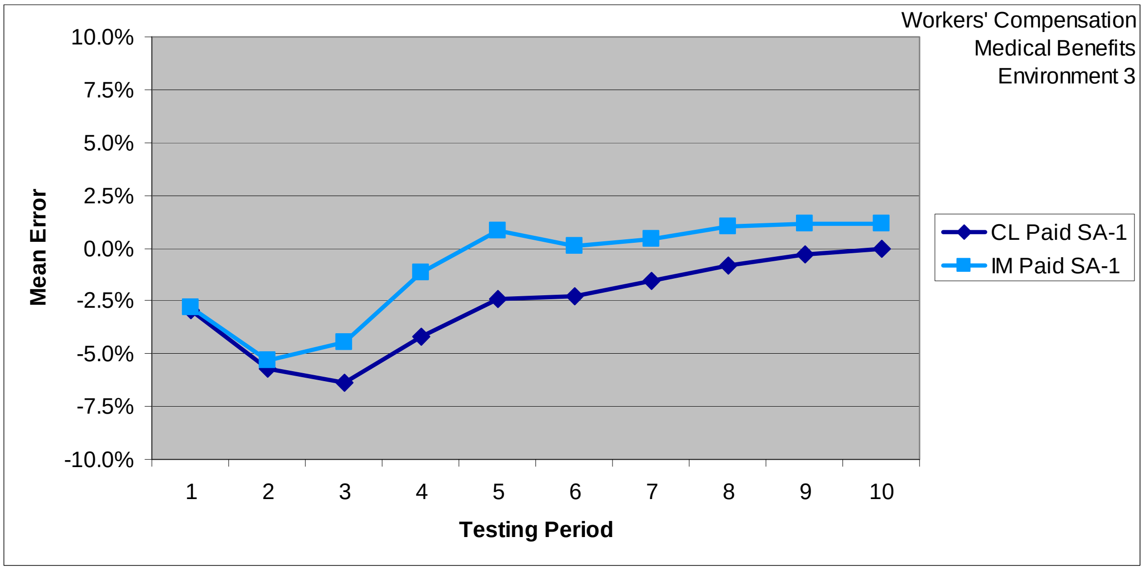


Figure 2: Comparison of incremental and cumulative methods during an accident year shift that distorts development.

Finally, Figure 3 illustrates that in the event of an accident year shift that does not distort development patterns (such as exposure growth), both incremental and cumulative methods are unaffected.

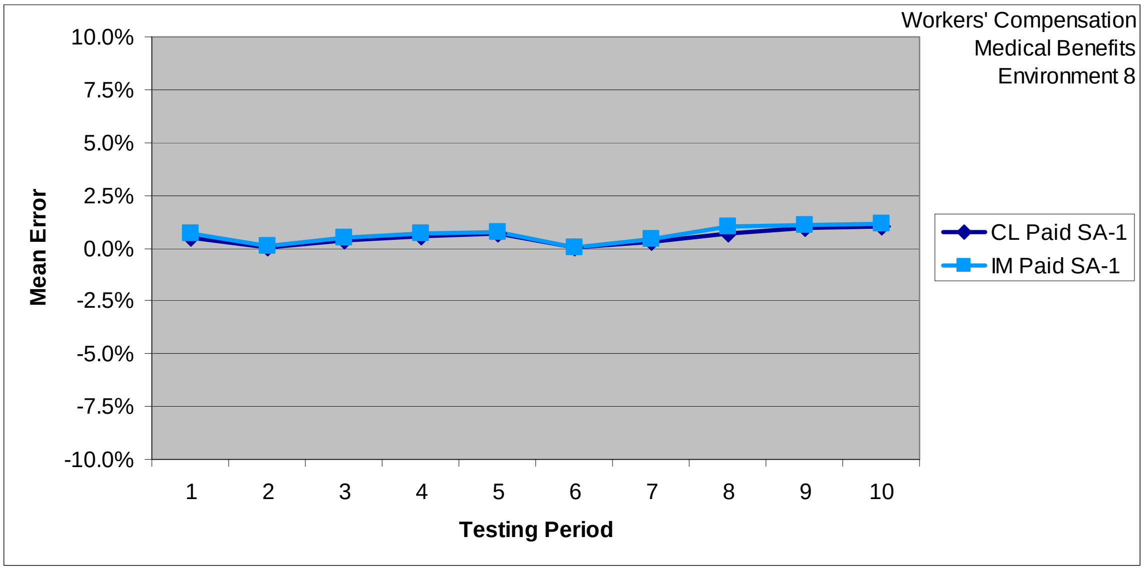


Figure 3: Comparison of incremental and cumulative methods during an accident year shift that does not distort development.

4.1.7 Deterministic versus stochastic analysis

The results of tests that do not take into account residual noise (e.g., tests based on well-behaved, deterministic data) may lead to conclusions that are not appropriate for application in the real world.[11] While enlightening and useful for assessing the reasonability of an approach, we believe that tests performed on deterministic data tend to over-recommend responsive methods, which are susceptible to noise, and under-recommend less responsive methods. When reasonable levels of noise are added, the accuracy of responsive methods is more adversely affected than that of stable methods. This conclusion may caution actuaries against evaluating methods using simplistic examples, which ignore the real-world noise dimension.

[11] Note that this is the only section in the paper where we show test results based on deterministic data. All other tests, shown before and after, are applied to stochastic data.

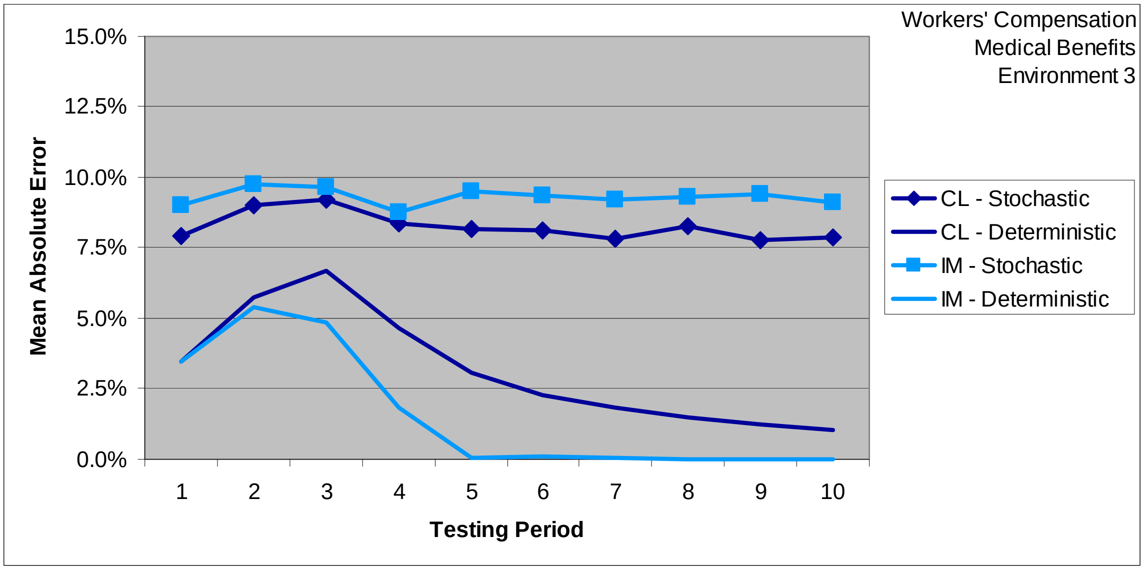


Figure 4: Comparison of the cumulative and incremental methods during both the stochastic and deterministic variants during an increase to a new plateau in the frequency of serious injuries.

Figure 4 shows an example of this phenomenon. Without any residual noise, the incremental method is more accurate and responds more quickly to the change than the cumulative method. However, after we add residual noise to the simulated data, the incremental method is more affected than the cumulative method. Furthermore, the noise in this particular situation is more important to the accuracy of the methods than the environmental change. This analysis invariably depends on the level of noise inherent in the data. When the data is noisy, such as in lines of business characterized by low frequency and high severity, approaches involving more stable methodologies and parameter selection are preferable.

4.1.8 Independence and bias

In cases where two independent methods are biased in opposite directions and produce similar magnitudes of error, a combined method based on the average of those two methods often outperforms either method individually, as the positive and negative errors offset. Additionally, it is helpful to know which methods are biased in opposite directions (and in which environments), as the best estimate is likely to fall between such methods.

4.1.9 Limitations of hindsight testing

Our tests of accuracy are designed to measure errors in the projected ultimate value that, for long-tailed lines, are not capable of being observed in practice. In practice, actuaries typically evaluate success by reviewing changes in *estimates* of ultimate loss over a shorter period (e.g., less than five years). Our analysis showed that such commonly used tests may lead the actuary to discard a good method, which, while it may appear to significantly over- or under-predict in the short term, in actuality performs quite well in predicting the ultimate value.

4.2 The Components of a Loss Reserving Function

Most loss reserving methods are built from the same basic component parts. For example, methods that rely on loss development factors (LDFs) use some type of average of recently observed factors. However, each loss reserving method has some unique aspects that makes it different from other loss reserving methodologies. Since we intend to study these unique aspects, we made sure to implement all methods as consistently as possible, so that the difference in test results directly represents the difference in the unique aspects of the method (e.g., when comparing the chain ladder method with the incremental multiplicative method, we parameterized the loss development factors the same way so that the comparison would only differentiate between structure of the methods).

However, during this process, we noticed that many seemingly different methods are, in practice, identical. The following sections present these results as well as results about characteristics of the various shared components.

4.2.1 The equivalence of Fisher-Lange and Adler-Kline

An excellent example of two distinct methods generating essentially identical results is that offered by the Fisher-Lange (FL) and Adler-Kline (AK) claims closure models. Although the authors describe different methods of computing future severities, the claims component is identical.

The FL method, as described in Fisher and Lange [6], is a frequency-severity approach that operates on report-year data. A key advantage of using report-year data is that the ultimate number of claims is fixed at the end of each report year. The only development is on loss amounts and future claims closure. However, in the absence of report year data, ultimate claim counts can be projected, and the FL method is equally applicable. This is the approach we took in our analysis. After making this modification, however, Fisher and Lange's closure ratios produce identical incremental closed claim counts as Adler and Kline's disposal ratios.[12] This can be shown algebraically, but Figure 13 and Figure 14 provide illustrations of this phenomenon.

4.2.2 The equivalence of the cumulative frequency-severity method with the chain ladder method

The cumulative frequency-severity approach (FS), as described by Friedland [7], projects ultimate loss by applying the chain ladder method separately to claim counts and claim severities. If this approach is parameterized using the latest set of development factors, it is algebraically equivalent to the chain ladder approach on cumulative loss. This relationship holds true whether or not the definition of claim counts is internally consistent and homogeneous. Furthermore, any other parameterization, if applied consistently to both the cumulative frequency-severity approach and the chain ladder approach, will produce results that are virtually identical.

This is not to say that the cumulative frequency-severity method is without purpose. Frequency, in particular, is often impacted by external factors that may not be reflected in the underlying data (e.g., changes in economic conditions and legislative changes). Often there is an advantage to incorporating information exogenous to triangle data when selecting future severities and closure ratios, especially if future frequency and severity are expected to differ from historical frequency and severity. Friedland [7] notes that "[frequency-severity methods] can be particularly valuable when an organization is undergoing changes in operations, philosophy or management."

4.2.3 The equivalence of the incremental additive method, Bühlmann's complementary loss ratio method and the chain ladder method

Both the incremental additive (IA) method and Bühlmann's complementary loss ratio (CLR) method project future incremental loss as a means of estimating the outstanding liability. The IA method computes these amounts based on the relationship of historical incremental loss to on-level exposure, and the CLR method trends forward historical incremental loss. These methods are algebraically equivalent to the chain ladder if parameters are based on the latest observation and loss trend is estimated using a link-ratio approach. By this, we mean that one trend factor is computed for each set of accident years and that these trend factors are calculated as the ratio of the cumulative loss at the latest period to the ratio of the cumulative loss at the earlier period. While this is not the only way (or the best way) to compute trend, it does indicate that both the incremental

[12] This equality only holds when parameterizing the methods using simple averages of one year. Longer-term averages and other types of averages will produce results that are slightly different.

additive method and Bühlmann's complementary loss ratio approach are to some extent intrinsically linked to the chain ladder methodology, and thus it may not offer an independent estimate of loss.

4.2.4 Selecting a projection base (paid loss, reported loss, case reserves, or exposure)

Each loss reserving method reviewed is based on one or more projection bases: paid loss, reported loss, case reserves, or exposures. Each of these projection bases has its own unique advantages and disadvantages that can be good predictors of a method's performance in various environments. For example, methods based on paid loss are immune to changes in case reserve adequacy. However, paid methods appear to be more susceptible to residual noise than methods based on reported loss, because paid methods lack the useful information provided by case reserves, and there is greater prediction error in the paid LDFs due to the larger magnitude of the factors.[13] Methods based on reported loss are quite susceptible to distortions in the reporting pattern caused by changes in case reserve adequacy or claim settlement rates. Methods based on case reserves can be even more distorted than reported methods during changing conditions, as they lack the stability provided by adding paid loss. However, methods based on case reserves are often the most responsive after a period of changing conditions, as they contain information about future loss amounts that is not distorted by volatility in historical amounts. Unlike loss-based methods, methods that rely solely on exposures (such as the budgeted loss method) are completely unresponsive to movements in loss amounts as they represent an a priori estimate of ultimate loss rather than a current estimate of future remaining payments. Generally, exposure-based methods produce stable estimates during changing conditions, but they can err wildly when there are significant changes in loss costs that are not reflected in the underlying exposures. However, as exposure-based methods are often independent of the other loss reserving methods, they are good candidates for establishing bounds within which ultimate loss is likely to be.

4.2.5 Incorporating loss trend

Most of the methods reviewed in this paper incorporate the concept of loss development (i.e., measuring changes in an accident year's losses from one evaluation period to the next). Some of the methods reviewed also incorporate the concept of loss trend (i.e., measuring changes in losses from one accident period to the next).

[13] This observation may depend on the evaluation age and the line of business. Also, in the real world, reported losses may be distorted by small undetected changes in case reserve adequacy, which may increase error in projections based on reported loss.

Several of the methods, including Bühlmann's complementary loss ratio method, Adler-Kline, Fisher-Lange, and the two Ghezzi methods, rely on trending forward historical loss severities from the loss triangle. For these methods, a trend rate is calculated for each evaluation age separately, by fitting a line to the log of the severity amounts. This has the advantage of capturing different trend rates by evaluation age, to the extent that they exist. However, separate trend rates are more susceptible to residual noise than a single trend rate for the entire triangle, especially when the trending period is limited to relatively few data points.

To parameterize the modified Bornhuetter-Ferguson method, which has a self-correcting loss ratio, we employed a three-year trend (i.e., four data points) on an accident year basis. For the incremental additive method, we chose a three-year trend measured on a calendar year basis (i.e., the trend observed based on calendar-year payments). The incremental additive method with a three-year trend often produces very similar results to the incremental multiplicative method based on the latest three years of observations. By contrast, using a long-term trend produces more stable results.

A short-term trend benefits from responsiveness following the end of an environmental change (i.e., when a new period of normalcy is reached), but it may result in wildly inaccurate results during a period of upheaval. A long-term trend, similar to long-term averages of development factors, produces more stable results but is slow to react to emerging conditions.

4.2.6 Comparison of short-term vs. long-term parameterizations

Short-term averages are more responsive than longer-term averages, which are more stable. Figure 5 shows Marker and Mohl's method parameterized using a one-year simple average and a three-year simple average during a three-year bubble in medical inflation.[14] Henceforth, we will generally show only one type of parameterization and work under the assumption that the observed errors will either be muted (and delayed) or intensified depending on whether a longer-term or shorter-term average is used, respectively.

[14] As mentioned previously, all environmental changes begin in the first testing period. Each environment is described in greater detail in Appendix C.

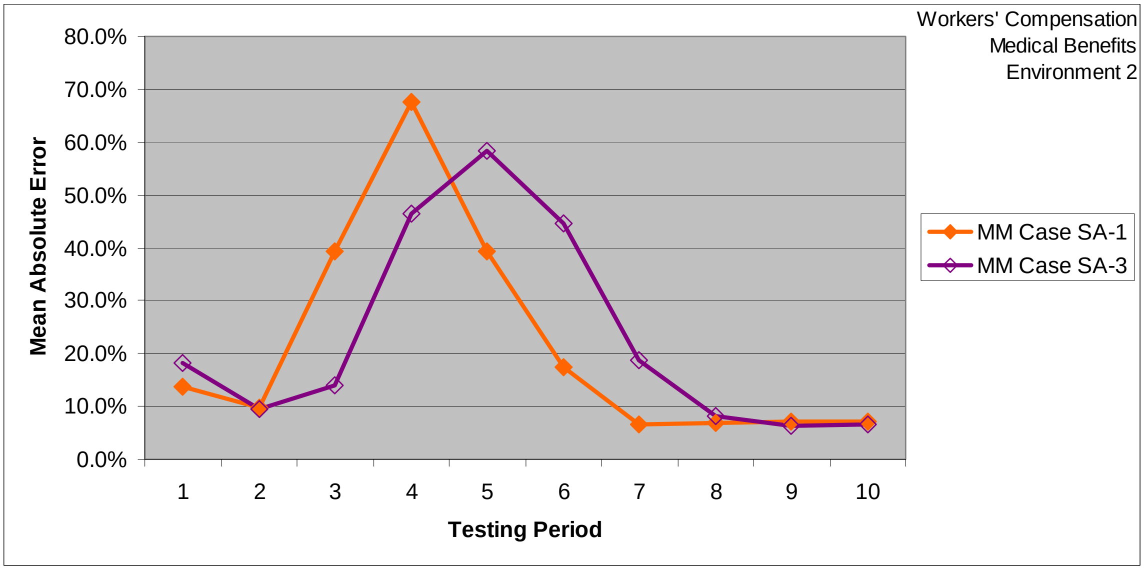


Figure 5: Comparison of a long-term parameterization to a short-term parameterization during a bubble in medical inflation.

4.2.7 Comparison of simple and volume-weighted averages

In our proxy data set, the volume of exposures is stable over time; because of this, there is little difference in accuracy by choosing development factors based on simple averages or volume-weighted averages. See Figure 6 for a comparison of methods based on simple and volume-weighted averages. For simplicity, for the remainder of the paper, we focus on methods that rely on simple averages (or more complex parameterization methods such as regression). However, in the real world, the actuary should be aware that simple averages can be distorted by individual accident years with a small volume of exposures, which are more volatile as a result.

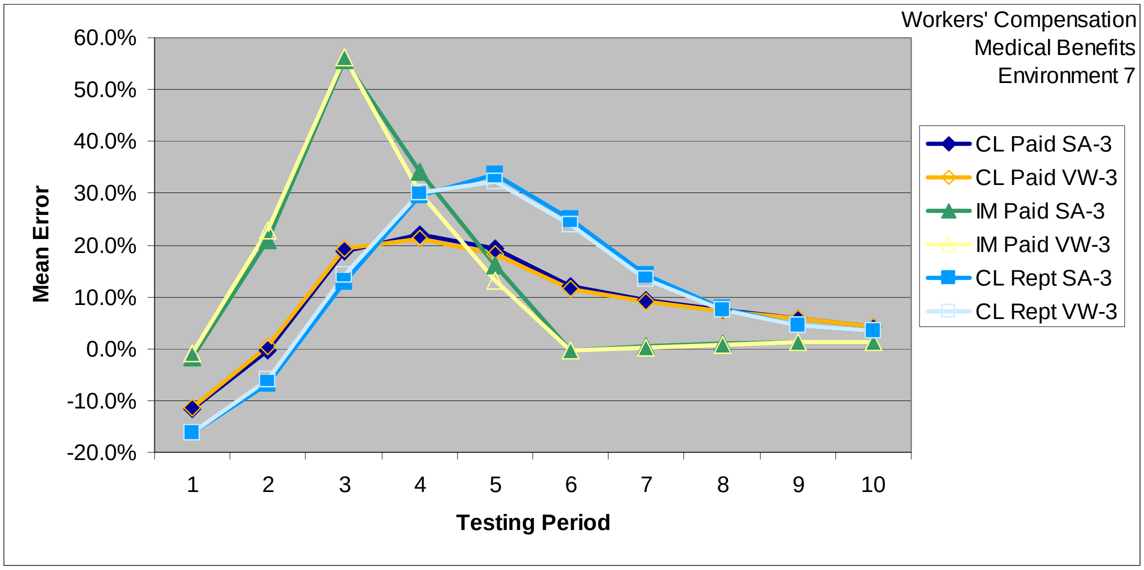


Figure 6: Comparison of volume-weighted and simple averages for the chain ladder method on paid loss, the chain ladder method on reported loss and the incremental multiplicative paid loss. This environment consists of a bubble in the rate of medical inflation coupled with an increase in the frequency of serious injuries.

4.3 Findings by Family of Methods

We grouped the various methods into several families based on certain characteristics (e.g., exposure-based methods, frequency-severity methods, incremental methods, regression methods, etc.). These classifications can be found in Appendix B. The following conclusions all pertain to one family or another and are meant to identify differences between methods within a family. By reviewing these results, we can draw conclusions that tie our understanding of how the methods are constructed to how accurately they perform in various environments.

4.3.1 Exposure-based methods

Figure 7 shows the mean error of methods based on exposures and/or paid loss in environment 2, a temporary three-year period of high calendar-year inflation. In the first year after the onset of inflation, all methods underestimate because they are unaware of the higher-than-expected inflation. Soon after the start of high inflation, the paid method overestimates the ultimate loss because it expects the higher inflation to continue indefinitely. The budgeted loss method (BL) never recognizes the change and therefore always underestimates. The Bornhuetter-Ferguson (BF) and Benktander (BT) methods, meanwhile, lie between the extremes. They still underestimate, but not to the same degree as the budgeted loss method. The modified Bornhuetter-Ferguson (MBF) closely follows the chain ladder method, because the trend underlying the MBFs expected loss ratio is based on projections of ultimate loss produced by the chain ladder method. The incremental additive method has a self-correcting trend rate, so that after inflation reverts to historical norms, the IA produces unbiased estimates.

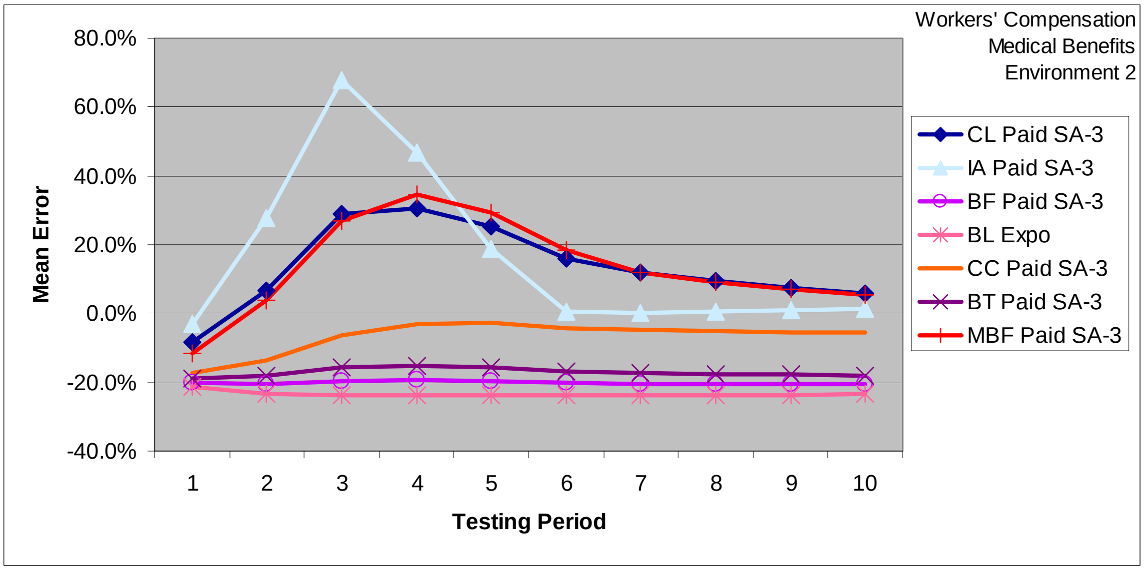


Figure 7: Comparison of various exposure-based methods on paid loss during a bubble in medical inflation.

Figure 8 shows the same methods based on reported loss instead of paid loss. Because the expected percentage reported at 12 months is higher than the expected percentage paid, the BF and BT on reported loss methods are relatively more responsive.

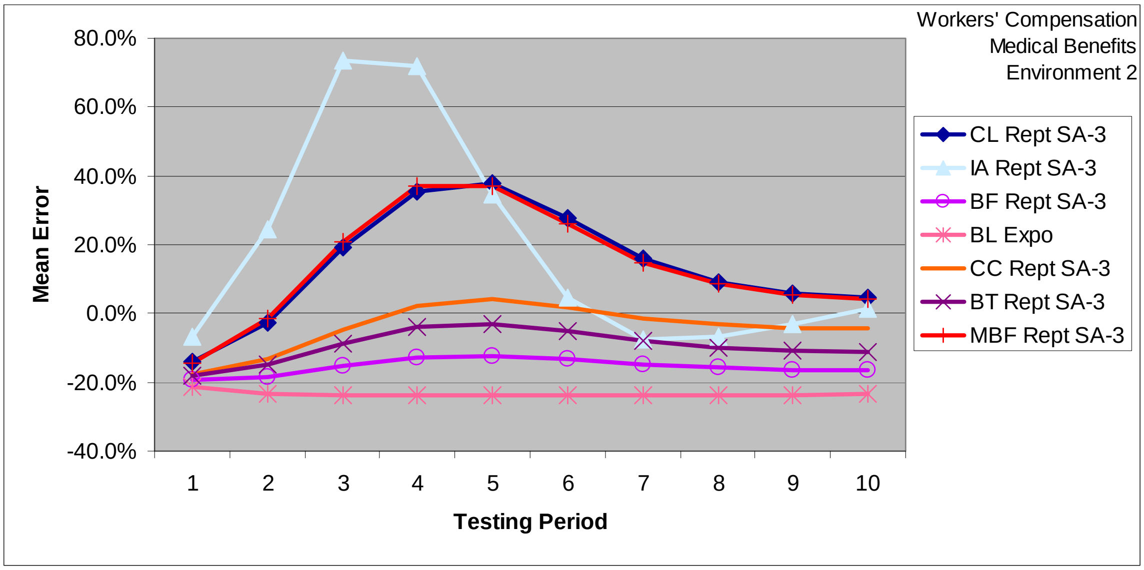


Figure 8: Comparison of various exposure-based methods during a bubble in medical inflation. Where Figure 7 shows the exposure-based methods on paid loss, this figure shows the exposure-based methods on reported loss.

Figure 9 shows the exposure-based methods when subject to an acceleration in claim settlement rates. All methods initially overestimate the ultimate loss, because in this environment, the faster claim closures result in lower ultimate losses. The BF and BT methods, by their nature, fall somewhere between the results of the BL and CL methods. Similar to the previous example, the IA method is the most responsive after the change.

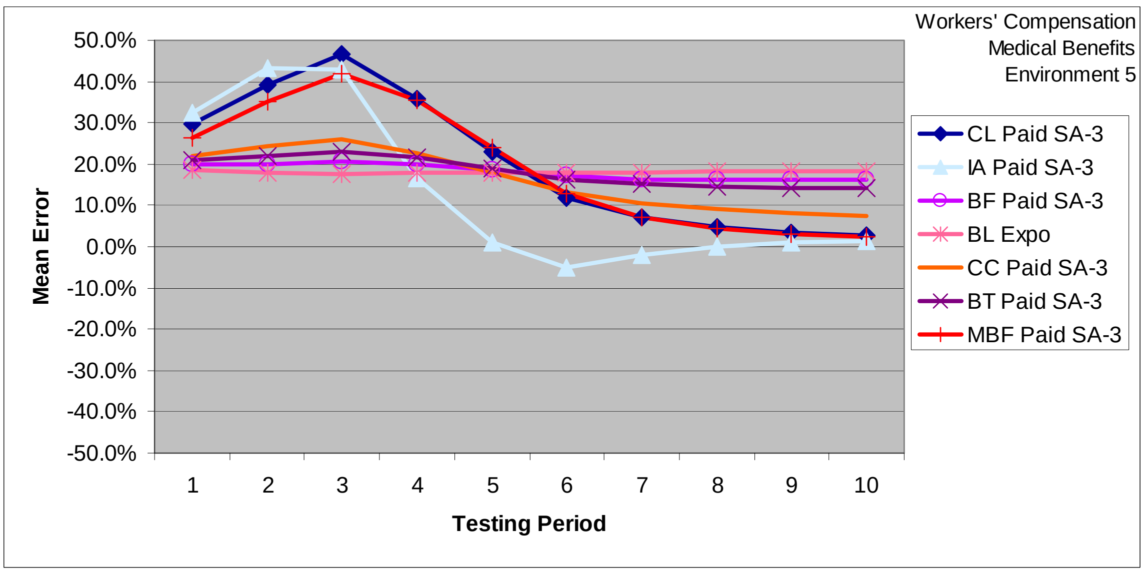


Figure 9: Comparison of various exposure-based methods during a permanent acceleration in claim settlement rates.

4.3.2 Regression-based methods

Regression-based loss reserving methods appear frequently in actuarial literature. The following are some general considerations regarding members of the regression family.

First, consider Brosius's least squares development (LS) method since it serves to highlight some advantages and disadvantages of regression methods. Figure 10 and Figure 11 make it obvious that LS is by far the worst method *during* the period of change, but that it responds much faster than any other method *after* the change, quickly becoming one of the most accurate methods. To understand this, consider how LS works. The LS method begins with the oldest accident years and uses data at the $n-1$ evaluation period to project data at the n^{th} evaluation period (or ultimate) by fitting a line through least squares. Subsequently, this projection is added to the vector of n^{th} evaluation (or ultimate) values and with the addition of another accident year, data at the $n-2$ evaluation period is used to project ultimate loss at the n^{th} evaluation period, and so forth. This approach is reasonably accurate as long as future actual observations are within the range of historical observations. But when future values fall outside of the range of history, the model must extrapolate and errors are increased. This is true of most regression methods; however, the problem is exacerbated by the LS method as the predictions are iteratively fed back into the model in such a way that the error propagates itself.

However, immediately after the change, when conditions stabilize, the LS and other regression methods correct themselves with varying degrees of responsiveness. This is an example of the stability/responsiveness trade-off as determined by the number of parameters in the model. In general (but not always), methods with more parameters are unstable during changing conditions (i.e., they are greatly affected by the changing conditions and produce inaccurate results), but very responsive after the conditions stabilize. As the number of parameters increases, the amount of variability in the dependent variable understood and explained increases as each successive parameter can mine for the residual relationship. However, when conditions are changing, these types of regression models overfit to the historical data (as described above) and produce more inaccurate results. Figure 10, in particular, provides an example of this. The LS method (2 parameters) and Murphy's least squares linear (Mur-LSL) parameterization (3 parameters) are less stable during the changing conditions and more responsive after than Murphy's least squares multiplicative (Mur-LSM) parameterization (1 parameter), and the chain ladder method based on a simple average of all observations (CL SA-All).

Consider now the multivariate (MV) method. This method actually performs very well during an increase in case reserve adequacy coupled with an acceleration in claim settlement rates as shown in Figure 10. However, the MV method performs rather poorly during a bubble in medical inflation as shown in Figure 11. In the simpler environment, the MV method overfits, however, in the more complex environment, the MV is able to combine disjoint pieces of information and perform relatively well.

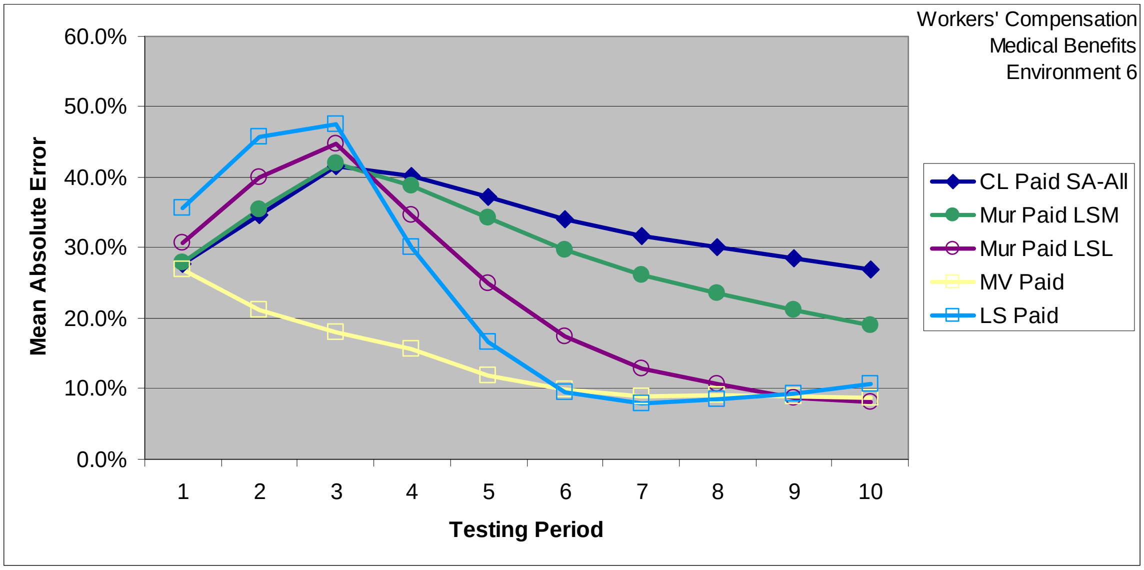


Figure 10: Comparison of various regression methods during a permanent acceleration in claim settlement rates coupled with a permanent increase in case reserve adequacy.

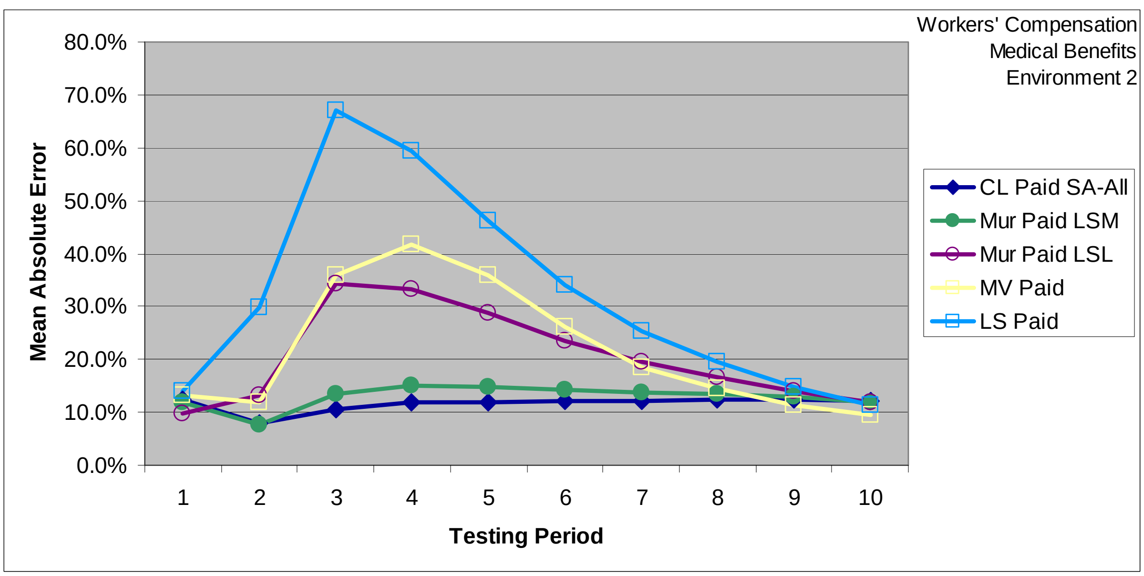


Figure 11: Comparison of several regression-based methods during a bubble in medical inflation.

If the actuary is able to find a regression method that does a good job of describing the loss process, then it may produce accurate results. However, regression methods just as often "overfit" historical data without providing a good prediction of future observations. As a cautionary note, Figure 12 compares each of Verrall's three log-linear models, as described in Narayan and Warthen [15], with the CL method in environment 7 (bubble in the rate of medical inflation coupled with an increase in the frequency of serious injuries). The first model (LL1) has parameters that vary freely by accident year and evaluation period. The second model (LL2) is restricted so that parameters vary only by evaluation period. The third model (LL3) is restricted further so that its parameters do not vary by accident period or evaluation period.

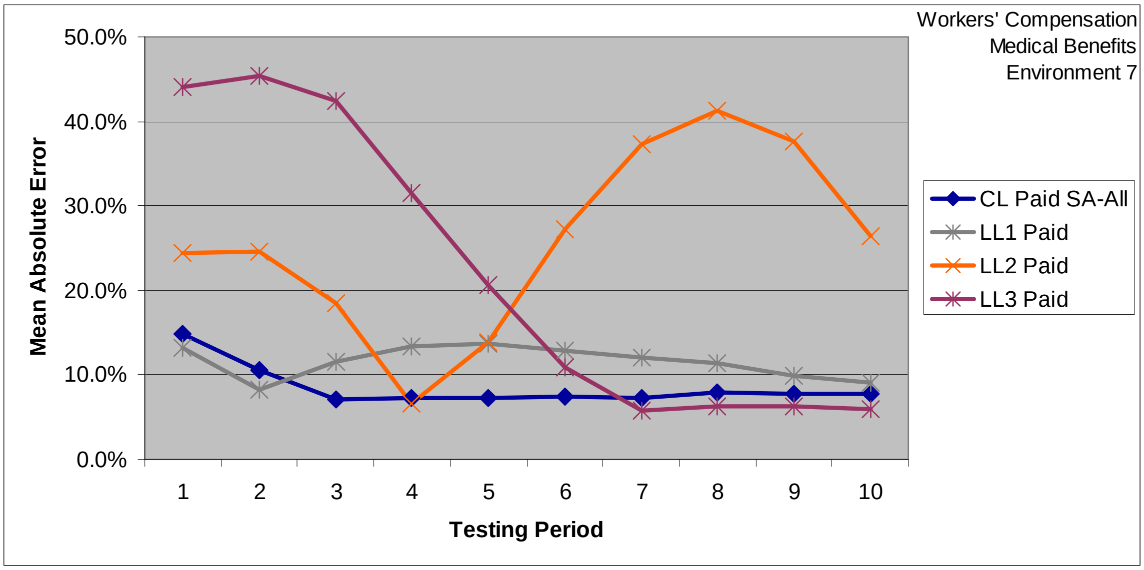


Figure 12: Comparison of Verrall's log-linear models during a bubble in medical inflation coupled with a permanent increase in the frequency of serious injuries.

What is immediately obvious about Figure 12 is the wide variation in results. This highlights the idea that while there are an infinite number of elegant regression models that adhere to theoretically desirable loss development properties, when these models are applied in practice to data that does not mimic those properties the results will be less than desirable. LL3, the simplest of these models, is unable to capture the complex interactions of the calendar year inflation with accident year increase in the frequency of serious claims and errs significantly until these changes work themselves out of the data. LL2 produces large and seemingly unpredictable errors, first underestimating the ultimate loss and then overestimating in later periods. LL1, the most complex of these models, is actually the most accurate during the change, however, it overestimates after the data have stabilized. At the end of the day, the actuary would have been better off using the CL method.

This section further highlights how important it is that the actuary gather both qualitative and quantitative insights from underwriters, claims administrators, and other data sources to improve understanding as to what disturbances underlie the data and, consequently, which methods and parameterizations are likely to over-, under- or correctly estimate future unpaid loss amounts.

4.3.3 Frequency-severity methods

In environment 2, a three-year bubble in medical inflation (see Figure 13), the FS method is distorted in a similar manner as the chain ladder method (not shown). The FL and AK methods start off well, but produce less accurate results a few years after the onset of higher inflation. This is mainly because these methods project future severities using exponential growth curves fit at each evaluation age, which are distorted by the kink in growth caused by a bubble in medical inflation.

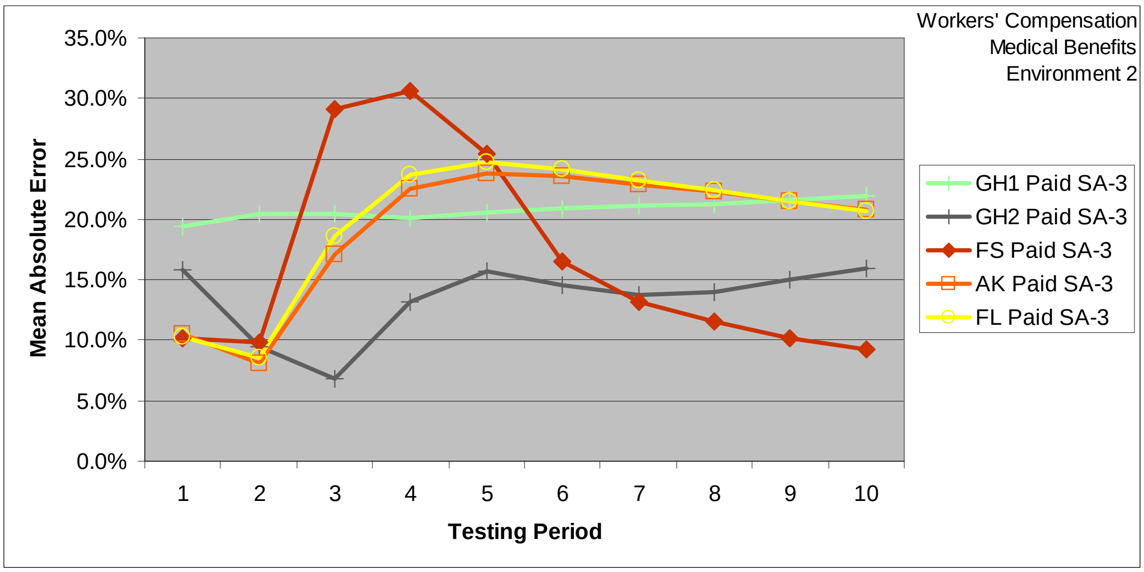


Figure 13: Comparison of various frequency-severity methods during a bubble in medical inflation.

In environment 6, a permanent change in case reserve adequacy combined with a permanent acceleration of claim settlement rates (see Figure 14), each of Ghezzi's methods (GH1 and GH2) shows its merit. This is because these methods are especially effective when data undergo a change that has little or no effect on actual ultimate loss, but the change serves to confuse and distort more traditional loss reserving methodologies. As mentioned previously, the FS method performs similarly to the chain ladder and does not offer any advantage over the AK and FL methods.

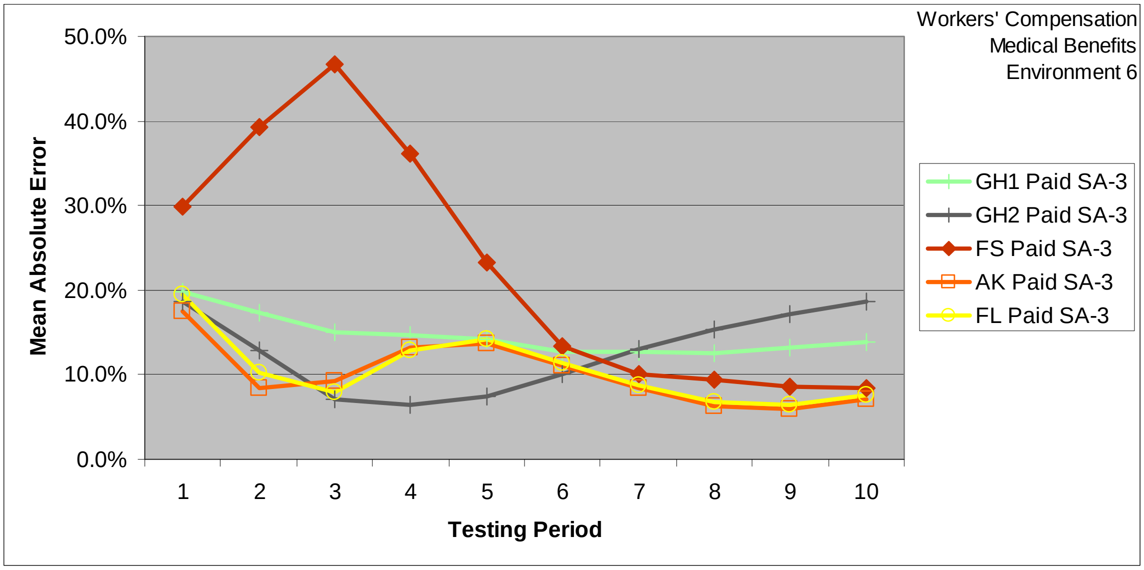


Figure 14: Comparison of various frequency-severities methodologies during a permanent increase in case reserve adequacy coupled with a permanent acceleration in claim settlement rates.

4.3.4 Berquist-Sherman adjustments

We use the phrase "Berquist-Sherman adjustments" to refer to the family of methods that adjust historical triangles prior to projecting. These methods are particularly accurate in environments where the historical change is similar to the one for which the adjustment corrects. Of course, if the emerging environmental change is different from the historical adjustment, then the accuracy of these methods may suffer.

The Berquist-Sherman adjustment for changes in case reserve adequacy (BSRA) method adjusts very well for changes in case reserve adequacy (see Figure 15). Furthermore, this method will also perform reasonably well during an acceleration in claim settlement rates (see Figure 16), where although there is no change in case reserve adequacy per se, there is a change in average outstanding case reserves. This may happen because the BSRA is a more stable (by construction) method and the adjustment for reserve adequacy somewhat dampens the high development factors that distort the chain ladder method on reported loss.

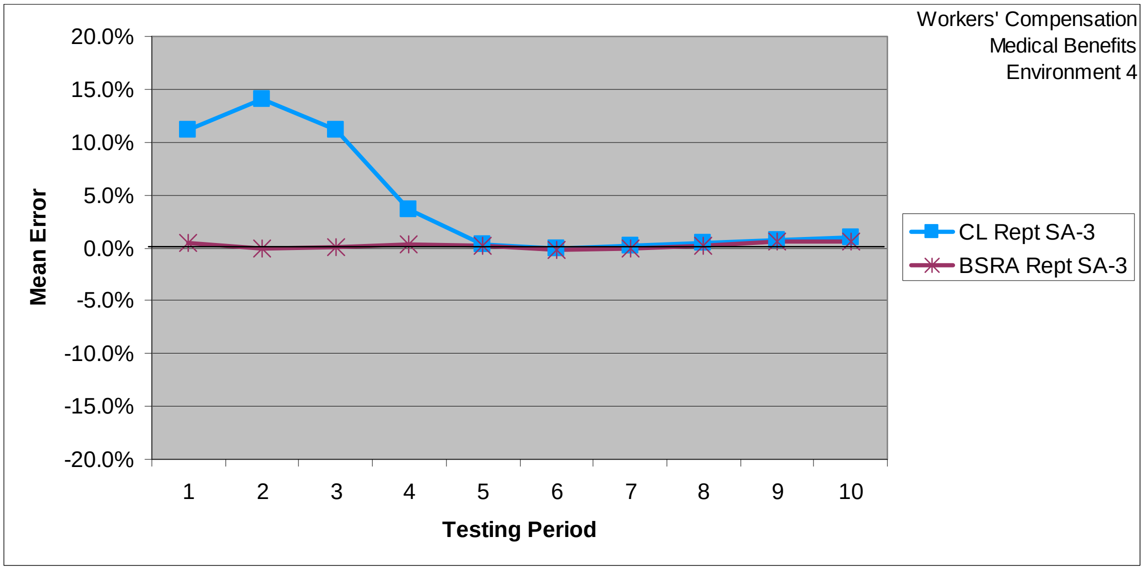


Figure 15: Comparison of the CL and BSRA method during a permanent change in case reserve adequacy.

The Berquist-Sherman adjustment for changes in claim settlement rates (BSCS) and the Fleming-Mayer adjustment for changes in claim settlement rates (FMCS) perform better than the chain ladder on paid loss when there are changes in settlement rates (see Figure 16). The BSCS and FMCS somewhat overreact to the change and underestimate the ultimate loss after the first testing period; neither method perfectly corrects for the change. This is perhaps evidence that in the presence of a change in the rate of claim settlement, the best estimate of ultimate loss lies somewhere between the BSCS and FMCS methods and the traditional CL method, which often in these situations are biased in opposite directions. Also shown is the BSRA, which beats the chain ladder but still overestimates the ultimate loss.

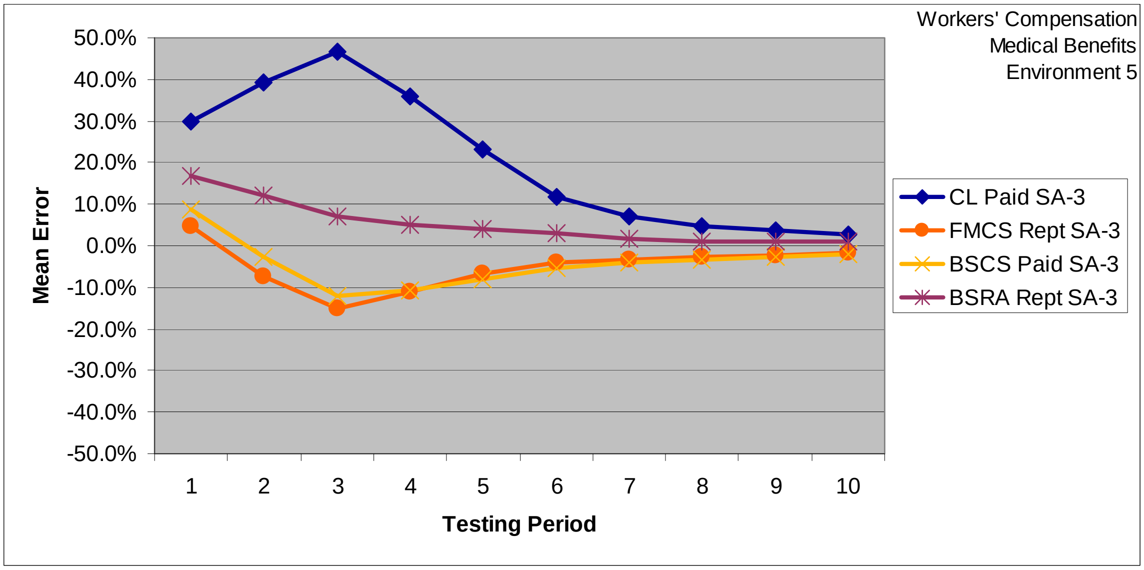


Figure 16: Comparison of the CL method with the Berquist-Sherman adjustments during a permanent acceleration in claim settlement rates to a higher plateau.

Figure 17 compares these methods during an increase in the frequency of serious injuries (environment 3). This change in the mix of claim types results in a change in claim settlement rates as well as a change in case reserve adequacy, although these changes manifest from one accident year to the next, and do not affect historical accident years. Similar to the previous example, the adjusted methods perform better than the chain ladder as they are able to correct somewhat for the environmental change. This is an interesting result, because it suggests that it may be worthwhile to incorporate methods that use Berquist-Sherman adjustments even if the actuary does not have a strong reason to believe that there has been a significant change in case reserve adequacy or claim settlement rates.[15]

[15] This conclusion warrants further investigation, because it is also possible that these methods may overreact to noise in the data that is not indicative of changing case reserve adequacy of claim settlement rates.

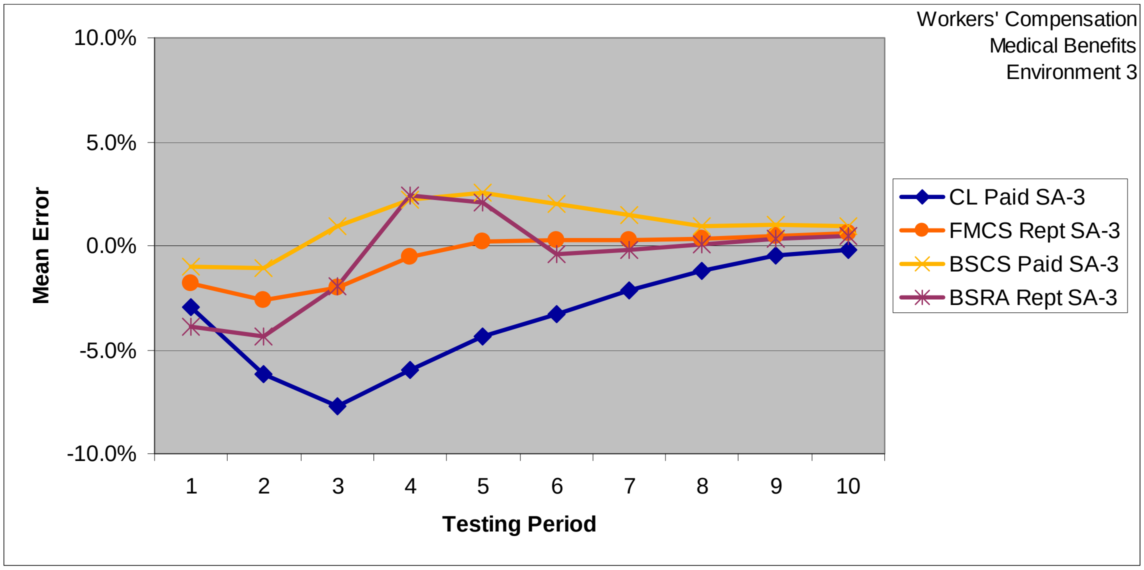


Figure 17: Comparison of the CL method with the Berquist-Sherman adjustments during a permanent increase in the frequency of serious injuries.

4.3.5 Case reserve methods

As would be expected, case reserve methods are adversely affected during a change in case reserve adequacy. Figure 18 compares Atkinson's case development (CD) method with the modified case development (MCD) method and Marker and Mohl's backwards recursive case development (MM) method during a permanent increase in case reserve adequacy. Methods that use case reserves as the projection basis are more adversely affected than the CL method on reported loss because they lack the ballast provided by paid amounts, which are unaffected by changes in case reserves. The MM method is the most adversely affected because the distortion in case reserves not only distorts future predictions of case reserves, but it also distorts future predictions of payments based on projected case reserves (i.e., the error is compounded in the iterative projections of paid and case loss).

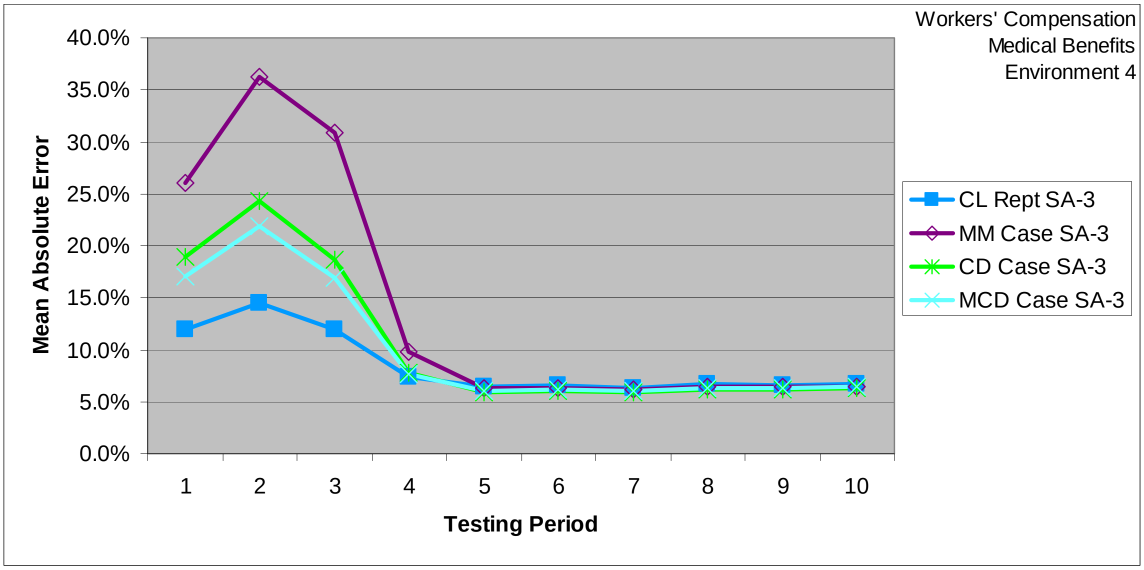


Figure 18: Comparison of case reserve methods during a permanent increase in case reserve adequacy.

What is not as obvious is that case reserve methods perform exceptionally well when there are similar distortions in both paid and reported triangles. Consider Figure 19, which compares the case reserve methods during a permanent increase in the frequency of serious injuries (environment 3). The CD and MCD methods project case reserves (which are unaffected in this environment) based on a function of the reported loss pattern and paid loss pattern. Both these patterns are lengthened due to the increase in frequency of serious injuries. However, the case development method is really only interested in the relative difference between the paid and reported pattern, not the nominal patterns. And since this difference is relatively unchanged in this environment, the CD and MCD methods are relatively unaffected.

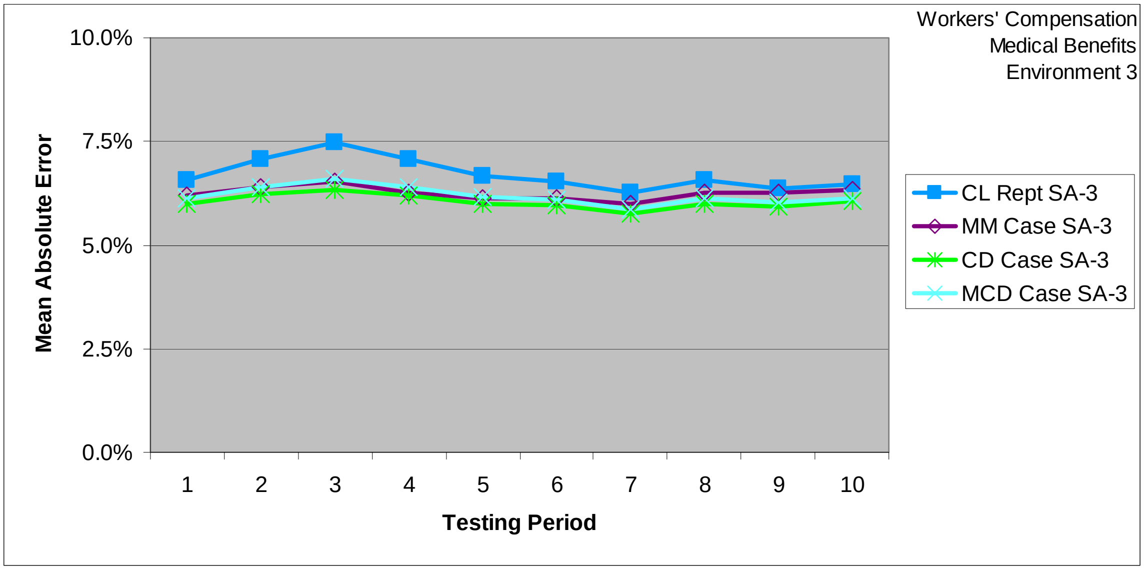


Figure 19: Comparison of case reserve methods during a permanent increase in the frequency of serious injuries.

The MM method is also relatively unaffected because an increase in the frequency of serious injuries drives up both the paid and case incremental severities as serious claims are more costly than average. However, because MM successively applies paid-on-prior case and case-on-prior case ratios, it is not as distorted since both numerator and denominator decrease at reasonably similar rates.[16] And the projection base, case reserves, adjusts to post-change levels more quickly than paid loss.

However, if reported loss patterns are significantly more distorted than paid loss patterns (or vice versa), then the case reserve methods will be distorted. Figure 20 illustrates this phenomenon by comparing the case reserve methods during an acceleration in claim settlement rates. Coupled with this acceleration in claim settlement rates is an increase in the average case reserve as the claims that remain open are the larger, more complex cases.

[16] This is an aspect of this environment, and may not apply in all situations with an increase in the frequency of large claims.

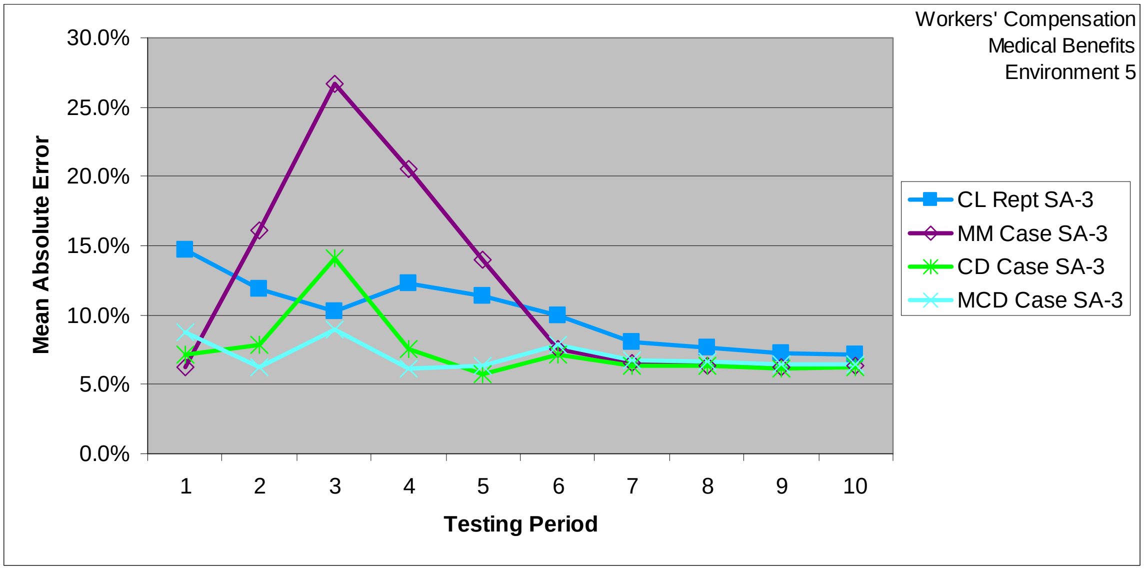


Figure 20: Comparison of case reserve methods during a permanent acceleration in claim settlement rates.

Finally, note that in Figure 18, Figure 19, and Figure 20 the MCD method performs marginally better than the CD method over all testing periods. This is because the MCD method incorporates information about the amount of loss paid to date, which the CD method ignores.

4.3.6 Joint paid-reported methods

The Munich chain ladder (MCL) method produces indications of ultimate loss that are nearly identical whether they are based on reported or paid loss amounts (see Figure 21, Figure 22, and Figure 23). However, those indications are often significantly less accurate than either the CL on paid loss or the CL on reported loss. There are several reasons for this phenomenon. The most obvious reason is that any distortion in loss development, whether it affects paid development (i.e., change in claim settlement rates) or reported development (i.e., change in case reserve adequacy) will always be captured as the MCL models paid and reported amounts simultaneously. For example, consider Figure 21, which compares the MCL and CL methods during a permanent increase in case reserve adequacy.

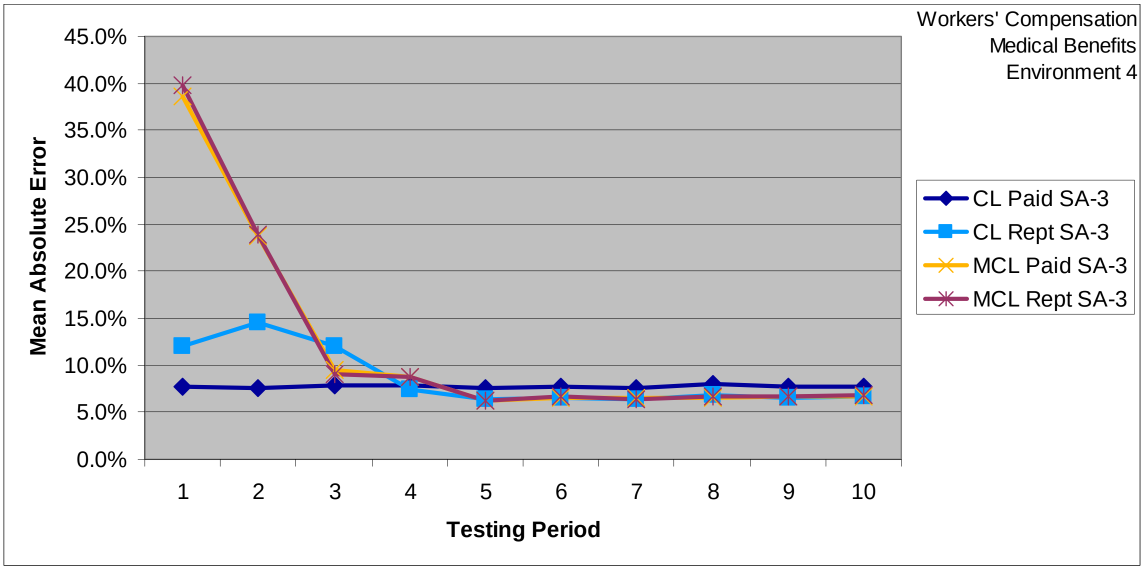


Figure 21: Comparison of the MCL and CL methods during a permanent increase in case reserve adequacy.

Furthermore, the Munich chain ladder appears to magnify distortions to paid development, as any distortion in paid development is implicitly reflected in reported development. To see this, consider Figure 22 where the error in the MCL is close to the combined error of the individual CL methods.

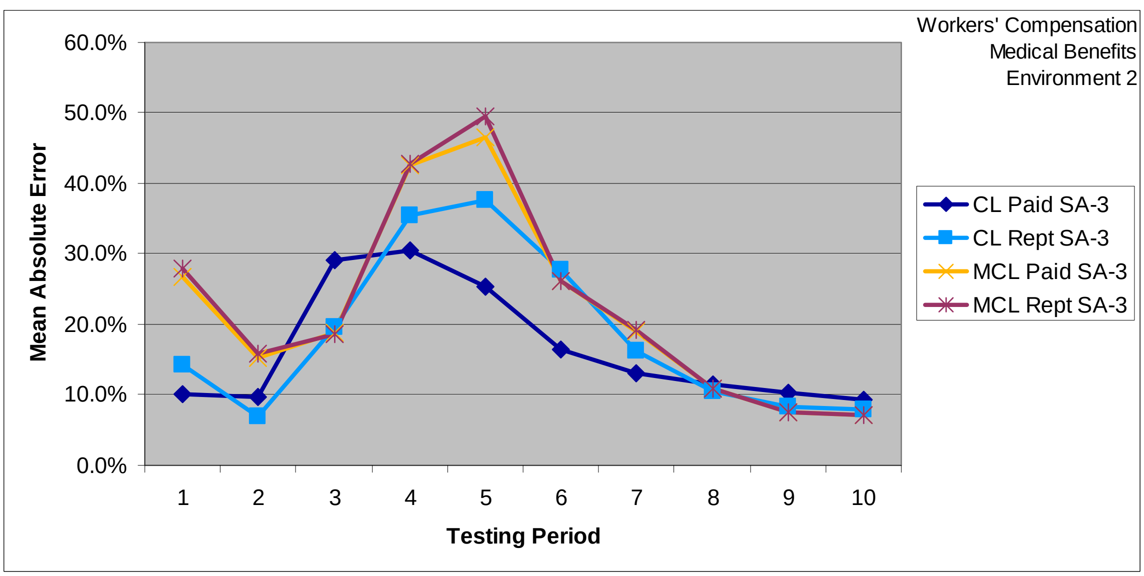


Figure 22: Comparison of the MCL and CL methods during a bubble in the rate of medical inflation.

Finally, note that in situations where there are no severe environmental distortions, although there is still residual noise, the MCL will produce more accurate paid and reported projections. Generally speaking, the MCL is useful to smooth out small distortions in paid and/or reported development that are not expected to be indicative of a larger shift in development. Figure 23 illustrates this phenomenon during a permanent acceleration in claim settlement rates. After the initial shock, the MCL responds much quicker than the CL on paid loss and is the most accurate method for testing periods 4 and subsequent (i.e., periods where there is no significant environmental distortion).

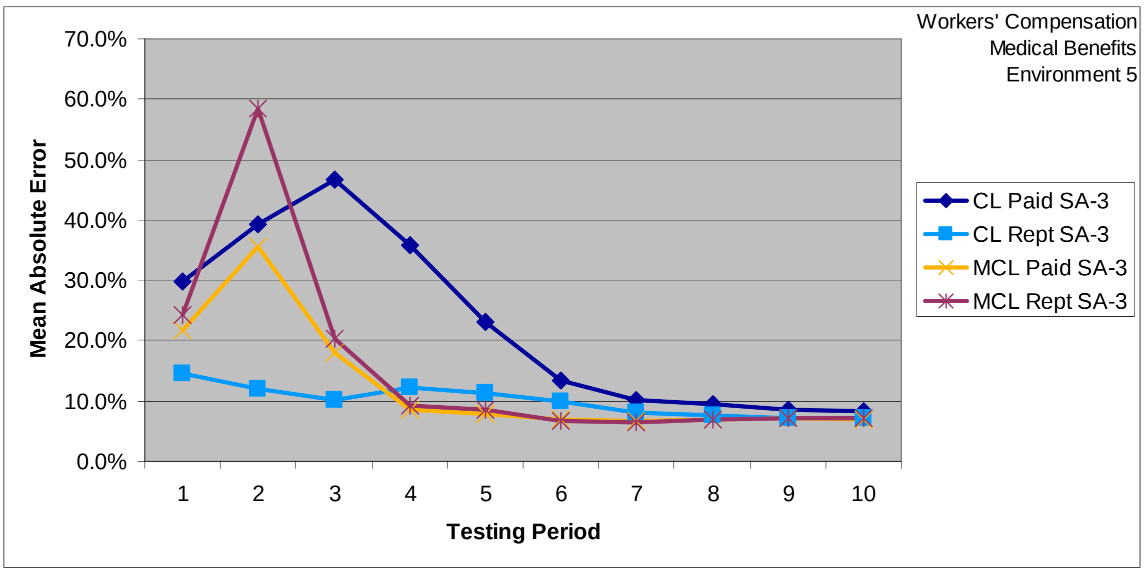


Figure 23: Comparison of the MCL and CL methods during a permanent acceleration in claim settlement rates.

4.4 Findings by Environment

4.4.1 Base environment (environment 1)

In the base environment, both the historical and future loss development patterns are stable, and the only source of error is residual noise. All methods perform similarly well (i.e., minimal mean errors), but no method is completely accurate, because of noise. The extent to which they differ shows their susceptibility to noise. For example, the chain ladder applied to paid loss is generally less accurate than the chain ladder applied to reported loss. This is not surprising, because paid loss development factors from age 12 to ultimate are significantly greater than reported LDFs, and larger factors leave more room for residual noise to distort development patterns. Not shown here are the exposure-based BL and BF methods. These, not surprisingly, are among the most accurate (close to

5% mean absolute error) because they are stabilized by giving weight to an unbiased a priori estimate of ultimate loss. Figure 24 shows these results.

The test results from the base environment illustrate the lowest potential level of mean absolute error possible for these methods under the residual noise levels assumed by our proxy data. Thus, for the rest of the environments, we should not expect to see levels of mean absolute error lower than we see here.

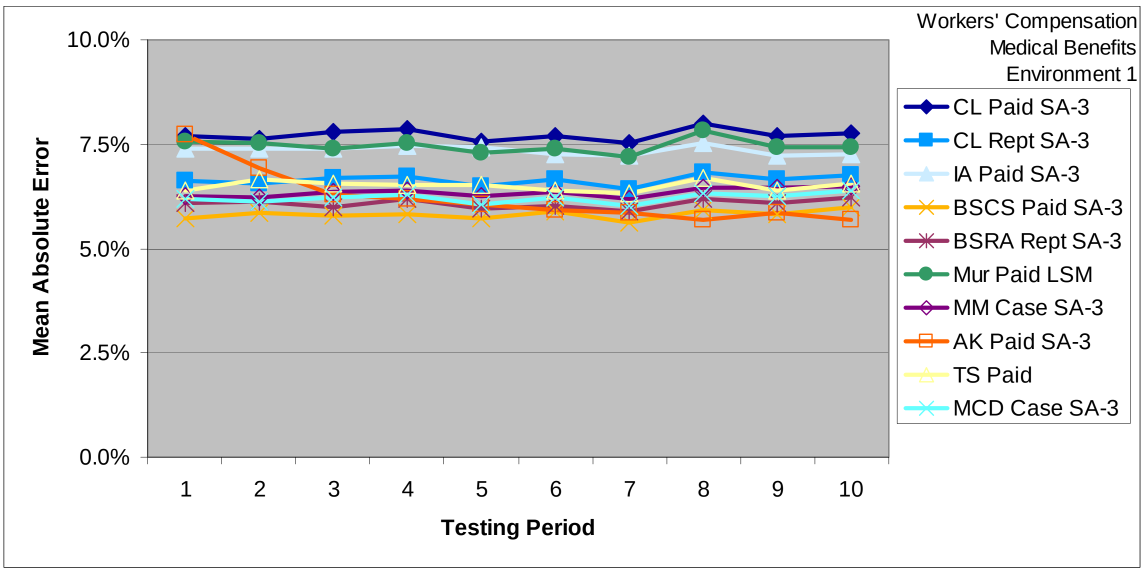


Figure 24: Comparison of various methods in the base environment.

4.4.2 Bubble in the rate of medical inflation (environment 2)

Figure 25 shows a comparison of various methods during a three-year bubble in medical inflation. Here paid severities are immediately affected; however, case severities respond more slowly as claims adjusters account for this new information. What is apparent is that methods based on case reserves or reported loss, such as the CL, MM, and MCD methods, show a significant delayed distortion, due to the lagged effect of higher inflation on case reserves that was assumed in the proxy data. Methods that perform well during the bubble include Murphy's least squares multiplicative (Mur-LSM) and Taylor's separation method (TS), possibly because these methods rely on more data points and are therefore more stable. The IA method performs poorly during the bubble, but it responds quickly after the rate of inflation reverts to its historical level, because the method relies on incremental payments and a short-term trend rate. If the actuary believes that recent inflation rates are likely to continue, then the IA with a short-term trend may be a good method. If the actuary believes that future inflation will be similar to historical long-term inflation, then the IA with a long-term trend may be preferable. Note that this environment is one of many where the data do not reveal the true nature of the change (i.e., specifically that it is only temporary rather than permanent) and other sources should be used to make an informed actuarial judgment.

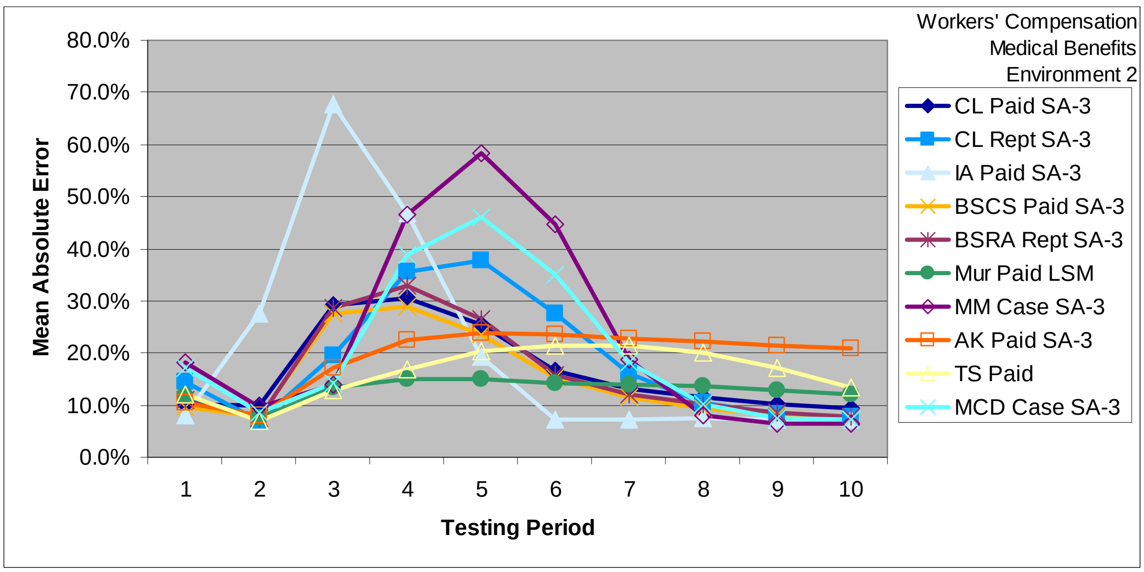


Figure 25: Comparison of various methods during a bubble in medical inflation.

4.4.3 Increase in the frequency of serious injuries (environment 3)

Environment 3 consists of a permanent increase in the frequency of serious injuries. Although claim severities are unchanged within each injury type, there are increases in ultimate claim counts, ultimate average severities, and ultimate losses.

Figure 26 isolates the mean error of the Adler-Kline claims closure model (AK) and the chain ladder method (CL). Both methods underestimate ultimate loss with the error gradually shrinking back to 0% error in the years after the change is complete. The CL underestimates because the increase in the frequency of serious injuries distorts the payment pattern, as serious claims report and close much slower than typical claims in our simulated data.

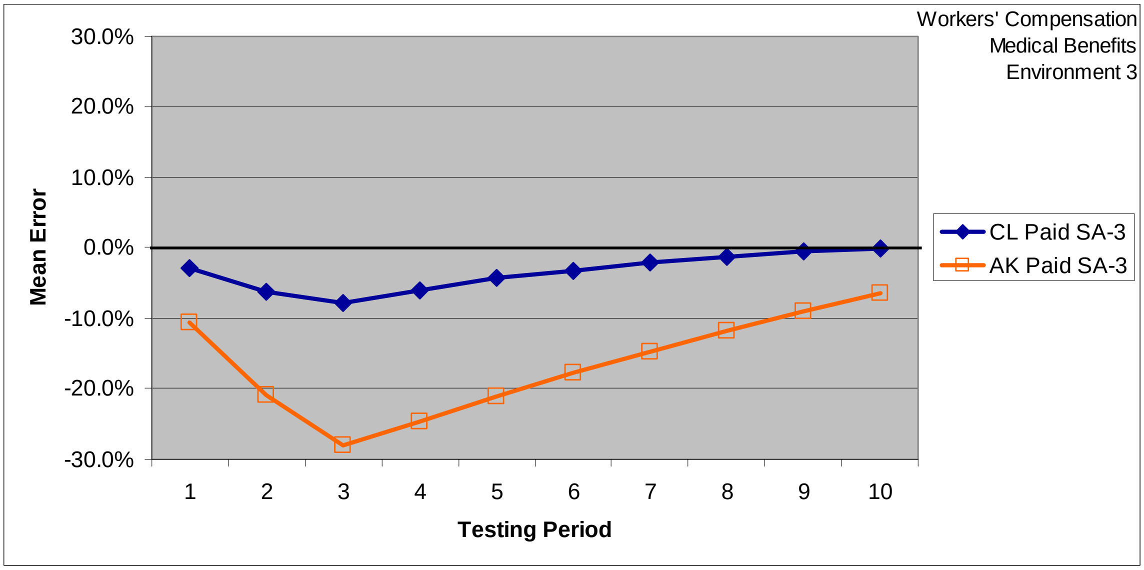


Figure 26: Comparison using the mean error statistic of the Adler-Kline claims closure model with the chain ladder method during and after an increase in the frequency of serious injuries.

To understand the additional error in the Adler-Kline method, note first that it relies on projecting incremental closed claim counts based on estimates of ultimate claim counts. Therefore, the underestimation in ultimate claim counts leads to an underestimation of future incremental closed claim counts. This problem is further exacerbated by applying the slowdown in claims closure pattern. The AK method interprets this slowdown in claims closure incorrectly, and allocates too many of the projected ultimate claims to earlier evaluation periods and too few to mature evaluation ages. And since loss severities are generally smaller at earlier periods and larger at later evaluation periods, the AK method further underestimates ultimate loss.

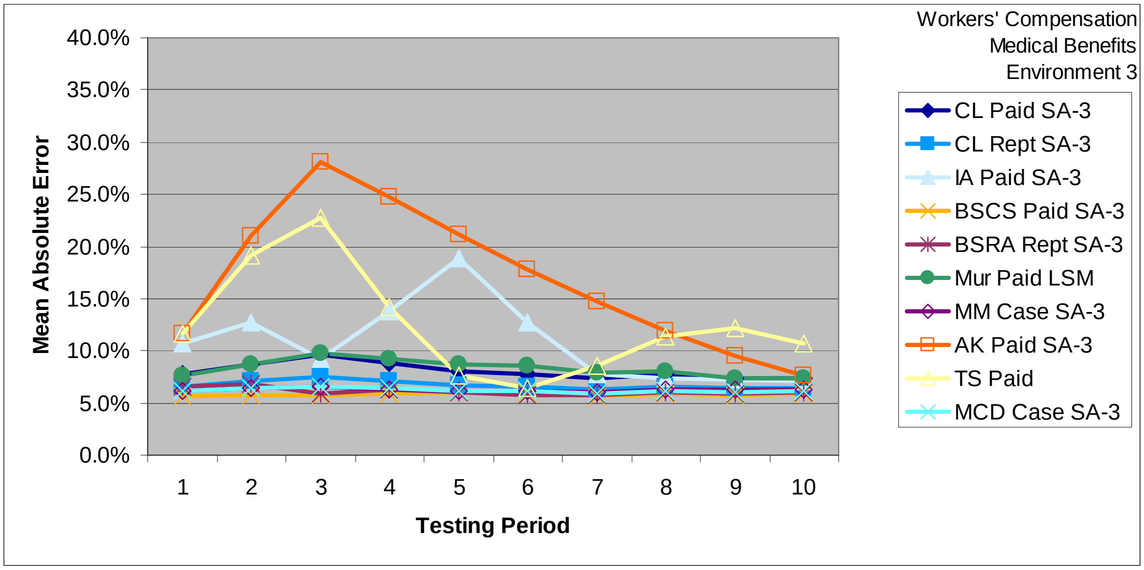


Figure 27: Comparison of various methods during a permanent increase in the frequency of serious injuries.

In Figure 27, we can see that the best performers in this scenario are the MM method, the MCD method, the BSCS and the BSRA method. As mentioned previously, the methods based on case reserves (MM and MCD) perform well in this environment, because although the loss payment and loss reporting patterns are distorted, there is less of a distortion in case reserve development.

It is interesting that both Berquist-Sherman adjustments, although intended to adjust for calendar year effects, perform well in this environment, in which there is a change in the mix of claim types from one accident year to the next. The BSRA method performs well as the increase in the frequency of serious injuries effectively mimics a change in case reserve adequacy, at least for the most recent accident years, and the BSRA method is able to immediately adjust to this new level of reserving. The BSCS method also performs well as the increase in the frequency of serious injuries effectively mimics a change in claim settlement rate that the BSCS method is able to model.

4.4.4 Increase in case reserve adequacy (environment 4)

Figure 28 shows the results for environment 4, in which case reserve adequacy permanently increases to a higher plateau, although paid and ultimate losses remain unchanged. Because only case reserves are affected in this environment, methods that do not incorporate case reserves are unaffected, and so we have excluded most of them from the graph. However, the best-performing method is the BSRA method, which, although it relies on case reserves, is able to correct for this misleading change by restating the historical triangle at the latest year of case reserve adequacy.

Furthermore, as the BSRA method smoothes the historical data, it adds an additional layer of stability minimizing the long-run error relative to other methods such as the chain-ladder approach.

Here the worst-performing method is the MM method since the distortion in case reserves not only distorts future predictions of case reserves, but it distorts future predictions of payments based on projected case reserves (i.e., the error is compounded in the iterative projections of paid and case loss).

The CL method on reported loss and the MCD method are also affected. The MCD method is more adversely affected since it applies the computed development factors to case reserves in isolation rather than to reported loss, which is somewhat stabilized by the paid component of reported loss.

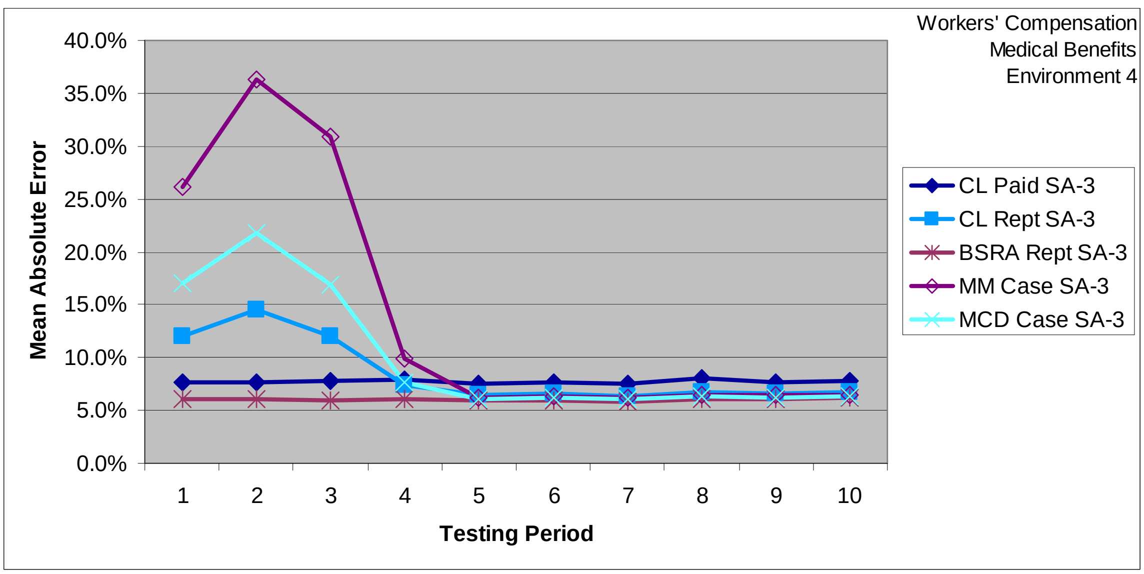


Figure 28: Comparison of various methods based on reported loss and/or case reserves during an environment of permanent increasing case reserve adequacy without an associated change in loss payments. The chain ladder on paid loss method is included as a base.

4.4.5 Acceleration in claim settlement rates (environment 5)

In this environment (see Figure 29), although there is a permanent acceleration in claim settlement rates, there is no change in the ultimate frequency or severity of claims. The FS method, which does not recognize this acceleration, overestimates the ultimate claim severity. In contrast, the AK method actually does quite well in this environment: since it models future incremental closed claims as a function of both ultimate counts and prior closed claims, it adequately responds to the acceleration in claim settlement rates.

In addition to the permanent acceleration in claim settlement rates, there is also a permanent increase in the average outstanding amounts (i.e., the claims that remain open are the more costly cases). This somewhat distorts those methods based on case reserve including the MM method and the CL method on reported loss. The MM method is more seriously distorted as it iteratively projects both case and paid amounts with paid amounts being very much affected in this environment (the case reserve development pattern is expected to change in this scenario). However, the MCD method is actually more accurate than the CL method on reported loss as it is able to adjust to the changing levels of case reserves. In addition to the MCD method, among the most accurate methods are the two Berquist-Sherman methods that are able to adjust for the several distortions in this environment.

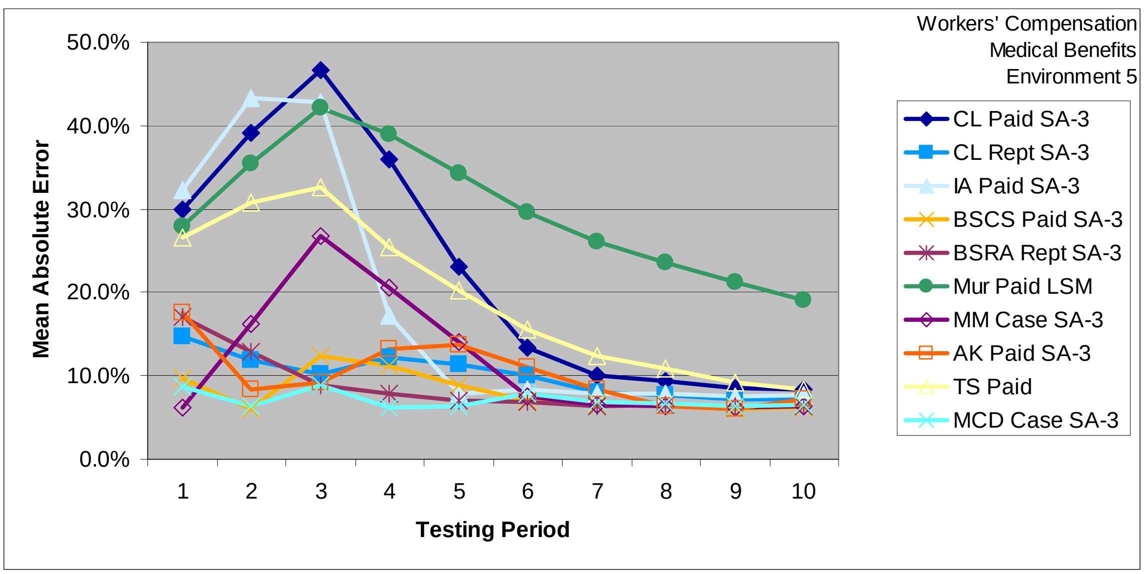


Figure 29: Comparison of various methods during a permanent acceleration in claim settlement rates.

4.4.6 Increase in case reserve adequacy with an acceleration in claim settlement rates (environment 6)

This environment is a combination of environment 4, an increase in case reserve adequacy, and environment 5, an acceleration in claim settlement rates.

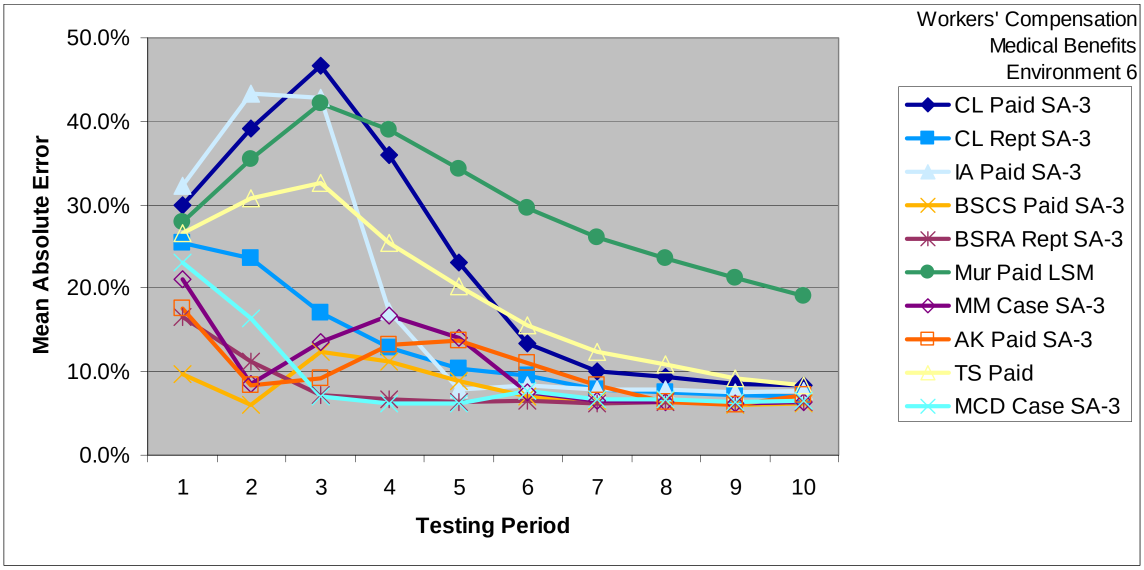


Figure 30: Comparison of various methods during a period of increasing case reserve adequacy, which plateaus at a permanently higher level coupled with a permanent acceleration in claim settlement rates.

As Figure 30 shows, the permanent change in claim settlement rates is the more significant of the two distortions, and so the results of environment 6 are similar to those of environment 5. As a result, the least accurate methods are those based on unadjusted paid loss. The Berquist-Sherman adjustments and the AK method are among the most accurate methods. Interestingly, the MM method is more accurate here than in either environment 4 or 5, because the biases created by the two environments offset each other. As shown in Figure 31, the higher average case reserves of environment 4 cause MM to overestimate, and the faster claim settlement rate of environment 5 cause MM to underestimate.

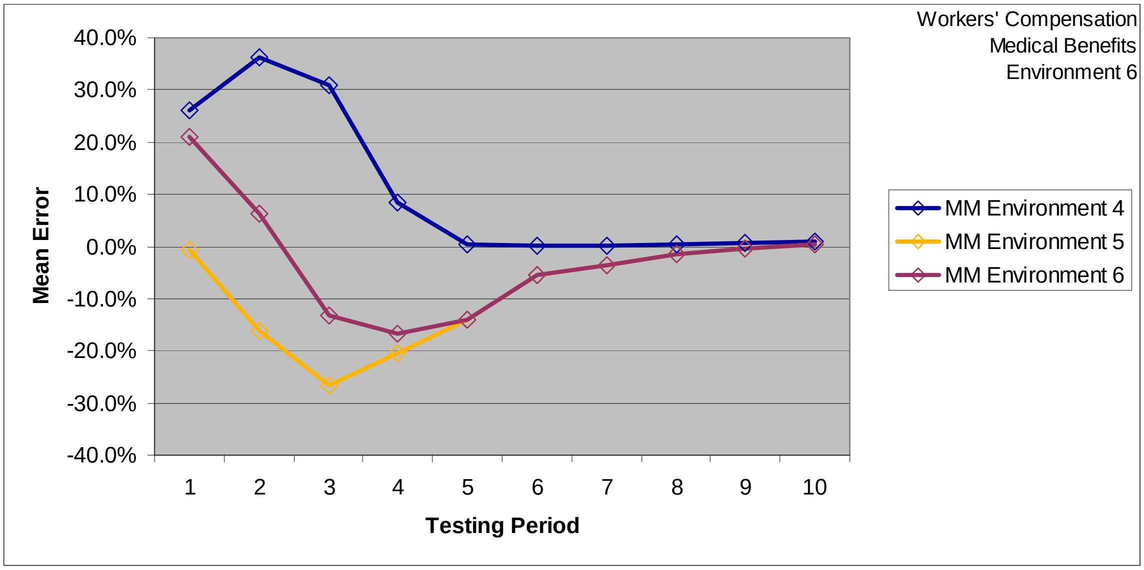


Figure 31: Comparison of various methods during a period of increasing case reserve adequacy, which plateaus at a permanent higher level coupled with a permanent acceleration in claim settlement rates.

4.4.7 Bubble in the rate of medical inflation with an increase in the frequency of serious injuries (environment 7)

This environment is a combination of environments 2 and 3. Here the IA and AK methods are distorted by the change in mix of claim types, and most other methods are distorted by the inflation bubble. One method that performs well during the period of change is Murphy's least squares multiplicative model. The BSCS and BSRA methods perform well in the first two testing periods, prior to being distorted by the inflation bubble.

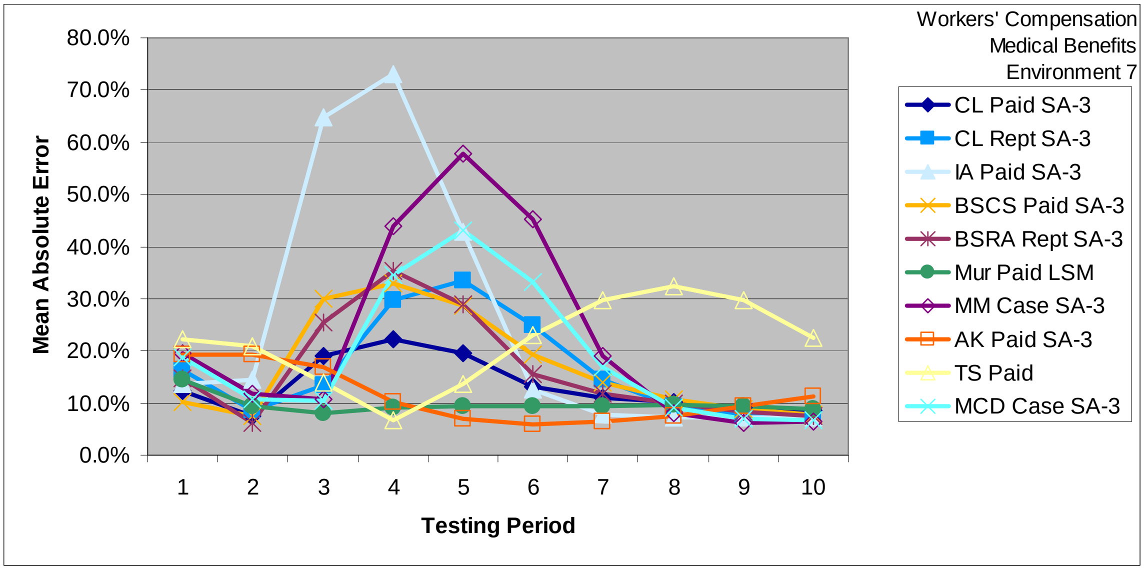


Figure 32: Comparison of various methods during a bubble in the rate of medical inflation coupled with an increase in the frequency of serious injuries.

4.4.8 Change in loss ratio (environment 8)

Figure 33 shows the mean error of various methods during and following a sharp, permanent change in the accident year loss ratio with no change in the mix of claims or the claims reporting, closing, payment, or reporting patterns. This simple environment, where loss doubles relative to exposures/premiums, provides insight into many of the traditional actuarial techniques. Note that the CL method is unaffected as this accident year shift has no effect on loss development patterns. The BL method is the worst-performing as it completely ignores loss information and relies entirely on the a priori estimate. The BF method performs slightly better than the BL method as it incorporates current loss amounts that reflect the higher level of loss. The BT method does slightly better as it gives twice as much weight to current loss amounts.

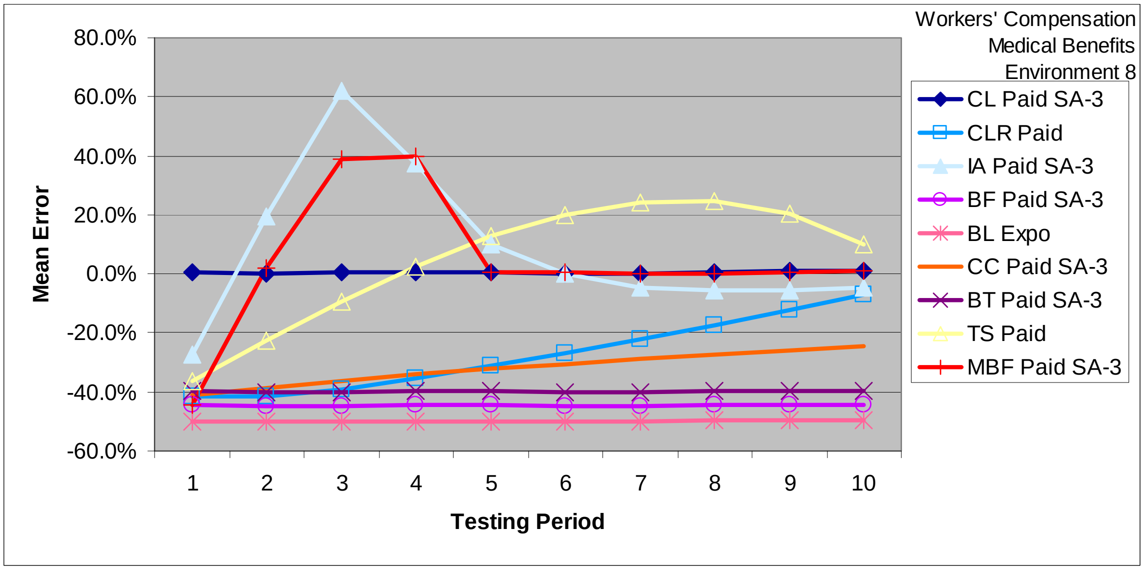


Figure 33: Comparison of various methods following a permanent shift in loss ratio.

The Cape Cod (CC) method is also based on exposures, and as a result it underestimates ultimate loss. Unlike the BL, BF, and BT methods, it does respond to changing conditions by re-evaluating the expected loss ratio. The response rate is gradual, because the expected loss ratio has been parameterized based on a long-term average.

The other methods shown (CLR, IA, TS, and MBF) incorporate loss trend, but in different ways. The modified Bornhuetter-Ferguson method initially matches the BF method, but in subsequent periods the expected loss ratio in the MBF is adjusted based on the observed loss trend in the latest four accident years. Once the accident year loss trend stabilizes, the MBF uses an accurate expected loss ratio and produces unbiased results. In this environment, a shorter-term trend rate would respond even more quickly. The incremental additive method relies on the loss trend observed based on payments in the latest four calendar years, and so the results are a bit worse than the MBF, because this environment is characterized by an accident year change. The CLR projects losses forward by fitting a trend line to columns of historical incremental paid losses, with trend rates calculated separately by evaluation age. Because the CLR uses long-term trend rates, it responds more gradually than the MBF or IA. Finally, Taylor's separation (TS) method uses a blended trend (based on both calendar year and accident year components), and as a result it responds more quickly than the CLR but more slowly than the MBF or IA.

5. CONCLUSION

Our observations and recommendations are intended to guide the actuary in evaluating the strengths and weaknesses of available loss reserving methods. While it is impossible to produce a fool-proof instruction guide for selecting actuarial methodologies, we envision that the practicing actuary will consider our recommendations within the context of the actuarial control cycle.

At the start of this control cycle, the actuary reviews a carefully selected set of diagnostics and leading indicators compiled to assist with detection and interpretation of trends in the system. The review of diagnostics and leading indicators should be accompanied by insights from other sources such as claims administrators, underwriters, and advisory organizations, in order to guide the identification of characteristics of the current and expected future environment. Although it may not be possible to pinpoint the environment, the actuary may be able to narrow down the possibilities and assess the volatility or level of noise in the underlying data.

After identifying characteristics of the environment and volatility of the data, the actuary can use general or specific observations from our analysis or similar work—either directly or by extrapolation of the conclusions contained within these sources—to identify suitable loss reserving methodologies. While we expect that the actuary's focus will first be on the expected accuracy of the various methods in the environment at the time (the priority of this paper), it is also important the actuary consider the relative importance of other criteria, such as bias, stability, and responsiveness.

Finally, the actuary should review the projections both in the short term using actual versus expected comparisons and in the longer term using hindsight testing. Actual versus expected analyses will allow the actuary to make minor corrections to optimize performance of the selected method (or methods). Hindsight testing allows the actuary to identify and subsequently correct for any systematic biases present in the data, the loss reserving methods considered, or the actuary's assumptions.

We hope that using the results of this paper, in the greater framework of the actuarial control cycle, will lead not only to more accurate projections of ultimate loss, but also help increase credibility of the actuarial profession by increasing documentation and arguments for selection of one methodology over another.

Acknowledgment

The authors would like to acknowledge Dave Bellusci for the helpful discussions about the intricacies of California workers compensation during the initial project that led to this paper; Bob Conger for his work as advisor and ensuring that our focus remained on the important rather than the nuanced; Tim Gault for the many hours he spent analyzing and refining the technical models underlying the results of this paper; Anne McKneally for her helpful edits to the many drafts of this manuscript thereby greatly aiding the readability of the final product; Alejandra Nolibos for her painstaking peer review and thoughtful reminders; and reviewers Jon Michelson and Jessica Leong for their thorough review and helpful commentary and suggestions.

Appendix A – Loss Reserving Methods

The following contains descriptions of various loss reserving methodologies considered and comments about their overall performance. The two-to-three letter parenthetical abbreviations are identifiers used to reference that specific method in our analysis.

A.1 Adler-Kline Claims Closure Method (AK)

The Adler-Kline claims closure method is a frequency-and-severity model where projected incremental closed claim counts are multiplied by projected incremental paid on incremental closed claim count severities. First, incremental closed claim counts are computed using disposal ratios; the disposal ratio is defined as the ratio of incremental closed claims to open claim counts. Open claim counts are developed by projecting reported claim counts to ultimate using the chain ladder method and then subtracting cumulative closed claim counts. Incremental closed claim counts are then projected by iteratively multiplying open claims by the disposal ratio and then updating the number of claims still open. Incremental paid on closed claim count severities are trended forward at each evaluation period. See Adler & Kline [1]. This method is typically applied in lines of business for which claims are reported rather quickly, such as coverages written on a claims-made basis.

A.2 Bornhuetter-Ferguson Method (BF)

The Bornhuetter-Ferguson method computes the outstanding loss as the product of the percentage of loss outstanding and an initial expected loss estimate. It sums this amount with the current cumulative loss amount to produce an estimate of ultimate loss. The initial expected loss estimate is the product of the a priori loss ratio with exposures. In our parameterization, the a priori loss ratio is equal to the loss ratio observed prior to the start of first environmental change. See Bornhuetter and Ferguson [4].

A.3 Budgeted Loss Method (BL)

The budgeted loss method computes ultimate loss as the product of an a priori loss ratio with exposures. In our parameterization, the a priori loss ratio is equal to the loss ratio observed prior to the start of first environmental change. See Brosius [5].

A.4 Berquist-Sherman Adjustment for Change in Claim Settlement Rate (BSCS)

This method adjusts actual experience to a common level of claim settlement speed first by computing claims closure ratios, i.e., the ratios of closed claims to ultimate claims (computed by developing reported claim counts to ultimate). Then, the latest diagonal of claims closure ratios is assumed for all historical diagonals. Adjusted paid loss on closed claim count severities are computed using log-linear interpolation between the actual paid loss on closed claim count severities and the actual claims closure ratio. This is to find the implied paid loss severity on closed claims associated with the new claims closure ratio. See Berquist and Sherman [3].

A.5 Berquist-Sherman Adjustment for Change in Case Reserve Adequacy (BSRA)

This method adjusts actual historical case reserves to a common level of reserve adequacy by de-trending the latest diagonal of the triangle of average case reserves per open claim by the trend in the average severity of paid loss on closed claims. See Berquist and Sherman [3].

A.6 Benktander Method (BT)

The Benktander method is a variant of the Bornhuetter-Ferguson method where instead of using the budgeted loss method as a prior, a credibility-weighted sum of the budgeted loss method with the chain ladder method is used. See Mack [11].

A.7 Stanard-Bühlmann or Cape-Cod Method (CC)

The Cape-Cod method is a variant of the Bornhuetter-Ferguson method where the a priori loss ratio is computed as the simple average of the historical loss ratios. The historical loss ratios are computed as the ratio of the latest diagonal of loss to used-up exposure. Used-up exposure is the product of exposure and the percent of loss developed for the year. See Friedland [7].

A.8 Atkinson Case Reserve Development (CD)

The chain ladder on case reserves method, as described by Atkinson [2], works by first selecting reported and paid loss development factors. Then, using the relationship between case reserves and paid and reported loss, the actuary derives case development factors from the paid and reported development factors. These case development factors are then applied to case reserves to project ultimate loss.

A.9 Chain Ladder Method (CL)

The chain ladder method we used is the traditional loss development method, whereby the change in cumulative loss from age to age is used to project the latest diagonal of the cumulative loss triangle.

A.10 Bühlmann's Complementary Loss Ratio Method (CLR)

Bühlmann's complementary loss ratio method computes incremental payments at each evaluation period by trending forward historical incremental payments at similar evaluation periods. These are then summed to provide an estimate of cumulative ultimate loss. See Pentikäinen and Rantala [17].

A.11 Fisher-Lange Claims Closure Method (FL)

Fisher and Lange's claims closure model is very similar to the Adler-Kline claims closure model. However, Fisher and Lange project incremental closed claims by using closure ratios that are the ratio of incremental closed claims to ultimate claims, where ultimate claims are computed by applying the chain ladder method to reported claim counts (as opposed to Adler and Kline who use disposal ratios that are the ratio of incremental closed claims to open claims). Incremental paid on incremental closed severities are trended forward at each exposure period. The product of the projected incremental closed claim counts with the projected incremental paid on incremental closed claim count severities then produces the reserve estimate. See Fisher and Lange [6]. Fisher and Lange advocate using report-year data because there is no development on reported claim counts in this situation, obviating the need to project ultimate claim counts. However, the method can equally apply to accident-year data by developing reported claim counts to ultimate.

A.12 Fleming-Mayer Adjustment for Change in Claim Settlement Rate (FMCS)

The Fleming-Mayer Adjustment for change in claim settlement rates (FMCS) is similar to the Berquist-Sherman adjustment for claim settlement rate (BSCS) in that it adjusts for changes in claim settlement rate. However, the FMCS applies to reported loss rather than paid loss. The paid component in the reported loss amounts are adjusted as they are in the BSCS; however, the case component is also adjusted, in a similar way as the paid component, to reflect the fact that changes in claim settlement often have a ripple effect onto outstanding case amounts. See Fleming and Mayer [7].

A.13 Cumulative Frequency-Severity Method (FS)

The basic frequency-severity approach included here works by projecting claim counts (frequency) to ultimate and projecting loss on claim count (severity) to ultimate using the chain ladder method. The product of these two then produces an estimate of ultimate loss. We have included both the "reported claim count/reported loss on reported claim count" variant as well as the "closed claim count/paid loss on closed claim count" variant. We choose this cumulative frequency-and-severity approach to contrast it with the various other incremental approaches included in our analysis including Fisher-Lange and Adler-Kline.

A.14 Ghezzi's Incremental Closed Claim Severity Method (GH1)

Future incremental closed claim counts are projected by applying the percentage of claims closed pattern to open claims. To compute open claims, ultimate claim counts are projected by applying the chain ladder method to reported claim counts. Future incremental paid loss on incremental closed claim count severities are computed by trending forward historical incremental paid on incremental closed claim count severities using exponential growth. To produce an estimate of outstanding loss, the actuary multiplies the vector of projected incremental paid on incremental closed claim count severities with the vector of projected incremental closed claims. The dot-product of these amounts produces an estimate of reserves. The key to this method (as well as Ghezzi's Ultimate Unclosed Claim Severity Method) is that only ratios prior to the significant environmental change are considered. See Ghezzi [9].

A.15 Ghezzi's Ultimate Unclosed Claim Severity Method (GH2)

Unpaid loss amounts are computed by estimating preliminary ultimate loss amounts using the chain ladder method on either paid or reported loss amounts (we used reported loss); and subtracting paid amounts. Similarly, unclosed claims are computed by estimating ultimate claim counts using the chain ladder method on either closed or reported claim counts (we used reported claim counts). Ghezzi's ultimate unclosed claim severity method then works by trending the ratios of unpaid loss to unclosed claims using exponential growth. The loss reserve is then computed by multiplying these trended ratios by current unclosed claim counts. The key to this method (as well as Ghezzi's incremental closed claim severity method above) is that only ratios prior to the significant environmental change are considered. See Ghezzi [9].

A.16 Incremental Additive Method (IA)

The incremental additive method uses both the triangle of incremental losses and the exposure vector for each accident year as a base. Incremental additive ratios are computed by taking the ratio of incremental loss to the exposure (which has been adjusted for the measurable effect of inflation), for each accident year. This gives the amount of incremental loss in each year and at each age expressed as a percentage of exposure, which we then use to square the triangle.

A.17 Incremental Multiplicative Method (IM)

The incremental multiplicative method uses incremental loss to compute incremental loss development factors, sometimes referred to as "decay ratios," which are defined to be the ratio of incremental loss at a later age to the incremental loss at an earlier age. With these loss development factors it is possible to extrapolate future incremental payments as means of squaring the triangle.

A.18 Verrall's Log-Linear Methods (LL1, LL2, LL3)

Pentikäinen and Rantala [17] include three log-linear regression models with varying numbers of parameters in their analysis. For comparison purposes, we have chosen to include these models, as described on page 184 of the above. For greater detail, we refer the reader to Verrall [25]. We refer to these three models as LL1, LL2 and LL3, respectively, based on the order of presentation in Pentikäinen and Rantala [17].

A.19 Brosius's Least Squares Development (LS)

Least squares development as described by Brosius iteratively regressed ultimate loss on cumulative loss at successive maturity starting with the maturity one evaluation period prior to ultimate. Each successive regression produces one more estimate of ultimate loss that is used in the next regression. See Brosius [5].

A.20 Modified Bornhuetter-Ferguson (MBF)

The modified Bornhuetter-Ferguson method is identical to the Bornhuetter-Ferguson method except that the initial expected loss estimate is the average of the prior years' ultimates (rather than the product of an a priori loss ratio with exposures). We have trended the initial expected loss estimate to adjust for growth in exposures. See Pentikäinen and Rantala [17].

A.21 Munich Chain Ladder (MCL)

The Munich chain ladder method we implemented is identical to the basic method described in Quarg and Mack [18]. However, we allowed the initial selection of parameters to be based on simple and volume-weighted averages of less than all years (e.g., volume-weighted average of latest 3 sets of observed data points).

A.22 Marker-Mohl Backwards Recursive Case Development Method (MM)

The backwards recursive case development method works by first computing the percentage of loss paid (to case reserves in the previous period) and the percentage of case reserves (to case reserves in the previous period) at each age. These are then iteratively applied to case reserves so as to produce case reserves and paid losses at each age. See Marker and Mohl [13]. This method is typically applied in lines of business for which claims are reported rather quickly, such as coverages written on a claims-made basis.

A.23 Murphy's Family of Chain Ladder Parameterizations (Mur)

As part of his 1994 work, Daniel Murphy outlines five possible parameterizations of the chain ladder method. We abbreviate them as follows: (1) LSL – least squares linear, (2) LSM – least squares multiplicative, (3) SA – simple-average development, (4) VW – weighted-average development, and (5) GA – geometric-average development. See Murphy [14] for exact solutions of each parameterization. The last three of these are equivalent to a chain ladder method using various types of averages of all historical observations.

A.24 Multivariate Regression Development Method (MV)

A great variety of multivariate regression models is suggested in the literature. While it would be extremely difficult to include all varying combinations of dependent and independent variables, we have, as a proxy, included the multivariate regression model. In this model, cumulative loss is regressed on case reserves and cumulative payments in the prior period to develop an estimate of loss development factors. These are then applied successively to the latest diagonal of loss (a la the chain ladder method) in order to project ultimate loss.

A.25 Taylor's Separation Method (TS)

Taylor's separation method (TS) attempts to "separate" the calendar year inflation effect from the evaluation period development effect. Our implementation is similar to that in Taylor [23].

A.26 Weller's Algebraic Method (WA)

Weller's algebraic method (WA) describes the claims reserve triangle as a system of linear equations involving various unknown parameters represented the percentage paid or reported at various evaluations. These linear equations can be iteratively solved to establish development factors that can be used to project ultimate losses. See Weller [26].

Appendix B – Loss Reserving Method Families

LOSS RESERVING METHODS CLASSIFIED BY FAMILY

Family	Methods
Exposure-Based Methods	(1) Budgeted Loss Method (BL) (2) Bornhuetter-Ferguson Method (BF) (3) Modified Bornhuetter-Ferguson Method (MBF) (4) Benktander Method (BT) (5) Cape-Cod Method (CC)
Regression Methods	(1) Least Squares Development (LS) (2) Murphy's *Least Squares Linear* Parameterization (Mur-LSL) (3) Murphy's *Least Squares Multiplicative* Parameterization (Mur-LSM) (4) Multivariate Regression (MV) (5) Verrall's Log-Linear Models (LL1, LL2, LL3)
Frequency-Severity Methods	(1) Adler-Kline Claims Closure Model (AK) (2) Fisher-Lange Claims Closure Model (FL) (3) Ghezzi's Incremental Closed Claim Severity Method (GH1) (4) Ghezzi's Ultimate Unclosed Claim Severity Method (GH2) (5) Cumulative Frequency-Severity Method (FS)
Case-Reserve Methods	(1) Marker-Mohl Backwards Recursive Case Development (MM) (2) Atkinson Case Development (CD) (3) Modified Atkinson Case Development (MCD)
Incremental Methods	(1) Incremental Multiplicative Method (IM) (2) Incremental Additive Method (IA) (3) Bühlmann's Complementary Loss Ratio Method (CLR)
Joint Paid-Reported Models	(1) Munich Chain Ladder (MCL)
Berquist-Sherman Adjustments	(1) Berquist-Sherman adjustment for case reserve adequacy (BSRA) (2) Berquist-Sherman adjustment for claim settlement rate (BSCS) (3) Fleming-Mayer adjustment for claim settlement rate (FMCS)
Miscellaneous	(1) Taylor's Separation Method (TS) (2) Weller's Algebraic Method (WA)

Appendix C – Environments

The following appendix contains descriptions of the various environments we considered. Each of these environments describes one or two changes which are common to the workers compensation line of business (as well as many other lines). However, the reader should note that these constructed environments are to some degree simplifications of the real world, which would include a combination of many of these changes in tandem.

C.1 Base environment (environment 1)

In the base scenario, exposures grow gradually at 1% from one accident year to the next, and claim frequency is constant. Claim reporting speed, claim payment speed, and claim closure speed are each consistent over time. Claim payments increase gradually with inflation, with case reserves moving in tandem (cost inflation is assumed at 5% per annum for medical on a calendar-year basis). The resulting claims and loss development patterns generally align with recent workers' compensation medical loss development patterns in California.

C.2 Bubble in the rate of medical inflation (environment 2)

In this environment, calendar-year medical inflation is 15% in the first three testing periods (as compared to 5% historically). For the fourth and subsequent testing periods, calendar-year medical inflation returns to the original 5% level. The changes apply consistently to all medical costs, independent of injury type. A practical example of this scenario would be that of runaway medical costs that are subsequently tamed by the implementation of treatment guidelines. Paid losses immediately reflect the increase and subsequent drop in medical inflation rates, whereas case reserves respond more gradually, lagging by three periods. Claim reporting speed and claim closure speed are unchanged from the base environment.

C.3 Increase in the frequency of serious injuries (environment 3)

In relation to the base environment, this environment features an approximate 75% increase in the frequency of serious claims, which occurs evenly over the course of three accident years (corresponding with the first three testing periods). Thereafter, the claim frequency remains at this elevated level. Severities for each injury type, patterns by injury type and the frequency of other injury types are unaffected.

C.4 Increase in case reserve adequacy (environment 4)

In this scenario, case reserve adequacy increases relative to the base scenario. The change occurs over the course of two calendar years (corresponding with the first two testing periods), after which case reserve adequacy remains at the higher level. This change applies consistently to all injury types. It does not affect any ultimate levels or the rate of payments or claim closures.

C.5 Acceleration in claim settlement rate (environment 5)

In this environment, the speed at which claims are settled increases. This change causes claims to be paid and closed earlier than in the base environment. The earlier closure of claims results in fewer payments at later ages, resulting in reduced ultimate losses compared to the base environment. The change applies to all injury types. In addition to speedier claim settlements, there is an increase in the incremental paid amounts per claim at each stage of development. Also, as the claims being settled are likely to be the less serious of the open claims, there are also small increases in the observed average case reserve per open claim (without any change in case reserving adequacy). The acceleration in claim settlement rates and the corresponding changes in severities take place over three years (corresponding with the first three testing periods). In the fourth year, claim settlement rates stabilize at rates observed in the third year (but still higher than historical norms). Similarly, paid severities and average case reserves on open severities stabilize at levels observed in the third year.

C.6 Increase in case reserve adequacy with an acceleration in claim settlement rates (environment 6)

This environment is a combination of the above fourth environment (permanent increase in case reserve adequacy) with the fifth environment (permanent acceleration in claim settlement rates). In addition to more claims being settled sooner with a higher average outstanding case severities (i.e., environment 5), case adjusters overreact to this shift and permanently over-reserve on the few large claims that remain open.

C.7 Bubble in the rate of medical inflation with an increase in the frequency of serious injuries (environment 7)

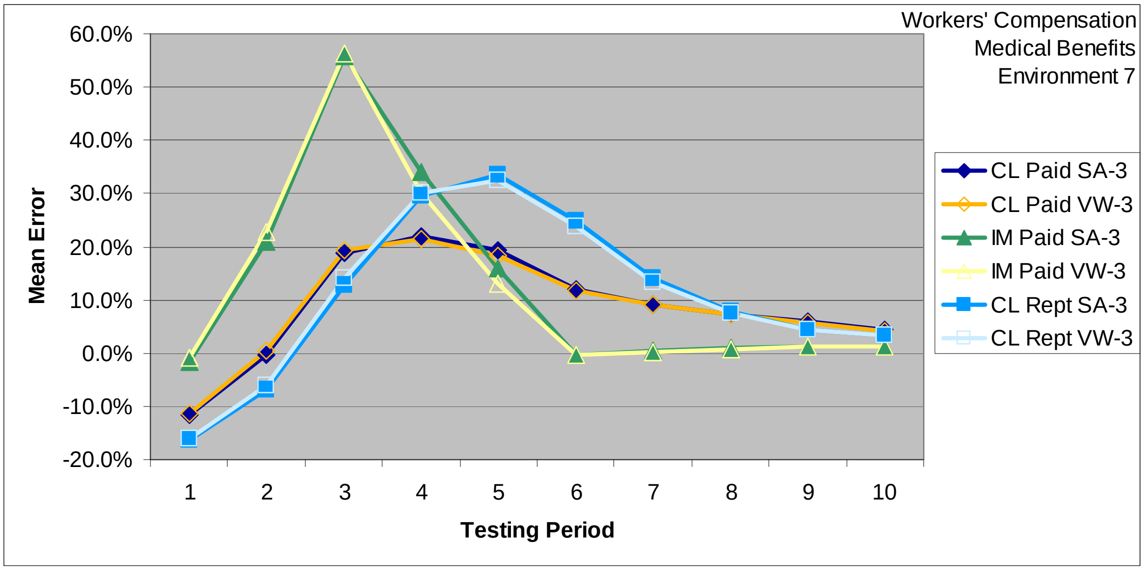
This environment is a combination of the above second environment (bubble in the rate of medical inflation) with the third environment (permanent increase in the frequency of serious injuries).

C.8 Change in loss ratio (environment 8)

In this environment, claim counts and losses increase relative to premium/exposures yet severities remain unchanged. This change occurs suddenly in the first testing period, and losses remain at the elevated level in subsequent testing periods. This simple environment models a change in loss ratio (i.e., as is present in loss ratio cycles) and is used to highlight (i) the effect of accident year changes that do not affect loss development patterns and (ii) the how exposure-based methods are adversely affected in the absence of correct a priori loss estimates.

Appendix D – Results graph interpretation

For the most part, we have used the same graph to present results. The following describes the various components of this graph and how it should be read.



(1) Y-axis: this will reflect either the "Mean Error" (i.e., measuring bias and accuracy) or "Mean Absolute Error" (i.e., measuring accuracy). Error is defined as the projected ultimate loss for the latest accident year (i.e., from age 12 to ultimate) minus the actual ultimate loss, expressed as a percentage of the actual ultimate loss. If a method overestimates an ultimate loss of 100 by 10, then the error would be 10%. Values close to 0 are optimal; values far from 0 indicate distortions.

(2) X-axis: testing period 1 represents the latest accident year one year after the start of the environmental change; testing period 2 represents the latest accident year two years after the start of the environmental change, and so on. So, if testing period 1 is accident year XX at 12/31/XX (at which point it is 12-months-old), then testing period 2 is accident year XX+1 at 12/31/XX+1 (at which point it is also 12-months-old). Most environmental changes take place over the first four testing periods with the data stabilizing in the fifth and subsequent periods. Each testing period measures the same test statistic, which is the error from age 12 to ultimate (i.e., the latest accident year).

(3) Right header: the right header provides three pieces of information. The first line indicates the line of business (i.e., workers compensation). The second line indicates the type of data (i.e., medical benefits). The third line indicates the environment being tested.

(4) Legend: the legend maps the lines with each method. Each method is described by four pieces of information. The first two-to-four letters indicate the loss reserving method (i.e., CL indicates the Chain Ladder method and IM is the incremental multiplicative method). The next word indicates to what data the method applies (i.e., paid, reported, case reserves, or exposure data). The next few letters describe the parameterization (i.e., SA indicates a simple average, VW indicates a volume-weighted average, and the number identifies the number of calendar years of loss development factors used to parameterize the loss reserving model.

(5) Reading the graph: ideally, we are looking for methods that perform well during the change (i.e., small errors in testing periods 1-3), but also important are responsive methods that quickly self-correct after the change (i.e., sharply sloped lines in testing periods 4-7), methods that are relatively stable throughout the testing period (i.e., flat lines in testing periods 1-10), and methods that are biased in opposite directions (i.e., lines that show mirror image mean errors above and below 0).

Appendix E – Accuracy Report Card

Grading of the methods' average accuracy over the second, third, and fourth testing periods in various environments based on a bell curve, with grade thresholds selected judgmentally (A is best; F is worst).

				Environment							
Family	Abbr.	Data	Param.	1	2	3	4	5	6	7	8
Chain-Ladder	CL	Paid	SA-3	D	C	C	C	F	F	B	B
Chain-Ladder	CL	Reported	SA-3	C	C	B	C	C	C	B	A
Exposure	BL	Exposure	N/A	A	C	F	A	C	C	F	F
Exposure	BF	Paid	SA-3	A	C	F	A	C	C	D	F
Exposure	BT	Paid	SA-3	A	B	D	A	C	C	D	F
Exposure	CC	Paid	SA-3	A	A	D	A	C	C	D	F
Exposure	MBF	Paid	SA-3	C	C	C	B	F	F	C	D
Exposure	BF	Reported	SA-3	A	B	D	A	C	C	D	F
Exposure	BT	Reported	SA-3	A	A	D	B	C	C	C	D
Exposure	CC	Reported	SA-3	A	A	D	B	C	C	C	D
Exposure	MBF	Reported	SA-3	B	C	C	C	C	C	C	D
Frequency-Severity	AK	Paid	SA-3	C	B	D	B	B	B	B	A
Frequency-Severity	FL	Paid	SA-3	C	B	D	B	B	B	B	A
Frequency-Severity	GH1	Paid	SA-3	F	C	D	D	C	C	D	C
Frequency-Severity	GH2	Paid	SA-3	B	A	D	B	A	A	C	A
Frequency-Severity	FS	Paid	SA-3	D	C	C	C	F	F	B	B
Frequency-Severity	FS	Reported	SA-3	C	C	B	C	C	C	B	A
Berquist-Sherman	BSCS	Paid	SA-3	B	C	A	A	B	B	C	A
Berquist-Sherman	BSRA	Reported	SA-3	B	C	A	A	B	A	C	A
Berquist-Sherman	FMCS	Reported	SA-3	B	C	A	C	C	B	C	A
Case	CD	Case	SA-3	B	C	A	F	B	A	C	A
Case	MCD	Case	SA-3	B	C	A	F	A	B	C	A
Case	MM	Case	SA-3	C	C	A	F	C	B	C	A
Incremental	CLR	Paid	N/A	F	C	D	C	D	D	C	F
Incremental	IA	Paid	SA-3	C	F	C	C	D	D	F	F
Incremental	IM	Paid	SA-3	D	F	C	C	D	D	D	B
Incremental	CLR	Reported	N/A	F	D	D	C	C	C	B	D
Incremental	IA	Reported	SA-3	C	F	C	D	B	B	F	F
Incremental	IM	Reported	SA-3	C	F	B	D	B	B	F	B
Joint Paid-Reported	MCL	Paid	SA-3	C	D	B	D	C	F	C	A
Joint Paid-Reported	MCL	Reported	SA-3	C	D	B	D	D	F	C	A
Regression	LL1	Paid	LS-All	D	B	C	C	F	F	A	B
Regression	LL2	Paid	LS-All	C	B	D	B	D	D	B	D
Regression	LL3	Paid	LS-All	F	C	F	F	A	A	D	F
Regression	LS	Paid	LS-All	D	F	C	C	F	F	F	C
Regression	Mur	Paid	LSL-All	D	D	C	C	F	F	C	B
Regression	Mur	Paid	LSM-All	C	B	C	C	F	F	A	B
Regression	MV	Paid	LS-All	C	D	B	C	C	C	C	B
Regression	LL1	Reported	LS-All	C	B	B	C	B	C	A	A
Regression	LL2	Reported	LS-All	B	C	C	C	B	B	B	C
Regression	LL3	Reported	LS-All	F	B	F	D	C	B	C	D
Regression	LS	Reported	LS-All	D	F	C	D	C	C	F	C
Regression	Mur	Reported	LSL-All	C	D	B	C	C	C	C	B
Regression	Mur	Reported	LSM-All	C	A	B	C	C	C	A	A
Regression	MV	Reported	LS-All	C	D	A	D	B	B	C	A
Miscellaneous	WA	Paid	All	D	C	C	C	D	D	D	D
Miscellaneous	TS	Paid	All	C	B	C	B	D	D	B	C
Miscellaneous	WA	Reported	All	D	D	D	F	C	D	F	F
Miscellaneous	TS	Reported	All	B	B	C	C	C	C	A	C

Appendix F – Bias Report Card

Methods' average bias categorized based on mean error over the second, third, and fourth testing periods (H+: 15% or greater, H: between 5% and 15%, U: between -5% and 5%, L: between -15% and -5%, L-: below -15%).

				Environment							
Family	Abbr.	Data	Param.	1	2	3	4	5	6	7	8
Chain-Ladder	CL	Paid	SA-3	U	H+	L	U	H+	H+	H	U
Chain-Ladder	CL	Reported	SA-3	U	H+	U	H	H	H+	H	U
Exposure	BL	Exposure	N/A	U	L-	L-	U	H+	H+	L-	L-
Exposure	BF	Paid	SA-3	U	L-	L-	U	H+	H+	L-	L-
Exposure	BT	Paid	SA-3	U	L-	L-	U	H+	H+	L-	L-
Exposure	CC	Paid	SA-3	U	L	L-	U	H+	H+	L-	L-
Exposure	MBF	Paid	SA-3	U	H+	L	U	H+	H+	H	H+
Exposure	BF	Reported	SA-3	U	L-	L-	U	H+	H+	L-	L-
Exposure	BT	Reported	SA-3	U	L	L-	U	H	H+	L-	L-
Exposure	CC	Reported	SA-3	U	L	L-	U	H	H+	L-	L-
Exposure	MBF	Reported	SA-3	U	H+	U	H	H	H+	H	H+
Frequency-Severity	AK	Paid	SA-3	U	H	L-	U	L	L	L-	U
Frequency-Severity	FL	Paid	SA-3	U	H+	L-	U	U	U	L	U
Frequency-Severity	GH1	Paid	SA-3	U	L-	L-	U	H	H	L-	U
Frequency-Severity	GH2	Paid	SA-3	U	U	L-	U	U	H	L-	U
Frequency-Severity	FS	Paid	SA-3	U	H+	L	U	H+	H+	H	U
Frequency-Severity	FS	Reported	SA-3	U	H+	U	H	H	H+	H	U
Berquist-Sherman	BSCS	Paid	SA-3	U	H+	U	U	L	L	H+	U
Berquist-Sherman	BSRA	Reported	SA-3	U	H+	U	U	H	H	H+	U
Berquist-Sherman	FMCS	Reported	SA-3	U	H+	U	H	L	L	H+	U
Case	CD	Case	SA-3	U	H	U	H+	L	U	H	U
Case	MCD	Case	SA-3	U	H+	U	H	U	H	H	U
Case	MM	Case	SA-3	U	H+	U	H+	L-	L	H	U
Incremental	CLR	Paid	N/A	U	H+	L-	U	H+	H+	L-	L-
Incremental	IA	Paid	SA-3	U	H+	U	U	H+	H+	H+	H+
Incremental	IM	Paid	SA-3	U	H+	L	U	H+	H+	H+	U
Incremental	CLR	Reported	N/A	U	H+	L-	H	H+	H+	L	L-
Incremental	IA	Reported	SA-3	U	H+	H	H	H	H	H+	H+
Incremental	IM	Reported	SA-3	U	H+	U	U	U	U	H+	U
Joint Paid-Reported	MCL	Paid	SA-3	U	H+	U	H	L	L-	H	U
Joint Paid-Reported	MCL	Reported	SA-3	U	H+	U	H	L-	L-	H	U
Regression	LL1	Paid	LS-All	U	H	L	U	H+	H+	H	U
Regression	LL2	Paid	LS-All	U	H	L-	U	H+	H+	L-	L-
Regression	LL3	Paid	LS-All	L-	L-	L-	L-	U	U	L-	L-
Regression	LS	Paid	LS-All	U	H+	L	U	H+	H+	H+	U
Regression	Mur	Paid	LSL-All	U	H+	L	U	H+	H+	H+	U
Regression	Mur	Paid	LSM-All	U	H	L	U	H+	H+	U	U
Regression	MV	Paid	LS-All	U	H+	U	H	H+	H+	H+	U
Regression	LL1	Reported	LS-All	U	H	L	H	H	H+	U	U
Regression	LL2	Reported	LS-All	U	H+	L-	H	H	H	L	L-
Regression	LL3	Reported	LS-All	L-	L	L-	L	L-	L	L-	L-
Regression	LS	Reported	LS-All	U	H+	U	H	H	H+	H+	U
Regression	Mur	Reported	LSL-All	U	H+	U	H	H	H+	H+	U
Regression	Mur	Reported	LSM-All	U	U	U	H	H	H+	U	U
Regression	MV	Reported	LS-All	U	H+	U	H	U	U	H+	U
Miscellaneous	WA	Paid	All	U	L-	L-	U	H+	H+	L-	L-
Miscellaneous	TS	Paid	All	U	H	L-	U	H+	H+	L	L
Miscellaneous	WA	Reported	All	U	L-	L-	H	H	H+	L-	L-
Miscellaneous	TS	Reported	All	U	H	L	H	H	H+	L	L

Appendix G – Abbreviations and notation

Miscellanea:
LDF, Loss development factor
SA, Simple average
SA-3, Simple average of the latest three observations
SA-All, Simple average of all historical observations
VW, Volume-weighted average
VW-3, Volume-weighted average of the latest three observations
WCIRB, Workers' Compensation Insurance Rating Bureau of California

Loss Reserving Methods:
AK, Adler-Kline claims closure method
AM, Weller's algebraic reserving method
BF, Bornhuetter-Ferguson method
BLM, budgeted loss method
BT, Benktander method
CC, Stanard-Bühlmann or Cape Cod method
CD, Atkinson chain ladder on case reserves method
CL, chain ladder method
CLR, Bühlmann's complementary loss ratio method
BSCS, Berquist-Sherman adjustment for the change in claim settlement rate method
BSRA, Berquist-Sherman's adjustment for reserve adequacy method
FL, Fisher-Lange claims closure method
FMCS, Fleming-Mayer adjustment for change in claim settlement rate
FS, cumulative frequency-severity method
GH1, Ghezzi's incremental closed claim severity method
GH2, Ghezzi's ultimate unclosed claim severity method
IA, incremental additive method
IM, incremental multiplicative method
LL1, Verrall's log-linear method #1
LL2, Verrall's log-linear method #2
LL3, Verrall's log-linear method #3
LS, Brosius's least squares development method
MBF, modified Bornhuetter-Ferguson method
MCD, modified Atkinson chain ladder on case reserves method
MCL, Munich chain ladder
MM, Marker & Mohl's backwards recursive case development method
MV, multivariate regression method
Mur-LSL, Murphy's family of parameterizations (least squares linear) method
Mur-LSM, Murphy's family of parameterizations (least squares multiplicative) method
TS, Taylor's separation method
WA, Weller's algebraic reserving method

6. REFERENCES

[1] Adler, Martin and Charles D. Kline, Jr., "Evaluating Bodily Injury Liabilities using a Claims Closure Model," CAS *Discussion Paper Program*, **1988**, 1-99.

[2] Atkinson, Rick, "Calculating IBNR based on Case Reserves," CAS *Forum*, Fall **1989**, 39-44.

[3] Berquist, James R. and Richard E. Sherman, "Loss Reserve Adequacy Testing: A Comprehensive Systematic Approach," *PCAS*, **1977**, Vol. LXVII, 123-184.

[4] Bornhuetter, Ronald L. and Ronald E. Ferguson, "The Actuary and IBNR," *PCAS*, **1972**, Vol. LIX, 181-195.

[5] Brosius, Eric, "Loss Development using Credibility," CAS *Exam Study Note*, **1993**, 1-19.

[6] Fisher, Wayne H. and Jeffrey T. Lange, "Loss Reserve Testing: A Report Year Approach," *PCAS*, **1973**, Vol. LX, 189-207.

[7] Fleming, Kirk G. and Jeffrey H. Mayer, "Adjusted Incurred Loss for Simultaneous Shifts in Payment Patterns and Case Reserve Adequacy Levels," CAS *Discussion Paper Program*, **1998**, 189-214.

[8] Friedland, Jacqueline F., "Estimating Unpaid Claims using Basic Techniques," CAS, **2009**.

[9] Ghezzi, Thomas L., "Loss Reserving without Loss Development Patterns: Beyond Berquist-Sherman," CAS *Forum*, Fall **2001**, 43-104.

[10] Jing, Yi, Joseph Lebens, and Stephen Lowe, "Claims Reserving: Performance Testing and the Control Cycle," *Variance*, **2009**, Vol. 3, pp. 161-193.

[11] Mack, Thomas, "Credible Claims Reserve: The Benktander Method," ASTIN *Bulletin*, **2000**, Vol. 30, No. 2, 333-347.

[12] Mahon, Mark J., "The Scorecard System," CAS *Forum*, Summer **1997**, 97-136.

[13] Marker, Joseph O. and James F. Mohl, "Rating Claims-Made Insurance Policies," CAS *Discussion Paper Program*, **1980**, 265-304.

[14] Murphy, Daniel, "Unbiased Loss Development Factors," *PCAS*, **1993**, Vol. LXXXI, 154-222.

[15] Narayan, Prakash and Thomas V. Warthen, "A Comparative Study of the Performance of Loss Reserving Methods through Simulation," CAS *Forum*, Summer **1997**, Vol. 1, 175-196.

[16] Peck, Edward F., "[Discussion] A Simulation Test of Prediction Errors of Loss Reserve Estimate Techniques," *PCAS*, **1995**, Vol. LXXXII, 104-120.

[17] Pentikäinen, Teivo and Jukka Rantala, "A Simulation Procedure for Comparing Different Claims Reserving Methods," CAS *Forum*, Fall **1995**, 128-156.

[18] Quarg, Gerhard and Thomas Mack, "Munich Chain Ladder: A Reserving Method that Reduces the Gap between IBNR Projections based on Paid Losses and IBNR Projections based on Incurred Losses," *Variance*, **2008**, Vol. 2, Issue 2, 266-299.

[19] Rollins, John W., "Performance Testing Aggregate and Structural Reserving Methods: A Simulation Approach," CAS *Forum*, Summer **1997**, 137-174.

[20] Schmidt, K.D., "A Bibliography on Loss Reserving." Available on: http://www.math.tu-dresden.de/sto/schmidt/dsvm/reserve.pdf.

[21] Skurnick, David, "A Survey of Loss Reserving Methods," *PCAS*, **1973**, Vol. LX, 16-58.

[22] Stanard, James N., "A Simulation Test of Prediction Errors of Loss Reserve Estimation Techniques," *PCAS*, **1985**, Vol. LXXII, 124-148.

[23] Taylor, Greg, "Separation of Inflation and Other Effects from the Distribution of Non-Life Insurance Claim Delays," *ASTIN Bulletin*, **1977**, Vol. 9, 219-230.

[24] Thorne, Joseph O., "Loss Reserve Adequacy: A Comprehensive Systematic Approach [Discussion]," *PCAS*, **1978**, Vol. LXV, 10-33.

[25] Verrall, Richard, "Statistical Methods for the Chain Ladder Technique," CAS *Forum*, Spring **1994**, 393-446.

[26] Weller, Alfred O., "An Algebraic Reserving Method for Paid Loss Data," CAS *Forum*, Fall **1995**, 255-280.

About the Authors

Tapio Boles is a consultant with the San Francisco office of Towers Watson. He is a Fellow of the Casualty Actuarial Society and a Member of the American Academy of Actuaries.

Andy Staudt is a consultant with the London office of Towers Watson. He is a Fellow of the Casualty Actuarial Society and a Member of the American Academy of Actuaries.

The authors can be reached at:

Tapio Boles: tapio.boles@towerswatson.com
Andy Staudt: andy.staudt@towerswatson.com