Looking Back to See Ahead: A Hindsight Analysis of Actuarial Reserving Methods

Susan J. Forray, FCAS, MAAA

Abstract

Thirty actuarial reserving methods are evaluated empirically against an extensive database of Schedule P data. The metric of method skill is used to evaluate the historical performance of the methods. Results are provided by company size and line of business. The effect of correlation on the usefulness of additional methods is considered. Results suggest the use of several methods not common in actuarial practice, as well as a refinement of weights typically assigned to the more common methods.

Keywords. reserving; reserving methods; management best estimate; suitability testing.

1. INTRODUCTION

Actuarial reserve analyses typically rely on a number of different estimation methods to develop indicated ultimate loss[1]. The paid and incurred (i.e., paid plus case) chain ladder methods are surely the most common. Other actuarial reserving methods[2] include the following:

- Backward Recursive
- Benktander
- Berquist-Sherman
- Bornhuetter-Ferguson
- Brosius
- Cape Cod
- Case Development Factor
- Claims Closure
- Frequency/Severity
- Hindsight Outstanding/IBNR
- Incremental Additive
- Incremental Multiplicative
- Loss Ratio
- Munich Chain Ladder

Of course most of these methods have both paid and incurred versions and many have several other variations as well.

Oftentimes, these methods diverge significantly, and actuarial judgment is used in selecting ultimate loss. A need exists for empirical evidence to support the use of particular methods over others.

[1] Within this monograph the term "loss" should be taken to refer either just to loss or more generally to loss and ALAE.
[2] Descriptions of these methods as used within the current analysis can be found in Appendix A.

1.1 Outline

The remainder of this paper proceeds as follows. Section 2 will provide an overview of the analysis, including the data available as well as a discussion of the metric. Section 3 will discuss the results of the analysis, including results by company size and line of business. Section 4 will discuss the effect of correlation between methods on the results and the practical implications of this correlation. Section 5 will provide additional discussion on the approach to the analysis, and, in particular, the metric selected. Lastly, Section 6 will offer some conclusive remarks.

2. OVERVIEW OF ANALYSIS

2.1 Data for Analysis

The current analysis relies on a large database of Schedule P[3] triangular data. Thirty methods, as listed in Appendix A, were applied to the triangular data given within Schedule P to develop indications of ultimate loss by coverage year and method for each line of business available for a given property & casualty insurance company. The most recent evaluation date available at the time of the analysis was December 31, 2010. Consequently this functioned as the date as of which "actual" ultimate loss would be determined. Indications based on data as of prior valuations were then evaluated against the ultimate loss as of this most recent valuation date. This is summarized in Table 1.

TABLE 1
VARIABLE COMBINATIONS FOR ULTIMATE LOSS DEVELOPMENT

Variables	Companies	Lines of Business	Methods	Evaluation Dates	Accident / Report Years
Description	P&C Writers	10-year triangle Schedule P lines	As Above, Including Variations	1996-2009	10 Years Preceding Evaluation Date
Approximate Number	3,100	16	30	14	10

[3] Schedule P is a section of the U.S. Statutory Annual Statement in which triangular data, including paid and case reserve loss and ALAE as well as closed and open claim counts, are reported. The current analysis uses only those lines of business for which 10-year triangles are provided in Schedule P.

Note that only individual companies were considered within this analysis. Company groups were excluded as their data would have overlapped with that of the individual writers. The indicated ultimate loss has been developed under an automated procedure for each of the above combinations for which data is available. Details of the calculations for each method are given in Appendix A.

In theory, the above combinations for which data is available could have resulted in as many as 208 million records of ultimate loss indications. In practice, most companies do not write many of the lines of business. Data for many of the companies is not available at all evaluation dates and, even when it is, the Schedule P triangles are occasionally found to be inconsistent at different evaluation dates and consequently unusable. Ultimate loss indications are available for 49 million of the above variable combinations.

2.2 Method Skill

The concept of method skill is a recent introduction to the actuarial profession, having first been discussed by Jing, Lebens, and Lowe in [8]. It will be discussed at a high level here. However, it may be useful for the reader to have further understanding of the calculation of method skill as applied to the analysis at hand. This is provided in Appendix B.

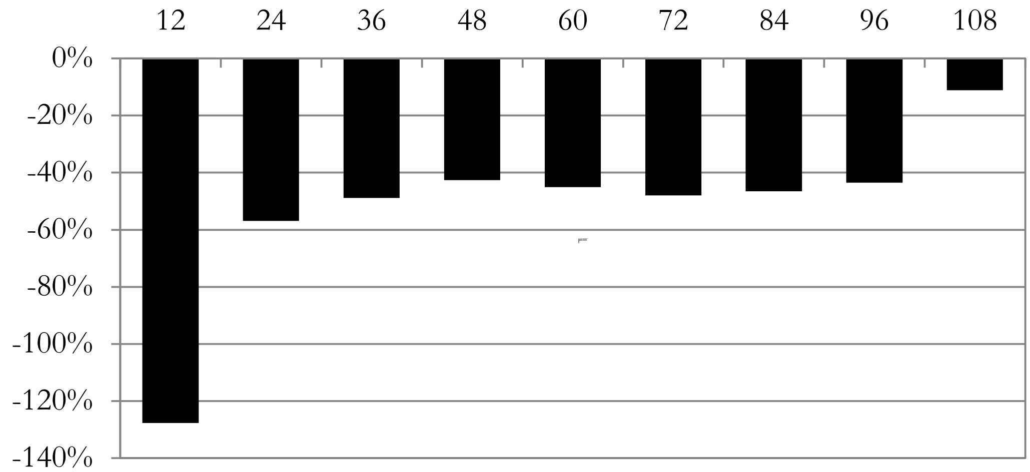
The skill of method m at development age d is calculated as

$$Skill_m^{(d)} = 1 - mse_m^{(d)} / msa^{(d)}$$

where $mse_m^{(d)}$ is the mean squared error of the method m at development age d and $msa^{(d)}$ is the mean squared anomaly of the data, also evaluated as of development period d, where anomaly is measured between the coverage years based on "actual" unpaid loss. Error is measured as the difference between the actual unpaid loss and the unpaid loss estimated by the method m as of development age d. These concepts are discussed further in Appendix B and in [8].

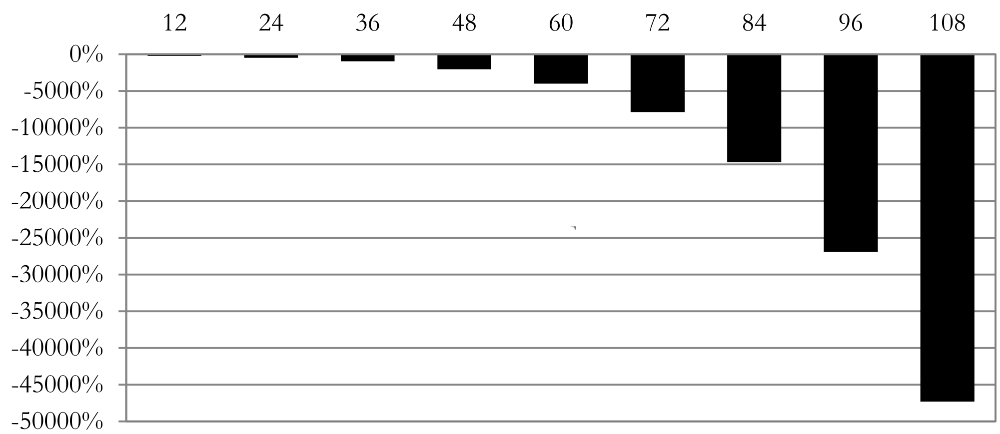
A method tends to exhibit one of two patterns with regard to its skill as the development age d increases (i.e., as more data for the coverage year at hand becomes available). The first pattern is exhibited by methods that reflect emerging experience, such as the LDF method. This pattern consists of an increase in the skill of the method, with the rate of increase typically declining over development periods and reaching an evaluation at which the skill has effectively plateaued. Chart 1 is an example of this pattern, displaying the median skill of the LDF-I method across all companies and lines of business within the analysis.

Chart 1
Median Skill – LDF-I Method
All Lines of Business, All Companies



The second pattern is a decline in skill as the development age *d* increases, which does not appear to level off. This pattern is typical of methods that do not respond to the experience of the given coverage year, such as the FS or LR method. Chart 2 is an example of this pattern, shown for the median skill of the LR1 method, again across all companies and lines of business within the analysis.

Chart 2
Median Skill – LR1 Method
All Lines of Business, All Companies



Note that in Chart 1 the skill increases noticeably from 96 to 108 months. Most likely this is not indicative of a "true" increase in skill at this evaluation, but is the result of the size of the triangles available from Schedule P, which terminate at 120 months. This causes the indicated skill at 108 months to be greater than the likely true skill at this evaluation.

2.3 Magnitude of Skill

Some commentary is warranted regarding the magnitude of indicated skill, in particular given the negative values shown in Charts 1 and 2. In general, skill should be viewed as a relative value with no inherent meaning of its own. In other words, the comparison of the skill of two methods can be very meaningful, but the skills themselves have no such meaning.

In practice, methods are seen frequently to have a negative skill. Mathematically, this means that the mean squared error in our methods typically exceeds the mean squared anomaly in our data. In other words, our methods are more volatile than the data itself, which may partly be the result of an insufficient volume of data.

A greater volume of (applicable) data would allow for greater stability in loss development factors and other method parameters, which would reduce the error in the methods. However, the additional data would not be fully correlated with the data previously available, and would consequently serve to reduce the anomaly in the overall data set. Thus it is not clear that skill would necessarily improve (or be positive) for larger data sets.

Additional insight can be gained by observing that the concept of skill originates in meteorology[4]. It seems reasonable that if we were to estimate tomorrow's high temperature as the average of the historical high temperatures for the same date, such a method might have a skill of 0%, with an expected mean squared error equal to the historical mean squared anomaly in the temperatures. Consequently a method that incorporates additional information, such as today's high temperature, into the prediction of tomorrow's high temperature would presumably represent an improvement and thus have positive skill.

We may assume that the meteorological method described above has 0% skill because the new information provided by tomorrow's high temperature will have very little impact on the calculation of the mean squared anomaly. This is strongly in contrast to a typical reserving scenario, in which estimates of development factors and loss ratios, to take two examples, often change significantly with the introduction of one new data valuation. Consequently, because the true anomaly of our data sets remains unknown, actuarial reserving methods will often exhibit negative skill.

2.4 How to Interpret an Improvement in Skill

It is helpful to have an intuitive understanding of skill and, in particular, what an increase in skill means regarding the volatility of a given method. As an example of this, consider as hypothetical examples Companies A, B, and C. These companies each write different volumes of what are otherwise similar books of business.

At a given month of development, Companies A, B, and C each have an expected unpaid loss ratio (i.e., unpaid loss relative to earned premium for the given coverage year) of 10%. Company A writes the most business, and Company A's data therefore exhibits the least variation. Company C writes the least business, and its data therefore exhibits the most variation. The mean squared anomalies of Companies A, B, and C's data at the given month of development are given in Table 2. This represents a fairly common range of mean squared anomalies when the expected unpaid loss ratio is approximately 10%.

---

[4] Additional discussion regarding the origins of skill can be found in [8].

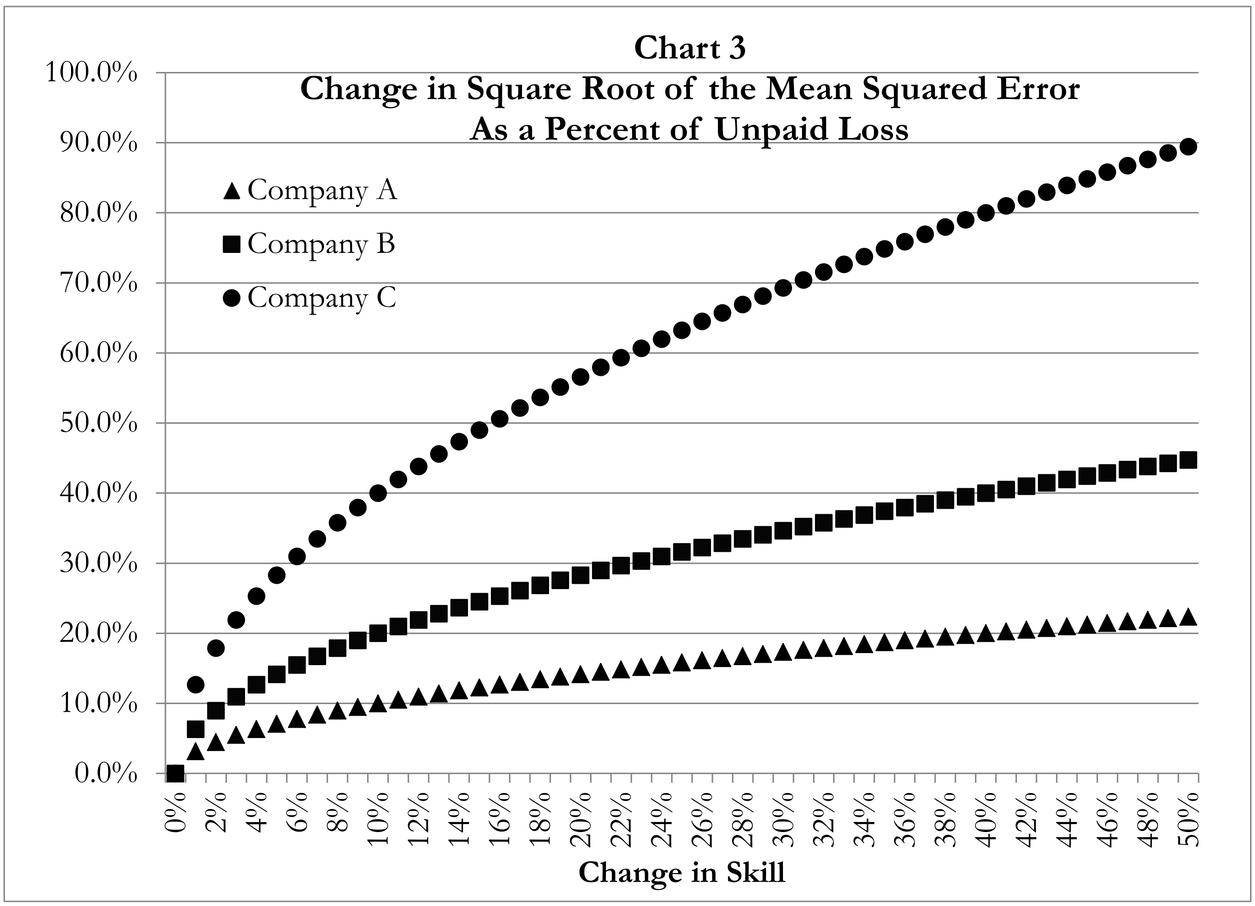
TABLE 2 MEAN SQUARED ANOMALY – HYPOTHETICAL EXAMPLES	
Company	**Mean Squared Anomaly**
A	0.1%
B	0.4%
C	1.6%

With this information we can calculate the impact of an increase in skill for a given method on the expected error in the unpaid loss, relative to the expected unpaid loss itself. The percentages in Table 3 are calculated algebraically based on the formula for skill. The square root of the mean squared error (RMSE) as a percent of unpaid loss is then equal to the RMSE as a percent of premium divided by the unpaid loss ratio.

TABLE 3 COMPANY A: CHANGE IN MEAN ABSOLUTE ERROR RELATIVE TO UNPAID LOSS, GIVEN A CHANGE IN SKILL			
Skill	**Mean Squared Error**	**RMSE* as a Percent of Premium**	**RMSE* as a Percent of Unpaid Loss**
1%	0.00%	0.3%	3.2%
2	0.00	0.4	4.5
5	0.01	0.7	7.1
10	0.01	1.0	10.0
15	0.02	1.2	12.2
20	0.02	1.4	14.1
30	0.03	1.7	17.3
50	0.05	2.2	22.4

* Square root of the mean squared error.

When expressed in this manner the impact of a change in skill becomes more apparent. For example, if we were able to improve the skill of an actuarial analysis for Company A by 50%, the RMSE within such an analysis would decline by approximately 22%, relative to the unpaid loss. The relationship between a change in skill for Companies A, B and C is shown in Chart 3. Thus for Companies B and C, whose books of business exhibit greater variation than Company A's, the reduction in RMSE is proportionally greater for the same change in skill than is the case for Company A. This implies, in a practical sense, that method skill becomes more important as the data becomes thinner.



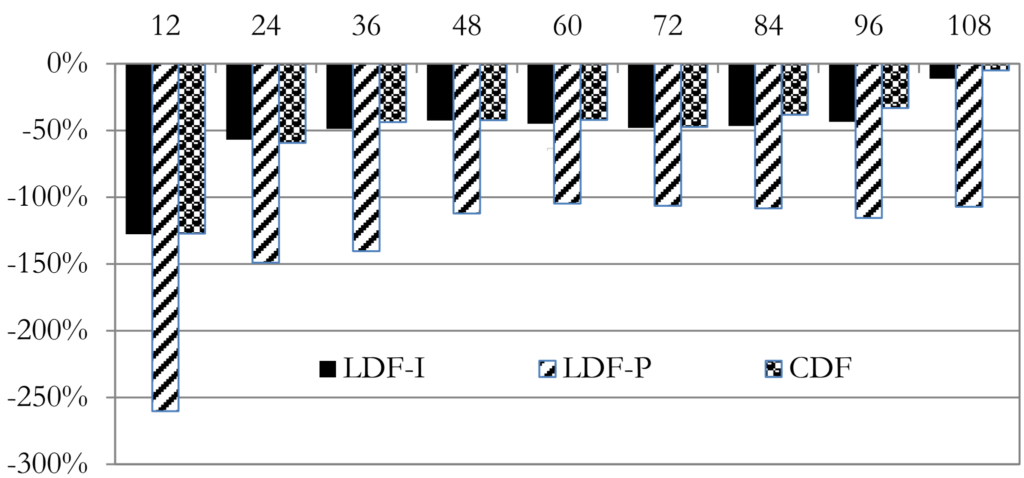
3. DISCUSSION OF RESULTS

The best-performing methods (i.e., the methods with the greatest skill) in the analysis were observed to satisfy the following two criteria:

1. Each relies at least in part on case reserves ("Criterion 1").
2. Amounts paid to date do not directly influence the indicated unpaid loss ("Criterion 2").

As a general example of this, consider Chart 4, which compares three common loss development methods, the LDF-I, LDF-P, and CDF methods. Results are shown across all companies and lines of business.

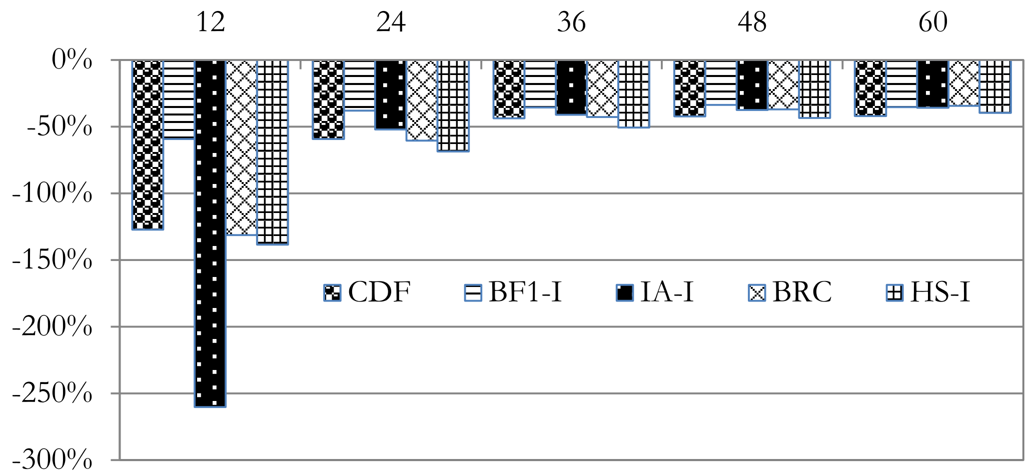
Chart 4
LDF-I, LDF-P and CDF Methods
Median Skill, All Companies, All Lines of Business



Thus the skill of the LDF-I method is seen clearly to exceed that of the LDF-P method. Results for the LDF-I method and CDF method are similar at earlier evaluations, although beginning at the 84-month evaluation the CDF method outperforms the LDF-I method. Differences in skill for the LDF-P and LDF-I methods exceed 100% at the first evaluation but remain close to or above 60% at later evaluations.

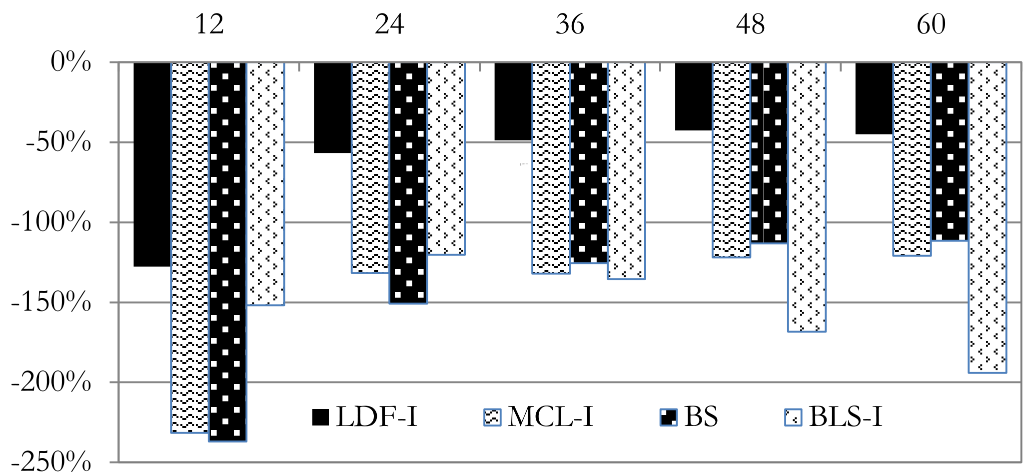
However, there are various methods that meet Criterion 1 and Criterion 2, and the CDF method is not necessarily the best of these. Consider the comparison given in Chart 5 across most of the methods that meet these characteristics. Of the five methods shown, the BF1-I method appears somewhat superior to the other methods considered. However, for any given company, this will depend on the applicability (more accurately, the skill) of the a priori loss ratio indications available.

Chart 5
Methods Satisfying Criteria 1 and 2
Median Skill, All Companies, All Lines of Business



It should be noted that the BT-I, BF2-I, CC-I, and IM-I methods are excluded from the above chart due to space constraints, although they each satisfy Criteria 1 and 2. The IM-I method significantly underperformed the other methods considered. Presumably this is due to the leveraged nature of the parameters on which this method relies. The other three methods on this list each underperformed the methods included in Chart 5, but only somewhat.

Chart 6
Incurred-Based Methods
Median Skill, All Companies, All Lines of Business



A similar comparison of the LDF-I method is made in Chart 6 with other methods that possess Criterion 1 but not Criterion 2. The LDF-I method is seen in Chart 6 to outperform the other incurred-based methods. At the earliest two evaluations, the BLS-I method outperforms the other incurred-based methods considered. However, the BLS-I method becomes too leveraged at later evaluations (beginning at 36 months) and underperforms the MCL-I and BS methods at this point.

The paid-based methods can be similarly compared, which shows that these methods generally underperform the LDF-P method. The paid-based methods that outperform the LDF-P method are the BF1-P, BT-P, and HS-P methods. However, these methods underperform the LDF-I method.[5] Note that these three methods satisfy Criterion 2 but not Criterion 1 (the remainder of the paid-based methods considered do not satisfy either of the criteria). Thus we could conclude that, while both criteria are important to method skill, Criterion 1 is more important than Criterion 2.

3.1 Results by Company Size

Results were similarly considered by company size, in which each company was segregated into a "Small," "Medium," or "Large" category. Companies were not permitted to migrate segments between evaluations, and were segregated according to their average annual net earned premium

[5] All three methods underperformed the LDF-I method at 24 months of development and subsequent. At 12 months of development, the BF1-P and BT-P methods outperformed the LDF-I method, while the HS-P method did not.

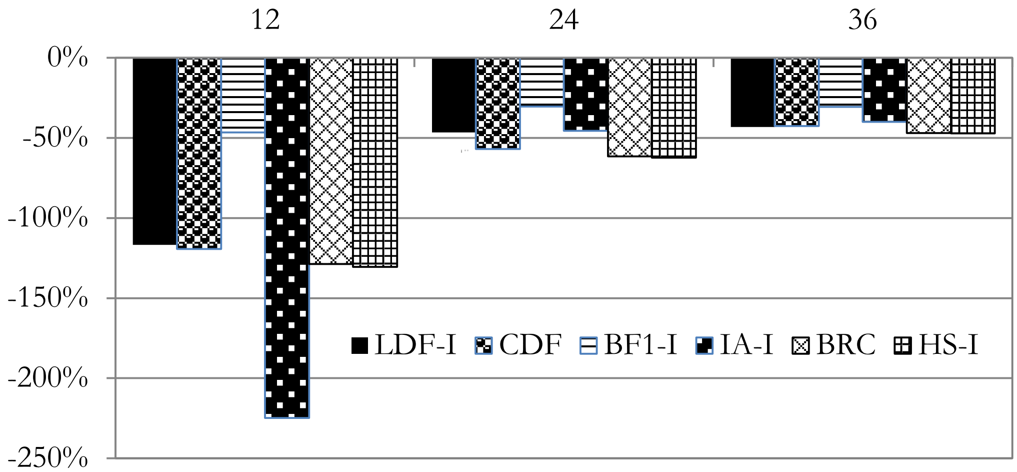
across all years considered. Table 4 provides the average 2010 net earned premium for the companies in each category.

TABLE 4 AVERAGE 2010 NET EARNED PREMIUM BY CATEGORY	
Company Size	**Average 2010 Net Earned Premium**
Small	$4.2 million
Medium	$17.5 million
Large	$350.0 million

In general, Criterion 2 was seen to be more important for small companies and less important for large companies, relative to all companies considered as a whole. In other words, for small companies, methods in which paid loss directly influences unpaid loss (e.g., the LDF-I method) are seen to have less skill, relative to other methods, than was the case when considering all companies as a whole. For large companies, methods such as the LDF-I method perform as well as methods such as the CDF method, and in some cases outperform these methods.

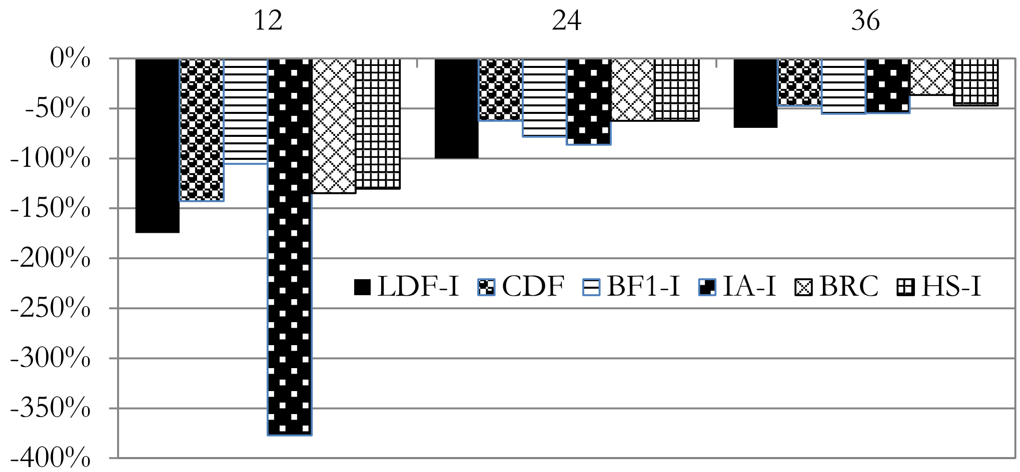
However, the general relationship between the various methods satisfying Criteria 1 and 2 that was present for all companies appears to hold when large companies are considered on their own. This is seen in Chart 7. Note that results in Chart 7 are shown through 36 months and that results subsequent to 36 months are similar.

Chart 7
Large Companies Only
Median Skill, All Lines of Business



Analogous results are given in Chart 8 for small companies. These show that for small companies methods that satisfy Criteria 1 and 2 outperform the LDF-I method. This outperformance is more apparent than it was when considering all companies as a whole. Presumably this suggests that small companies are more affected by the leveraged nature of methods such as the LDF-I method, in which paid loss to date that is greater than or less than the historical average paid loss affects the indicated unpaid loss for the book of business. Intuitively this seems reasonable, as small companies will exhibit more variable experience as a whole, including amounts paid to date.

Chart 8
Small Companies Only
Median Skill, All Lines of Business



3.2 Results by Line of Business

Results were also considered by line of business, where line of business was defined by the segments used within Schedule P. In general the results discussed above did not vary significantly by line of business. However, there are two apparent exceptions to this observation, specifically for the Homeowners and Workers Compensation lines of business.

Chart 9 displays the observed results for Homeowners coverage. In general the directional relationships between the methods are similar as when all lines are considered in tandem. However, the relative performance of the CDF method as compared to the LDF-I method is clearly different, with the CDF method outperforming the LDF-I method (as well as the LDF-P method). This may be the result of the fast-paying nature of Homeowners coverage for most claims.

When all lines of business were considered together, we saw that at later months of development (in particular, at 84 months of development and subsequent) the CDF method outperformed the LDF-I method. At this evaluation, the majority of claims are paid for most lines of business. For the Homeowners line of business, this point in time (when the majority of claims are paid) is reached much earlier. In general, it appears that the CDF method and other methods satisfying Criteria 1 and 2 may be most useful at "later" evaluations (where the definition of "later" varies by line of business according to the rate of payments).

Chart 9
Homeowners Line of Business
Median Skill, All Companies

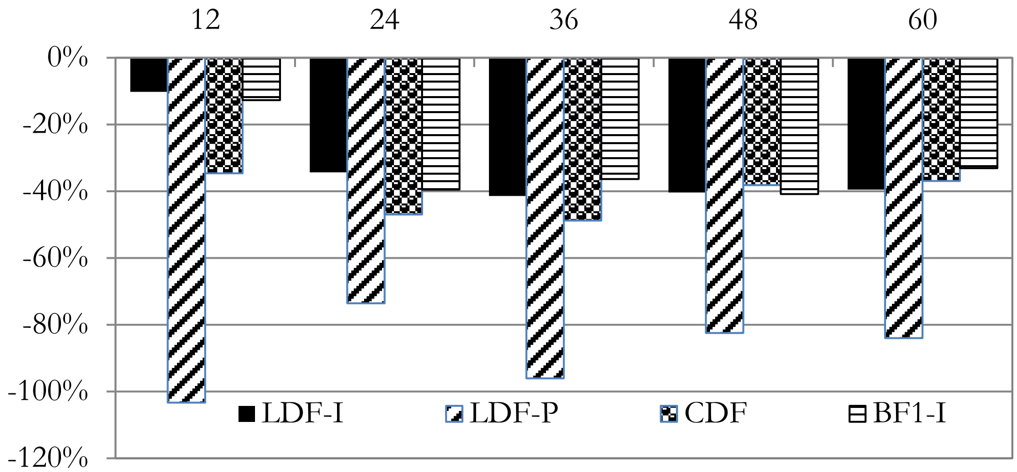
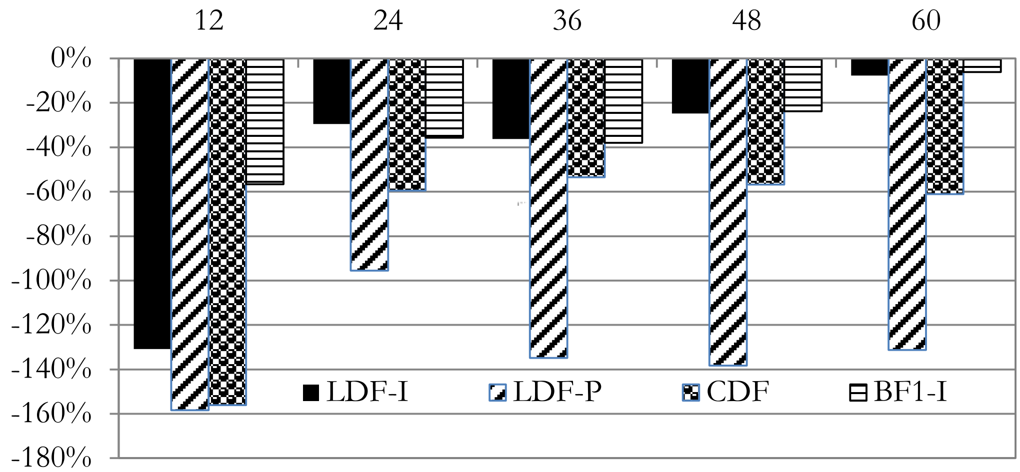


Chart 10 demonstrates a very different situation for Workers Compensation. For this line of business, the LDF-I method clearly outperforms the CDF method. Two possible reasons exist for this observation. The first is that the rate of payment for Workers Compensation claims is slower than for property & casualty coverages considered as a whole, and certainly slower than for Homeowners claims. This suggests a reason analogous to that observed for Homeowners, and that at a "later" evaluation the CDF method would outperform the LDF-I method for Workers Compensation. (This "later" evaluation would requisitely be at some point past 108 months, as the LDF-I method outperforms the CDF method for Workers Compensation at all evaluations made available by the Schedule P data.)

Chart 10
Workers Compensation Line of Business
Median Skill, All Companies



The second reason is the regular pattern of payments exhibited by Workers Compensation claims, as this rate of payments is largely determined by legislation. For most other lines of business, the rate of payments can be raised by efforts to settle claims more quickly, and the observed rate of payments can be altered by a particular large claim. Workers Compensation payments are largely immune to this phenomenon, as even a claim with a large medical component will often exhibit a rate of payments similar to other claims. Consequently the disadvantage of the LDF-I method that was observed earlier – that this method is easily biased up or down by the presence of absence of a large claim or larger than usual number of paid claims – does not apply to Workers Compensation.

While the number of claims with amounts paid to date for Workers Compensation in a particular coverage year may be greater or lesser than average, the paid loss on these claims is predictive of unpaid loss. Any Workers Compensation claim open 12 months after its accident date is generally a claim on which payments have been made to date and on which payments will be made for several years into the future (and perhaps for the lifetime of the claimant). Similarly, any Workers Compensation claim closed within this timeframe will generally have had a relatively small amount of payments (relative to the claims remaining open), making the impact of paid loss on such claims on unpaid loss within the LDF-I method largely immaterial. While the larger paid loss on claims remaining open would impact the indication of unpaid loss, these paid amounts, separated by an accounting date, would be reasonably expected to be highly correlated. Thus, under this reasoning, the CDF method is not necessarily a "worse" method for Workers Compensation, but the LDF-I method is a relatively "better" method, when compared to its performance for other lines of business.

This reasoning begs the question as to why the LDF-P method underperforms for the Workers Compensation line of business, similar to its performance for other lines. It is possible that if a much larger triangle of payments were available for Workers Compensation, the LDF-P method would not underperform, and might even outperform the LDF-I method. As discussed in Appendix A, to develop paid loss to the same level as incurred loss, it is necessary within the LDF-P method to apply a ratio of incurred-to-paid loss to the paid loss developed to a 10th report.

The ratio of paid-to-incurred loss will typically be greater than the ratio that is applied for other lines of business, and will typically be subject to greater volatility than for other lines of business. This ratio would not be necessary in the more common reserving scenario in which a larger triangle of paid Workers Compensation loss was available. Consequently, particular caution should be taken in inferring the applicability of these results for Workers Compensation onto the skill of methods in a reserve analysis for which a greater history of paid loss data is available.

4. EFFECT OF CORRELATION

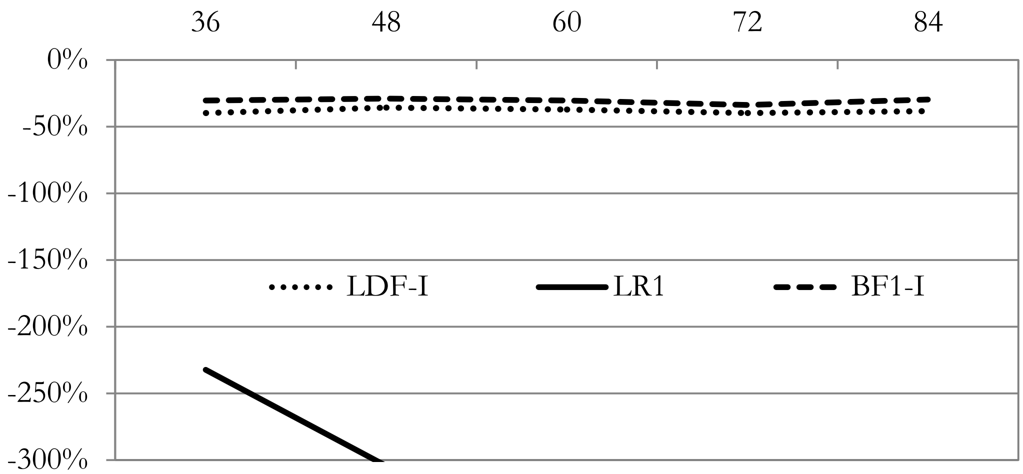
An interesting observation can be made in comparing the skill of three methods highlighted previously: the LDF-I, LR1, and BF1-I methods. Chart 11 shows the skill of these three methods on a logarithmic scale[6]. A logarithmic scale was used due to the vast disparity between the skill of the LR1 method and the other two methods (see Chart 2 and prior discussion concerning the LR1 method, and note that despite the logarithmic scale the skill of the LR1 appears within the range of the chart for only the first two evaluations).

Recall that the BF1-I method is a weighted average of the LDF-I methods and LR1 methods, in which the weights are determined by the loss development factors of the LDF-I method. Yet the skill of the BF1-I method is clearly greater than the skill of either of the two methods of which it is comprised. This is the result of correlation, and in particular the observation that the LR1 and LDF-I methods are partially but not fully correlated. Consequently the BF1-I method takes more

[6] More precisely, the translation used given a median skill of S to a logarithmic (median) skill of L was $L = -\text{Ln}\,[-(S-1)]$. Thus, since S must be less than or equal to 1, L was well-defined for all values of S. Negating the value of the logarithm ensures that L will exhibit the same property as S, in that if the skill of a given method is greater than another, its logarithmic skill will be greater as well.

information into account than either of these methods on their own, and the weighting of these two methods is such that the resulting observed skill is greater.

Chart 11
Logarithmic Skill
All Companies, All Lines of Business



This observation holds for any method calculated as the weighted average of other methods. As additional examples, consider Charts 12 and 13. The first of these charts compares the skill of the LDF-I and LDF-P methods with a new method for estimating unpaid loss calculated as 90% of the unpaid loss indication from the LDF-I method and 10% of the unpaid loss indication from the LDF-P method. For earlier evaluations, this new method performs comparably to the LDF-I method. For later evaluations (beginning at 60 months, as shown in Chart 12), the new method represents an improvement in skill. (Note that a 50/50 weighting of the LDF-I and LDF-P methods produced a method that underperforms relative to the LDF-I method.)

Chart 12
Weighted Average of the LDF-I and LDF-P Methods
Median Skill, All Companies, All Lines of Business

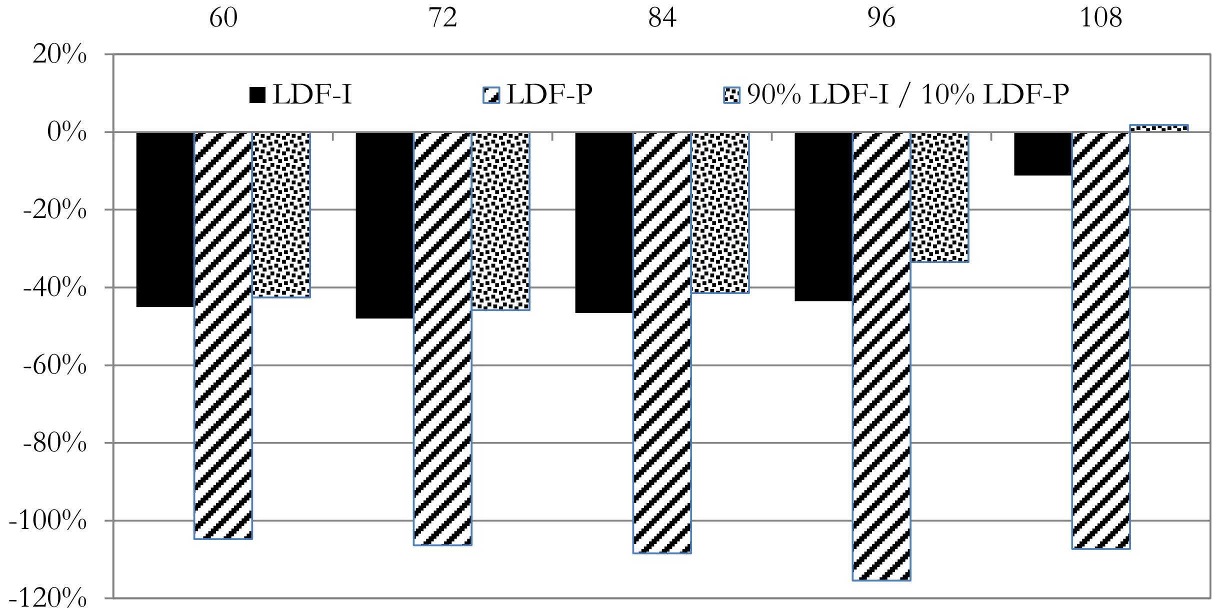
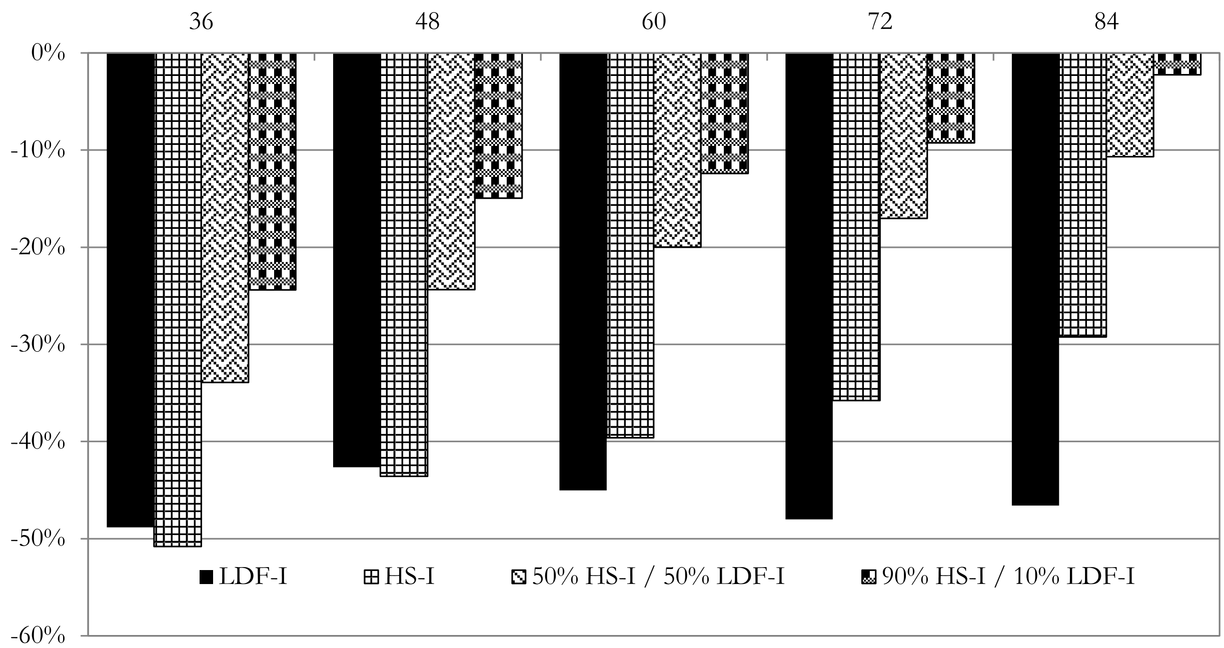


Chart 13 similarly shows the skill of four methods, the LDF-I and HS-I methods, a 50/50 weighting of these two methods, and a 90/10 weighting of these two methods (in which 90% of the weight is given to the HS-I method). It is easily seen that each of the weighted average methods outperform the two component methods (results before 36 months and after 84 months are similar to these respective observations). This suggests that the use of multiple methods (provided they are properly chosen) is very important to the skill of any actuarial analysis. The LDF-I and HS-I methods are two of the best-performing methods from the analysis, yet we are easily able to improve on the skill of these methods with a straight average of the two, and able to improve even more by judgmentally refining the weights.

Chart 13
Weighted Average of the LDF-I and HS-I Methods
Median Skill, All Companies, All Lines of Business



5. APPROACH TO ANALYSIS

Some discussion regarding the selection of method skill as the appropriate metric for the current analysis is warranted. Various techniques have been suggested for use in evaluating the hindsight performance of actuarial methods. These have historically fallen in three categories:

1. "The Scorecard System" compares the indicated ultimate loss for an individual entity or data set, where the comparison is made either from one evaluation to the next or from a given evaluation to the "true ultimate" loss, once that is known. References [10] and [16] are both examples of this technique. An advantage of this technique is its simplicity and ease of explanation, but a disadvantage is that it provides no way of aggregating observed results for a single method or entity, or across multiple methods or entities. For this reason the technique has typically been used for single entities or data sets only.
2. Calculating the mean and standard deviation of the prediction errors, and similar statistical calculations related to the prediction error's distribution, is another evaluation

method. Here, prediction error is defined as the difference between true ultimate loss and the ultimate loss indicated by a given method as of a given evaluation. References [3], [12], [13], and [15] are examples of this technique. In each of these monographs, triangular data sets are simulated under a set of assumptions and standard actuarial methods are applied to the simulated data in a mechanical fashion. Hence, given the underlying assumptions, the distribution of prediction error is readily known, provided that a sufficient number of triangular data sets are simulated.

The assumptions used in the simulation are typically specific to a given line of business or amount of business written, however, and consequently the results may not be applicable to other lines of business or situations when a greater or lesser amount of business is written. For example, claims may be assumed to be reported promptly after occurrence, or it may be assumed that there is a significant lag in claim reporting, and both assumptions can have a significant impact on the performance of claim-based methods. Similarly, the amount of business written will clearly impact the standard deviation of prediction error, and some methods may be impacted more than others.

For the analysis considered here, multiple lines of business were analyzed across various company sizes. Consequently any attempt to derive statistics concerning the distribution of prediction error would either need to consider each line of business and company size grouping separately, or would need to attempt to normalize for these differences. Even if data were segregated, differences in company size would still exist between different companies in the group, and any attempt at additional segmentation to mitigate this issue would likely result in a statistically insignificant sample size. Any attempt at normalization across companies or lines of business would ideally consider the amount of business written as well as the inherent volatility of that business.

3. Method skill has the advantage of normalizing for differences in premium earned and the resulting volatility in unpaid loss by company (a disadvantage in the discussion of the distribution of prediction error above). Taking the form of a single numeric value can be seen as both an advantage and a disadvantage. Method skill will not provide additional information on the distribution of prediction error, and in particular will not indicate whether a method is biased. Taking the form of a single numeric value, however, allows more easily for comparison across different companies and lines of business.

   One final aspect of method skill that should be considered is that, by itself, it provides no statistical significance of its indication. Observations regarding statistical significance could be added to an analysis such as this by segregating data triangles randomly into two or three groups, then comparing the indicated skill for each group. While this random segmentation is outside the scope of the current analysis, segmentation of results has been done by company size and line of business, as discussed above. The general similarity of results by line of business (with the exception of the Homeowners and Workers Compensation lines) suggests a meaningful level of statistical significance for the results, although this cannot be measured with precision.

6. CONCLUSIONS

The results of this empirical analysis suggest three conclusions:

A. In most situations, the methods with greatest skill are those satisfying Criteria 1 and 2, first defined in Section 3 above:

   1. Each relies at least in part on case reserves ("Criterion 1").

   2. Amounts paid to date do not directly influence the indicated unpaid loss ("Criterion 2").

B. These methods (those possessing the greatest skill) are not commonly in use.

C. The weighting schemes most commonly in use are not supported by the current analysis.

Consider that the LDF-I and LDF-P methods are ubiquitous throughout reserve analyses. While the best reserve analyses will almost always contain other methods as well, significant weight is typically given to these two methods. However the above results suggest that many more valuable methods exist, and that the LDF-P method in particular should receive little to no weight in most analyses. This is noticeably in contrast to the 50/50 weighting used in many reserve analyses for the LDF-I and LDF-P methods.

Where not already in use, we could greatly improve our analyses by use of methods satisfying Criteria 1 and 2. In particular, the CDF, IA-I, BR, and HS-I methods would all serve in many cases to enhance and improve our work. Various versions of the BF method are commonly in use now, although this analysis suggests that greater weight should likely be given to the BF methods than is typical currently. Often BF methods are viewed as appropriate for "middle" years, yet the current analysis suggests that giving weight to the BF methods can improve our analyses for all years.

Additional work in the area of method weights would be helpful, as this paper has merely touched the surface of that topic. Results shared here suggest that the selection of method weights is significant to the skill of an analysis, perhaps more so than the methods themselves. Consider, for example, that little improvement in skill is gained by switching from the LDF-I method to the HS-I method. However, a 25% additive improvement in skill is gained at all evaluations by using a 90/10 weighted average of the HS-I and LDF-I methods, respectively.

Another topic not addressed here is the weighting of more than two methods, as would be done in practice. It is likely that as any method is added to an analysis already consisting of several methods, the incremental skill achieved by giving weight to such a method will decline as the number of methods already incorporated into the weighted average increases. However, the number of methods required before this incremental skill becomes minimal is an open question, as is whether this number could be reduced by appropriate selection and weighting of the methods. Additional research in this area is welcome.

Acknowledgment

The author would very much like to thank:

- For their many hours programming the analysis discussed here, Paul Keuler, Linda Gritter, and Shaun Cullinane (who also provided helpful comments on the paper itself);
- For assistance with graphs as well as the underlying analysis, Brooke Mote-Nirode; and
- For comments and review of the analysis itself, Eric Wunder, Mark Mulvaney, and Jeff Courchene.

Appendix A – Loss Reserving Methods

The following provides a list of the methods considered in the analysis, including the abbreviation used to refer to each method (note that for methods for which there are paid and incurred versions, multiple abbreviations are given). Also included is any relevant information as to how the method is applied within the current analysis, given the data limitations of Schedule P. As discussed further in Appendix B, the methods outlined below develop indications of loss at a 10th report (i.e., the last evaluation included within the Schedule P triangles) rather than indications of loss at ultimate.

1. **Backward Recursive Case Development (BRC)**

   This method is discussed by Marker and Mohl in [11]. The paid-on-prior-case and case-on-prior-case factors selected for our analysis are each the weighted average of the columns of these factors as given by the triangles, where the weights are proportional to the prior case. At a 10th report, we have assumed a paid-on-prior-case factor of 1.00 and a case-on-prior-case factor of 0.00.

2. **Benktander (BT)**

   The Benktander method, discussed in [9], is often referred to as the "iterated Bornhuetter-Ferguson method." In the BT method, a priori loss is equal to the indication from the BF method (in our case, BF1-I for the incurred method, and BF1-P for the paid method). The calculation of indicated loss then proceeds as described for the BF method, with calculations of the percent unpaid for the BT-P method and the percent IBNR for the BT-I method.

3. **Berquist-Sherman Case Adjustment (BS)**

   The BS method is the first of the two methods given in [2], in which an adjustment is made to the incurred loss in the prior diagonals of a given triangle for assumed changes in case reserve adequacy. This adjustment is made by de-trending the average case reserve along the most recent diagonal of the triangle (in the case of the current analysis, at a rate of 5.0% per annum). The result is multiplied by the number of open claims within prior diagonals in order to obtain an indication of case reserves from prior diagonals at the approximate level of case reserve adequacy as the most recent diagonal. Incurred loss development factors are then developed and applied to loss along the most recent diagonal as for the LDF-I method.

4. **Bornhuetter-Ferguson 1 (BF1)**

   The first of the BF methods included in the analysis uses the indicated loss from the first loss ratio method (LR1), described below, as the a priori indicated loss. The percent unpaid and percent IBNR are then calculated as described in [4], producing both paid (BF1-P) and incurred (BF1-I) versions of this method.

5. Bornhuetter-Ferguson 2 (BF2)

The second of the BF methods is an iterative procedure in which the a priori indicated loss is based on the weighted average loss ratios of preceding accident years, as based on the BF2 method indications for these years. The oldest accident year in the triangle, as well as any other accident year for which loss ratios of older accident years are not available, relies on the same a priori loss ratio as the BF1 method. Both paid (BF2-P) and incurred (BF2-I) versions of this method are calculated.

6. Brosius Least Squares (BLS)

The BLS method considers that there may be both additive and multiplicative aspects of loss development. Thus the method iteratively develops both a multiplicative loss development factor, to be applied to losses paid or incurred to date, and an additive factor, to be included subsequent to the multiplication. The factors are based on a least squares regression, where the incurred loss ratio at a 10^{th} report is the dependent variable and the paid or incurred loss ratio at the given evaluation is the independent variable. The use of loss ratios rather than loss is a difference from the methodology as presented in [5], and was done so as to normalize for changes in exposure across accident years. Both paid (BLS-P) and incurred (BLS-I) versions are included.

7. Brosius Least Squares – Weighted (BLSW)

Having observed certain indications produced by the BLS method, we sought to enhance the reliability of this method by giving more credibility in the regression process to years with greater premium, and presumably greater exposure. The Weighted Brosius Least Squares method that resulted uses a regression process weighted by premium, in contrast to the unweighted regression used in the BLS method itself.

8. Cape Cod (CC)

The Cape Cod method is very similar to the BF method, but develops a priori loss under the assumption that in total across accident years it should be equal to the CC method indication. For the CC method as included in this analysis, we have assumed the same loss ratio for each accident year (i.e., unlike certain of the loss ratio methods discussed below, there is no a priori difference assumed by year). Both paid (CC-P) and incurred (CC-I) versions of the method are included.

9. Case Development Factor (CDF)

The CDF method is based on the loss development factors from the LDF method, discussed below. In the CDF method an indicated unpaid-to-case ratio is derived from the relationship between unpaid loss and case loss implicit in the selected paid and incurred loss development factors. This factor is then applied to the case reserve to derive an indication of unpaid loss, which is added to paid loss to date for an indication of loss incurred through the 10^{th} report.

10. Frequency/Severity (FS)

The FS method is based on a projection of reported claims at a 10^{th} report and a severity applied to these claims. Reported claims are based on the company's triangular reported claims data (i.e., Section 3 of Part 5 of Schedule P for the given line of business) developed to a 10^{th} report using weighted average reported claim development factors. Given the relatively favorable performance of the LDF-I method as well as its general acceptance within actuarial practice, we took the LDF-I method to be the "preliminary" selected method for use in selecting severities.

Thus the severity for each accident year is calculated as the incurred loss at a 10^{th} report indicated by the LDF-I method divided by the indicated reported claims at a 10^{th} report. For a given accident year, a severity is selected based on the weighted average severities of all prior accident years, where the weights are proportional to the projected reported claims. In this process, the severities are trended to the accident year in question at a rate of 5.0% per annum.

11. Hindsight Outstanding/IBNR (HS)

The HS method is similar to the FS method in that it relies on an equivalent projection of reported claims as well as a preliminary selected loss method (also the LDF-I method). However within the HS method, the projection of reported claims is used to calculate a triangle of "hindsight outstanding" claims, which are the difference between the projection of reported claims at a 10^{th} report and closed claims to date. Similarly, the preliminary selected loss method is used to calculate a triangle of hindsight outstanding loss, which is the difference between the preliminary method loss projections and the paid or incurred loss to date. Thus the difference represents unpaid loss for the HS-P method and IBNR loss for the HS-I method.

The ratios of the values within the hindsight outstanding loss triangle to the corresponding values within the hindsight outstanding claims triangle produces a triangle of hindsight outstanding severities (unpaid severities for the HS-P method and IBNR severities for the HS-I method). For a given accident year, severities from the preceding years are trended at 5.0% per annum to the accident year in question. A weighted average of these severities, where the weights are proportional to hindsight outstanding claims, is selected.

The weighted average hindsight severity is then applied to the number of projected outstanding claims for the given accident year to produce indications of unpaid loss for the HS-P method and IBNR loss for the HS-I method. These are then added to paid loss or incurred loss, respectively, to derive indications of incurred loss at a 10^{th} report. This method is also referred to as the "ultimate unclosed claim severity technique" within [7].

12. Incremental Additive (IA)

In this method, incremental (i.e., calendar year) changes in paid or incurred loss are observed by accident year and compared to the premium for that year. A weighted average ratio of incremental loss to premium is selected, where the weights are proportional to the premium. These ratios are accumulated to derive an IBNR-to-premium or unpaid-to-premium ratio at the given evaluation. The ratios are applied to premium to derive IBNR or unpaid loss itself, then added to incurred loss or paid loss, respectively, for the IA-I and IA-P methods. So that the IA-P method will produce an indication of incurred loss at a 10^{th} report, the unpaid-to-premium ratio at a 10^{th} report is set equal to the case-to-premium ratio at a 10^{th} report of the earliest year in the triangle.

13. Incremental Claims Closure (ICC)

The incremental claims closure method is described by Adler and Kline in [1]. In this method, reported claims at a 10^{th} report are projected based on the reported claims triangle and weighted average reported claims development factors selected from this triangle (as above for the FS and HS methods). A closing pattern is then selected based on historical weighted average incremental closed-on-prior-open factors, where the weights are proportional to the number of claims open. These factors are then applied iteratively to project incremental closed claims, with the difference between the projected reported claims at the 10^{th} report and the projected closed claims at the 10^{th} report being the number of claims projected to close after the 10^{th} report.

As the next step, historical incremental paid loss is compared to incremental closed claims to derive incremental paid loss per closed claim by time period. These amounts are then trended at 5.0% per annum to the relevant time period and a weighted average of the indications selected (where the weights are proportional to the number of closed claims). Prospective incremental paid loss by accident year is then projected as the product of the projected incremental closed claims and the projected paid loss per closed claim, each for the same time period. Ultimate loss is then the sum of these projections with paid loss to date. Within the current analysis, claims that are projected to close after the 10^{th} report are assumed to have a severity equal to that of the claims that close between the 9^{th} and 10^{th} reports, but trended one additional year.

14. Incremental Multiplicative (IM)

The incremental multiplicative method is similar to the incremental additive method in that both methods consider incremental loss triangles. However, the IM method calculates development factors that are ratios of incremental loss in one time period to the incremental loss in the preceding time period. Weighted averages of these development factors are calculated, where the weights are proportional to the incremental loss in the preceding time period.

The development factors are then applied iteratively to project incremental loss in subsequent time periods. Projections of unpaid loss and IBNR loss are derived for the IM-P and IM-I methods, respectively, by accumulating the indications of incremental paid and incremental incurred loss by time period. These projections of unpaid loss and IBNR loss are added to paid loss to date and incurred loss to date, respectively, to derive distinct indications of ultimate loss.

Within the IM-P method, a tail factor from paid loss at a 10^{th} report to a level reflecting incurred loss at a 10^{th} report is selected based on the oldest accident year in the triangle and the assumption that the case loss within this accident year will be paid as is. In other words, the tail factor is the case loss for this year divided by the incremental paid loss for this year in the time period preceding the 10^{th} report. If incremental paid loss for this time period is zero, then such a ratio is undefined and assumed to be zero for purposes of our analysis.

15. Loss Development Factor (LDF)

The LDF methods are based on the calculation of historical loss development factors from the paid and incurred triangles. The weighted average loss development factor from all available years within the triangle is applied to loss at the given evaluation date to derive indicated loss at a 10^{th} report. Both paid (LDF-P) and incurred (LDF-I) versions of this method are included within the analysis. For the paid method, a tail factor to develop the losses from paid at a 10^{th} report to incurred at a 10^{th} report is equal to the incurred-to-paid ratio at a 10^{th} report for the earliest year in the triangle.

16. Loss Ratio – Based on A Priori Assumption (LR1)

Three versions of the loss ratio method are included within our analysis. Each relies on net earned premium by calendar year, consistent with the use of net paid and incurred loss within the triangles. The first of these (LR1) is based on a priori industry indications of the loss ratio for the given coverage year. These loss ratios were derived from historical A.M. Best Review & Preview reports.

17. Loss Ratio – Based on Preliminary Selected for Prior Years (LR2)

The remaining two loss ratio methods are each based on the use of preliminary selected incurred loss at a 10^{th} report, which for both is set equal to the results of the LDF-I method, consistent with the preliminary selected loss in the FS and HS methods. For the LR2 method, the loss ratio for a given accident year is set equal to the weighted average of the loss ratios produced by the preliminary selected method within the preceding accident years of the triangle, where the weights are proportional to net earned premium. This loss ratio is then multiplied by net earned premium for the given calendar year to derive indicated incurred loss at a 10^{th} report for the LR2 method.

18. Loss Ratio – Based on Preliminary Selected for Most Recent Three Prior Years (LR3)

The LR3 method is very similar to the LR2 method, but rather than relying on all preceding accident years within the triangle, relies on at most the preceding three accident years. Thus this method is more responsive to recent loss ratio experience, but potentially more volatile.

19. Munich Chain Ladder (MCL)

The MCL method is described by Quarg and Mack in [14]. Similar to the LDF method, discussed above, there are paid (MCL-P) and incurred (MCL-I) versions of the MCL method. In practice, these indications often converge on each other, although the indications are rarely equal. Due to the convergence of the two methods, no adjustment factor is included in the calculation of the MCL-P method, which is distinct from the LDF-P method.

Appendix B – Example of Method Skill Calculation

An extensive discussion of the calculation of method skill is provided in [8], which I will not replicate here. However, given the limitations of Schedule P data as well as the volume of such data considered within the analysis discussed here, there were certain requisite judgments that needed to be made in implementing the necessary calculations. First, some background on terminology and the requisite calculations is appropriate.

Using terminology similar to that given in [8], the skill of a method m applied to a given data set as of development period d is calculated as

$$Skill_m^{(d)} = 1 - mse_m^{(d)} / msa^{(d)}$$

where $mse_m^{(d)}$ is the mean squared error of the method and $msa^{(d)}$ is the mean squared anomaly, a property of the data independent of the method. Both $mse_m^{(d)}$ and $msa^{(d)}$ are calculated as of development period d and can be expected to vary by development period. Typically the method m will be a function of the triangular matrix $[C_{i,j}]$, in which rows would most often represent coverage years and columns the development periods. Each $C_{i,j}$ would typically represent cumulative paid loss in coverage year i through development period j. The method m could also be a function of an analogous triangular matrix of case reserve loss, in addition to the matrix $[C_{i,j}]$.

For each coverage year i, the method m produces an estimate of ultimate loss as of development period d of $\hat{C}_i^{(d,m)}$. This should be distinguished from C_i, the true ultimate loss for the coverage year i. Given earned premium of E_i for year i, the error of method m for year i as of development period d is

$$Error_{i,m}^{(d)} = [\hat{C}_i^{(d,m)} - C_i] / E_i.$$

The mean squared error for the method as of development period d is then

$$mse_m^{(d)} = \textstyle\sum_i P_i \times [Error_{i,m}^{(d)}]^2 / \sum_i P_i$$

where the sums are taken over all coverage years i. This is a weighted average in which the weights are the percent paid for each coverage year at the "actual" evaluation, denoted by P_i.

The calculation of the anomalies requires a weighted average of the "actual" unpaid loss ratios for the given book of business, which is

$$ULR = \textstyle\sum_i \{P_i \times C_i \times [1 - P_i] / E_i\} / \sum_i P_i.$$

Then the anomaly of the unpaid loss ratio for a given coverage year i is

$$A_i = C_i \times [1 - P_i] / E_i - ULR$$

and the mean squared anomaly for the data set is

$$msa = \textstyle\sum_i [P_i \times A_i^2] / \sum_i P_i.$$

The above calculations require the ultimate loss C_i to be known. In practice, however, this is often not known, especially for more recent coverage years. Hence the calculation of skill incorporates the use of what is effectively a credibility-weighting procedure, in which each of the averages that is taken above is a weighted average in which the weights are proportional to P_i, the portion estimated to be paid (assuming C_i is correct) for each coverage year.

While any claims data set will exhibit uncertainty in $\{C_i\}$, this issue is somewhat more pronounced for any claims data set consisting of Schedule P triangles, such as the data set underlying the current analysis. This is because the triangles within Schedule P contain at most 10 years of data, and quite often the ultimate loss for a given coverage year remains unknown at 10 years of development. Therefore, within this analysis, we have requisitely based the calculations of skill discussed above on the paid plus case loss as of 10 years of development. In other words, the "ultimate" loss for purposes of the current analysis is the paid plus case loss as of 10 years of development, and no "tail" factor has been included in the analysis to estimate development subsequent to this evaluation. We believe the use of paid plus case loss as opposed to paid loss or an amount in between the two is reasonable given the consistent historical adverse development on case reserves (and IBNR) for the U.S. property & casualty industry as a whole after 10 years of development (see [6]).

The paid plus case loss at 10 years of development is known with certainty for any coverage year that has reached this maturity. This amount can be forecast for less mature coverage years by the various methods. However these forecasts will typically vary between the methods. Consequently a forecast of C_i must be selected for any coverage year of less than 10 years maturity. Within the current analysis we have varied the selected forecast of C_i according to the method m whose skill is being calculated.

For example, in calculating the skill of the LDF-I method, C_i is forecast based on the most recent indication of C_i based on this method. In calculating the skill of the LDF-P method, the forecast of C_i would differ, and be based upon the most recent indication as given by the LDF-P method. An alternative approach would be to select a fixed method for purposes of defining C_i. However, this would presumably bias the analysis in favor of that method, and so has not been chosen. Varying the selection of C_i serves to place the methods on a more equal footing with each other within the analysis.

Varying C_i by method works for any method whose indication converges on the true ultimate loss as the development period increases. This is true for most of the methods considered within the current analysis, but does not hold for the FS method or any of the LR methods. For these methods, in the more recent years where the paid plus case loss at 10 years of maturity is unknown, we used the forecast of this amount as given by the LDF-I method. This method was selected given its standard acceptance as well as its generally favorable performance within the current analysis.

7. REFERENCES

[1] Adler, Martin, and Charles D. Kline Jr., "Evaluating Bodily Injury Liabilities Using a Claims Closure Model," *CAS Discussion Paper Program*, May 1988, 1-67.

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

[3] Boles, Tapio, and Andy Staudt, "On the Accuracy of Loss Reserving Methodology," *CAS E-Forum*, Fall 2010, 1-62.

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

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

[6] Forray, Susan J., "Mind the Gap—Recognizing Perceived and Ultimate Liabilities," *Contingencies*, Nov./Dec. 2011, 64-67.

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

[8] Jing, Yi, Joseph Lebens, and Stephen Lowe, "Claim Reserving: Performance Testing and the Control Cycle," *Variance*, 2009, Vol. 3, Issue 2, 161-193.

[9] Mack, Thomas, "Credible Claims Reserves: The Benktander Method," *ASTIN Bulletin*, 2000, Vol. 30, Issue 2, 333-347.

[10] Mahon, Mark J., "The Scorecard System," *CAS Forum*, Summer 1997, Vol. 1, 97-136.

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

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

[13] Pentikäinen, Teivo, and Jukka Rantala, "A Simulation Procedure for Comparing Claims Reserving Methods," *ASTIN Bulletin*, 1992, Vol. 22, Issue 2, 191-216.

[14] 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*, Fall 2008, Vol. 2, Issue 2, 266-299.

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

[16] Wacek, Michael G., "A Test of Clinical Judgment vs. Statistical Prediction in Loss Reserving for Commercial Auto Liability," *CAS Forum*, Winter 2007, 371-404.

Abbreviations and notations

ALAE, allocated loss adjustment expense
BRC, backward recursive case development
BT, Benktander
BS, Berquist-Sherman case development
BF, Bornhuetter-Ferguson
BLS, Brosius least squares
BLSW, Brosius least squares – weighted
CAS, Casualty Actuarial Society
CC, Cape Cod
CDF, case development factor
FS, frequency/severity
HS, hindsight outstanding
I, incurred
IA, incremental additive
IBNR, incurred but not reported
ICC, incremental claims closure
IM, incremental multiplicative
LDF, loss development factor
LR, loss ratio
MCL, Munich chain ladder
P, paid

Biography of the Author

Susan J. Forray is a Principal and Consulting Actuary in the Milwaukee office of Milliman. She is a Fellow of the Casualty Actuarial Society and a Member of the American Academy of Actuaries. Susan has provided actuarial assistance to a spectrum of risk-bearing entities in both the private and public sectors, ranging from large multi-line commercial insurance companies to self-insured programs. She has also been active in the Casualty Actuarial Society, having served on the Examination Committee, the editorial board of the journal *Variance*, and the Committee on Professionalism Education. She is a frequent presenter at industry symposia, and her work has been published in the CAS *Forum*, *Best's Review*, *The Physician Insurer*, and *Contingencies*. She can be reached at susan.forray@milliman.com.