By Susan J. Forray

Tried and true may be familiar, but is it the best option?

A Fresh Look at Actuarial

THERE ARE AT LEAST 30 DIFFERENT ANALYTIC METHODS that actuaries rely upon when developing indicated loss reserves for property and casualty (P/C) companies. Two methods stand out as (by far) the most popular: the paid chain ladder and the incurred chain ladder (i.e., paid plus case).

Why are these two methods the most common? Part of the answer must be their ease of use. But another reason, I believe, is simple familiarity. Actuaries use these tools nearly every day and have for a long time; this has led to the paid and incurred chain ladder becoming default reserving methods.

Until recently, there have been no studies or hard evidence to determine whether these two methods, in addition to being the most popular, are indeed the most reliable. Results from different methods can diverge significantly, so actuaries would benefit from an empirical analysis to support their choice of analytic methods.

An Evidence-Based Review

With the help of others, I decided to make an attempt to fill this need by testing 30 different analytic methods, including the paid and incurred chain ladder. We collected Schedule P triangular data from more than 3,000 individual P/C insurance companies over a 15-year period and then compared loss at the end of that time against projections arrived at using all 30 methods.

We focused on annual statements from 1996 to 2010, each of which contained triangular data for the preceding decade, and developed indications of ultimate loss by accident year and method for every line of business available within each company.

The most recent evaluation date available at the time of our analysis was Dec. 31, 2010, so this functioned as the date at which "actual" ultimate loss would be established (for older accident years, this date was the last evaluation for the year included in Schedule P).

We also applied to the analysis a concept relatively new to the actuarial profession known as "method skill." Originally developed in the field of meteorology, method skill provides an approach to analyzing different methods that normalizes across a broad range of widely different data sets. The result is a single number for establishing a method's relative "skill" at using historical data to develop reliable projections.

The results of the analysis were surprising. In brief, we found that the methods exhibiting the greatest skill over time were not the most popular but rather those that best satisfied the following two criteria:

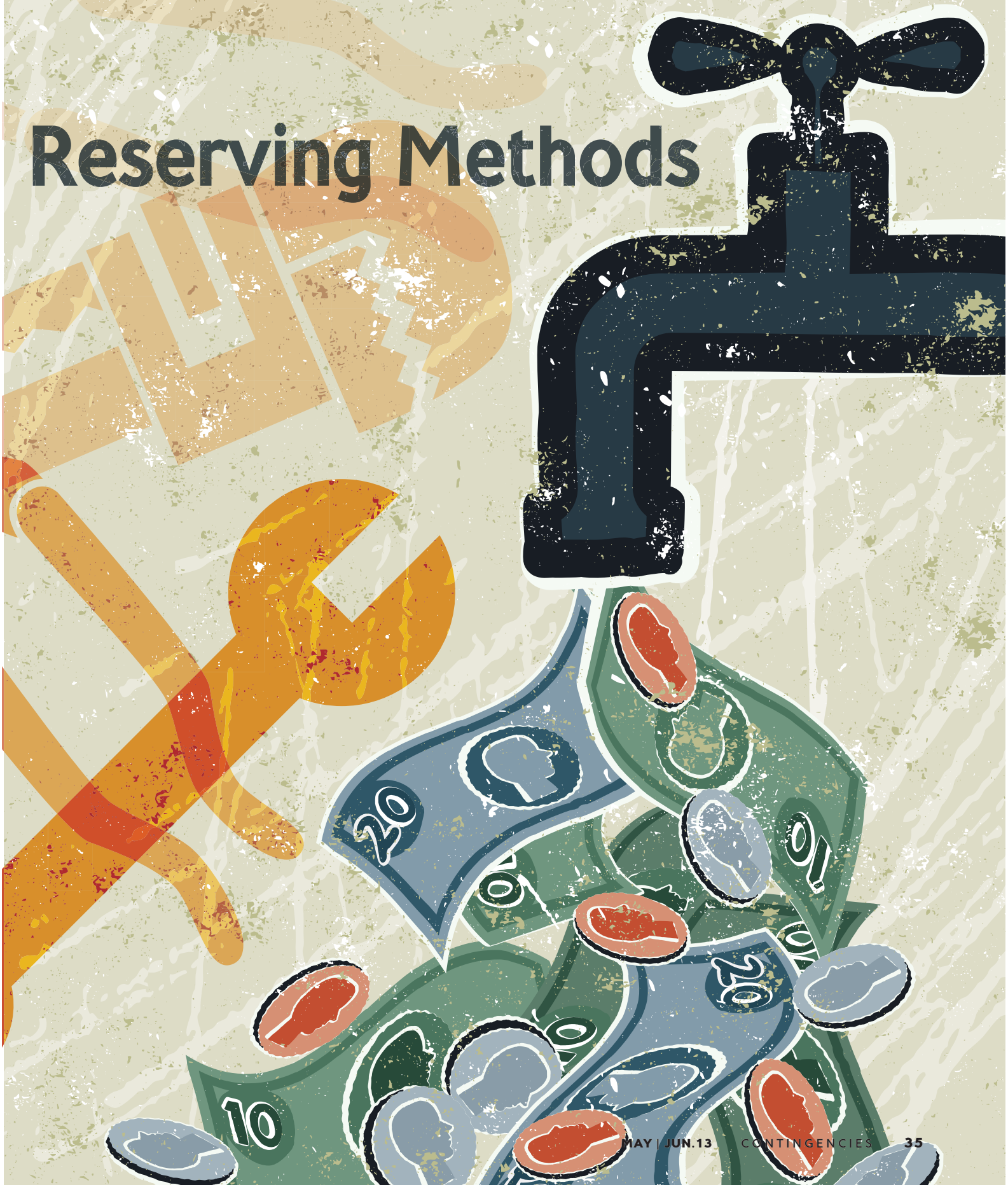
1. They relied, at least in part, on case reserves in their evaluations.
2. The paid-to-date data they used did not directly influence the indicated unpaid loss.

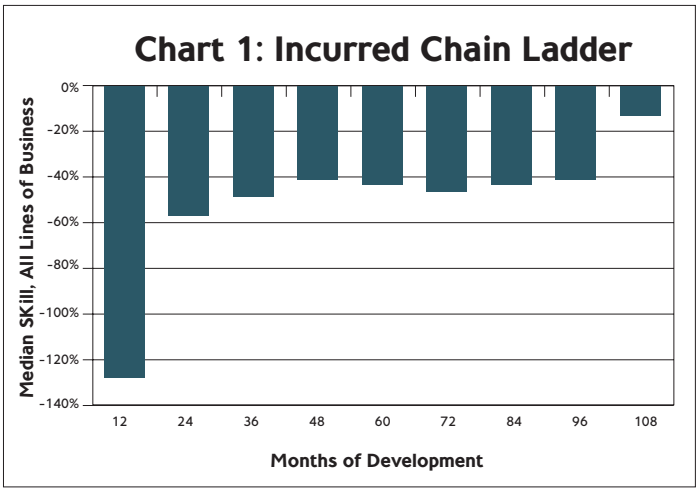
We identified different methods satisfying both criteria, each of which exhibited greater skill than the incurred chain ladder—and significantly greater skill than the paid chain ladder.

Defining Method Skill

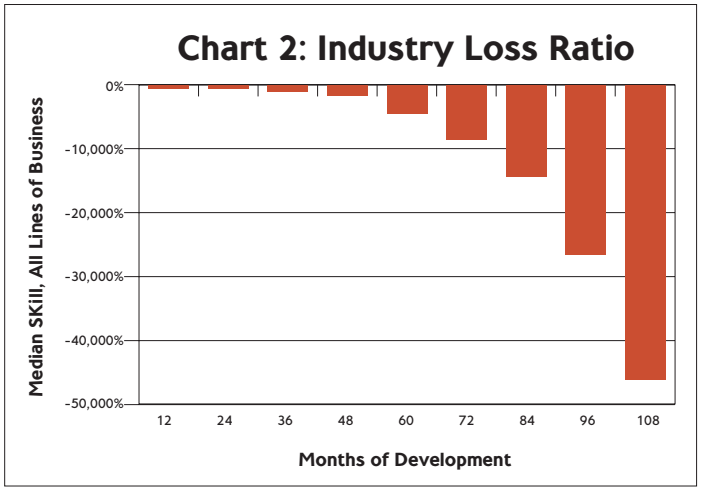
According to *Merriam-Webster's Dictionary*, skill means "the ability to use one's knowledge effectively" or "doing something competently." In other words, skill implies a high level of human knowledge and execution. This common understanding of the word's meaning contributes to some confusion surrounding its use in the context of method skill.

Reserving Methods





Source: Milliman analysis based on data provided by SNL Financial Services



Source: Milliman analysis based on data provided by SNL Financial Services

If you always do what you've always done, you'll always get what you've always gotten.

In the original Old Norse language, the word "skill" meant to make a difference or a distinction. That would probably be the more accurate definition for our uses here. Method skill, first developed by meteorologists, rates the relative accuracy of different methods based on historical weather data by comparing the results of these predictions against the actual weather as it later manifests.

When it comes to P/C reserve development, the skill of a method at any particular development stage is calculated as 1 minus the mean squared error of the method divided by the mean squared anomaly of the data. "Anomaly" is measured between the accident years as the difference between the "actual" unpaid loss and the average unpaid loss for all accident years at the same evaluation. "Error" is measured as the difference between the actual unpaid loss and the indicated unpaid loss, as estimated by the method at hand.

The highest possible method skill rating for both meteorological and actuarial purposes is 1, with no limit as to how low into negative numbers a method can score. P/C reserving methods will frequently exhibit a negative skill number. This does not mean or imply that the method is without skill; the rating is entirely relative to the skill level of other methods. In other words, the comparison of the skill of two methods can be very meaningful, but the skills themselves have no such meaning.

In meteorology, where questions ("What will tomorrow's high temperature be?") result in a tighter range of possible outcomes than we face in reserving, method skill ratings are typically in the positive range between 0 and 1. But in the actuarial world of P/C reserve development, where significant uncertainty exists beyond what is represented in the historical data, method skill ratings are frequently negative.

A look at Charts 1 and 2 shows median skill levels across all lines of P/C business for two different methods.

Note that in Charts 1 and 2 the incurred-chain-ladder method, which accommodates for new loss data as they become available, improves in skill over time, while the loss-ratio method, based solely on a priori assumptions, exhibits increasingly lower skill numbers.

Results in general suggest that each reserving method follows one of these two intuitive patterns:

- Methods that respond to increasing knowledge of the given accident year tend to improve their skill over the first few evaluations, then reach a point when the skill levels off;
- In contrast, the skill of methods that fail to respond to this increasing knowledge becomes more negative as the accident year ages.

Several methods consistently outperformed the incurred chain ladder, including the following:

- Bornhuetter-Ferguson (incurred);
- Incremental additive IBNR (incurred but not reported);
- Backward recursive;
- Case reserve chain ladder;
- Hindsight IBNR.

Chart 3 shows the median skill, across all months of development and for all lines of business, for the incurred-chain-ladder

method and for the five methods listed above, considered together. Some variation in underlying results exists by method and line of business. However, the general observation that these methods have outperformed the incurred chain ladder (after 12 months of development) holds true across data sets and points in time.

For comparison, Chart 3 also includes the median skill of the paid-chain-ladder method. Despite its common use, the paid chain ladder is seen to significantly underperform the other methods.

Beyond the Familiar

The results of our study suggest that there are many more valuable methods for reserve analysis beyond the incurred- and paid-chain-ladder methods, and that the paid chain ladder, in particular, should not receive the weight it often does. Of course, this is a general observation, and a particular company's circumstances always should be considered in selecting methods for any reserve analysis.

The most significant finding of this study is that we all could improve our analyses greatly by focusing on methods that satisfy the two criteria mentioned above: relying, at least in part, on case reserves, and exhibiting no direct influence by amounts paid to date on indicated unpaid loss.

The only method in common use that satisfies both these criteria is the Bornhuetter-Ferguson method. Based on our analysis, greater weight should be given to this or similar methods than currently is typical.

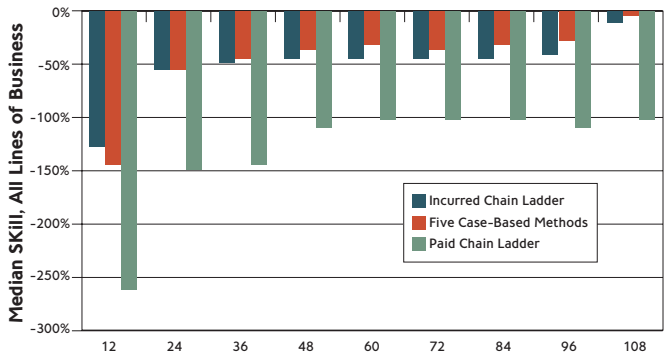
Chart 4 provides the relative skill of the Bornhuetter-Ferguson method, illustrating the effect that the use of multiple methods can have on an analysis. Consider that the Bornhuetter-Ferguson method is a weighted average of the incurred chain ladder and (in this case) an industry-based loss-ratio method, so that its result falls in between these two distinct approaches. Yet its skill exceeds that of either underlying method, despite the very poor performance of the loss-ratio method itself.

The ongoing challenge is identifying which new methods to select among the handful indicated and how to weight them against those more common methods already in our actuarial toolbox. This is an area for possible future work.

Whichever methods actuaries ultimately decide to use when performing reserve analyses, this study strongly suggests we should all consider methods beyond the familiar chain-ladder approach.

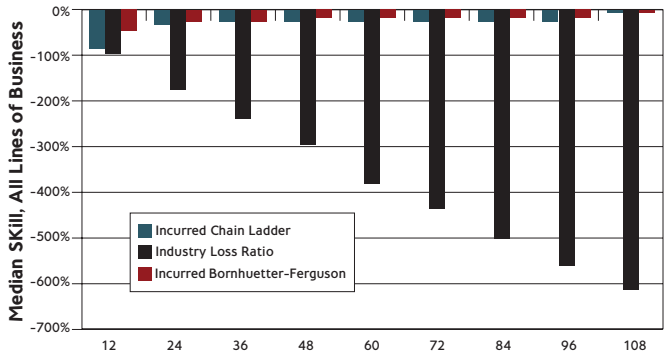
A quote variously ascribed to Henry Ford, Mark Twain, and, sometimes, Albert Einstein seems particularly appropriate in this context: "If you always do what you've always done, you'll always get what you've always gotten."

Chart 3: Comparison of Incurred, Paid, and Case-Based Methods



Source: Milliman analysis based on data provided by SNL Financial Services

Chart 4: Incurred Bornhuetter-Ferguson



Source: Milliman analysis based on data provided by SNL Financial Services

SUSAN J. FORRAY, a fellow of the Casualty Actuarial Society and a member of the Academy, is a principal and consulting actuary with the Milwaukee office of Milliman. She is grateful for the assistance of Shaun Cullinane, Paul Keuler, Linda Gritter, and Brooke Mote-Niroda with this work. Forray can be reached at susan.forray@milliman.com.

Resources

Jing, Yi, Joseph Lebens, and Stephen Lowe, "Claim Reserving: Performance Testing and the Control Cycle," *Variance,* 2009. http://www.casact.org/education/clrs/2009/handouts/GS1ClaimReserving.pdf

Forray, Susan, "Looking Back to See Ahead: A Hindsight Analysis of Actuarial Reserving Methods," CAS *E-Forum,* Summer 2012. http://www.casact.org/pubs/forum/12sumforum/Completesummer12.pdf