DISCUSSION OF PAPER PUBLISHED IN VOLUME LXXXIX

TESTING THE REASONABLENESS OF LOSS RESERVES: RESERVE RATIOS

C. K. KHURY

DISCUSSION BY CHARLES A. BRYAN

Abstract

Mr. Khury's paper advocates using various reserve ratios to test the reasonableness of loss reserve estimates. This review expands upon these ideas by discussing the practitioners for whom these techniques will be most useful, the practical decisions required to apply Mr. Khury's concepts, and a statistical technique to evaluate whether ratios derived from the loss reserve estimates are reasonable relative to other available data.

Mr. Khury has provided us a paper whose basic idea is "compilations of histories of reserve ratios are likely to reveal stable patterns that can be useful in testing loss reserves for reasonableness." Actuaries have long used reserve ratios as part of the process of constructing loss development factors, but have not generally used the ratios discussed in this paper to evaluate retrospectively the reasonableness of the reserves in the context of many years' experience and many development points. The paper is useful in that it adds to the actuary's collection of tools and techniques available to reach conclusions on the adequacy of loss reserves and the judgments made in setting those reserves. Mr. Khury's paper emphasizes that loss reserves should behave in a stable manner and that it is the job of the actuary to determine what that behavior is and to understand and explain any variance from the expected behavior. The purpose of this review

is to provide some historical context, comment on where this technique might be most useful, provide an approach that will make these techniques easier to apply, and evaluate some of the statements made in the paper.

The actuary has played the primary role in setting and evaluating loss reserves since the late 1980s when many of the requirements for actuarial opinions were implemented. The most significant impetus was the NAIC solvency agenda, which in turn was motivated by the United States House of Representatives *Failed Promises* report (A Report of the Subcommittee on Oversight and Investigations of the Committee on Energy and Commerce, February, 1990) and the A. M. Best *Insolvency Study: Property/Casualty Insurers 1969–1990*. Both these reports helped convince the NAIC and its members that strengthening the analysis of loss reserves is critical to improving the regulatory approach to solvency. The A. M. Best report noted, "Deficient loss reserves (intrinsically linked with inadequate product pricing) and rapid growth were the most dominant causes of insolvencies." To address this issue, the NAIC adopted requirements for an actuarial opinion backed up by an actuarial report. The Casualty Actuarial Society strengthened the Casualty Loss Reserve Seminar to respond to training needs in support of the primary role of the actuary. The Actuarial Standards Board adopted documentation standards that specified the type of information that must be in the supporting report.

It is helpful to remember that the setting of reasonable reserves for life contingency products is fundamentally different than setting reasonable reserves for property and casualty products. The methods for life contingency products emphasize standardized reserve requirements such as those embedded in the Commissioner's Reserve Tables. The methods for property and casualty products emphasize the training and judgment of individual actuaries and, in fact, there are no standardized tables against which P&C reserve adequacy is measured. Therefore, the first question the reader should ask is: "Does the Khury paper provide

some evidence that there may be a more standardized approach to judging the reasonableness of loss reserves?" My view is that the paper does point the way to introducing a useful set of standardized ratios that can flag where the reviewer of loss reserves should spend his or her time. Therefore, I classify these tools as important in an efficiency sense.

The second question, then, that the reader should ask is, "In what context and for what situations can these ratio tools be most useful?" The paper emphasizes that these ratio techniques are useful when reviewing reserves for reasonableness and not in setting reserves. This reviewer agrees with the author's conclusions and sees four key reasons why use of these techniques is not appropriate for setting reserves:

1. The historical information embedded in the ratios may not be the same as the most current information available, and reserve setting should use the most current information.

2. The process of setting the reserves must not be bound by what has happened in the past. The essence of a thorough setting of reserves requires the actuary to have no preconceived notion of what a reasonable reserve level will be and so not be influenced in judgments by an expected result.

3. The use of these techniques in setting reserves would encourage too high-level a view on the reserves and provide an excuse for inadequate analysis of the data.

4. P&C companies are sufficiently different from one another and even different over time that the use of historical ratios in setting reserves may not be accurate.

Even with these four limitations, it is appealing to consider situations where we know enough about a line of business or a company to at least put bounds around what reasonable results

are. I see seven persons/situations where it is most likely that these tests for reasonableness will be helpful:

1. The chief actuary responsible for the overall adequacy of the reserves: Typically, a large company has several independent business units and each of these business units has its own actuary who sets the reserves. The chief actuary receives reports, backup material, and reserve valuations from each business unit. The chief actuary can use these ratio techniques to provide a reasonability check on the valuations from the business units before accepting the estimates as reasonable.

2. Regulatory authorities: These authorities currently receive actuarial opinions and actuarial reports for each company but often have insufficient time to delve deeply into any particular reserve valuation until long after the end of the reporting period. The IRIS ratios are available to red flag any unusual historical development, but these ratios contain misleading information if the current reserve valuation is inadequate. I can envision a situation in which a company would submit a reserve ratio review report with its reserve valuation and would provide a mandatory explanation of any significant departures from the historical reserve ratios. This report would allow the regulatory actuary to focus on red-flagged situations first.

3. Rating agency analysts: It may be useful to request a reserve ratio analysis and an explanation of any outliers as part of the analytical information collected by the rating agency.

4. A legal proceeding, in which a court or an arbitrator must determine if there was a reasonable basis for questioning the adequacy of a prior reserve level: The factual situation might include a very inadequate reserve, appropriate data and correct calculations at the time the reserve was

set, no significant change in court decisions affecting the coverages provided subsequent to the reserve setting process, and no other substantial reason identified for the reserve deficiency. The ratio techniques can provide a view on the question "did the resulting reserve seem so unlike the expected reserve that additional analysis or skepticism was warranted?"

5. Internal Revenue Service agents: If the historical reserves are out of line with the ratios, then the agent may have a basis for beginning discussions with the company. In fact, it may be possible to automate the ratio tests and so help the agent focus on companies that may have overstated reserves.

6. External auditors assessing the reasonableness of the reserves: While the auditor generally has a specialist, usually an actuary, available to help in the reserve evaluation, it is beneficial for the audit partner himself or herself to have some techniques available to reach a view on the reserves. The calculation of the ratios is straightforward and requires no complicated mathematical formulas.

7. A company that has seen large reserve developments: This is the "crashed airplane situation" where we are interested in seeing if we could have detected problems prior to the severe financial distress.

With these contexts in mind, it is important to consider in detail the approach the paper suggests and how the author has selected historical data and the tests applied. My emphasis in this discussion is whether or not the selections the author recommends are available to and useful to the seven groups delineated above. This discussion requires digging deeper into the following two areas: 1) data selection for establishing benchmarks, and 2) using benchmarks to identify outliers.

1) Data Selection

The first area to examine is the question of what data and segmentation of data should be used for establishing a set of benchmark ratios. The choices to be made include (a) industry data or company data, (b) all lines combined or specific lines, (c) direct or net data, and (d) which ultimate estimates to use.

(a) Industry data or company data

While the company data may be more relevant to the company's particular situation, my election would be the use of industry data. Industry data has the following advantages:

1. The larger database should be more stable.
2. The ratio comparisons will provide insight into what may be different about the company, if anything.

Historical industry data normally requires using Schedule P information because of its standard format and availability for all licensed U.S. companies. Data can be obtained directly from the Annual Statements or from compilations such as those prepared by A. M. Best and other data service providers.

For most lines of business, the use of Schedule P data is perfectly acceptable. However, there have been changes in the Schedule P definitions of lines of business and expenses over time, and any analysis should consider these changes and comment on them. For example, the older Schedule P data used allocated loss adjustment expense. The newer Schedule Ps use the "defense and cost containment" categorization instead of "allocated loss adjustment expense." By-line applications will generally be much more useful than all lines, because it is easier to fit a company's situation with a by-line analysis than it is with an all lines analysis.

There are many situations in which ratios developed from Schedule P will not be useful. For example, many loss reserve

problems arise after the 10-year time period displayed in Schedule P. We are all familiar with environmental, asbestos, and other mass tort claims that arise more than 10 years after the policy was written. We also encounter this issue with long tailed workers compensation and unlimited PIP.

(b) All lines versus specific lines

Benchmark ratios can be easily established for Schedule P lines of business. Schedule P now contains detailed loss reserve development detail for many lines of business, and even a cursory review for reasonableness will be much more informative if done by line of business. By-line applications will generally be much more useful because it is easier to fit a company's situation with a by-line analysis. I suggest reviewing the largest lines of business to cover a pre-determined portion, say 80%, of the total business. The threshold can be established based on the importance of the review. For some companies, selecting one line of business may fulfill the 80% requirement. For other companies, fulfillment may require selecting and checking several lines of business.

(c) Direct versus net

Ratios should be calculated separately for direct, assumed, and net business. Such a calculation will reveal any unusual effects of reinsurance on the results. For example, if direct and assumed data provided values consistent with the model, but net data produced outliers, then one area of inquiry is the effect of reinsurance.

(d) Selection of ultimates

Where applicable, the most useful benchmarks would be ratios restated using the most recent estimate of ultimate losses, which should be the most accurate estimate. The latest estimate

is available from the most recent year Schedule P. While the restatement will not normally affect the premium, paid losses, or reported losses, it will affect the IBNR.

2) Using Benchmarks for Identifying Outliers

How should stability and acceptable ratio values be evaluated, and which values should be flagged as outliers? In layman's terms, "stable" can mean all the numbers look similar in a visual inspection. Such a definition may be satisfactory for some contexts, such as for a review by a chief actuary prior to signing off on reserves set by other actuaries.

However, other contexts require a more rigorous definition. Future papers can address the most appropriate way to define "stable" and "reasonable values." For purposes of this review, I will use a Normal distribution to identify outliers.

For example, Table C of the paper provides historical ratios of IBNR to premium for the Reinsurance industry. For 2 years of development and 10 sample years, we have the sample values 29, 27, 30, 28, 25, 27, 27, 28, 25, and 23. Visually, this appears to be a tight distribution of numbers. If we calculate a value of 20 for the most recent year, should that flag the reserve for further analysis? Although visually, the observer can see that the 20 falls outside the range of historical results, it is useful to have a mathematical test to apply so that we will have the maximum consistency on the decision. The test would allow us to set a tolerance level in advance of calculating the current year result and so avoid any bias.

I select the situation in which the 24-month ratios of IBNR to premium are distributed normally. The mean value for the sequence of numbers in Table C as stated in the above paragraph is 26.90. The corresponding sample standard deviation is 1.97. A common convention is to define outliers as those values that are more than two standard deviations from the mean. In this case, we would define as outliers any values less than 22.96 or greater

than 30.84. Therefore, a value of 20 is an outlier and we would red flag the most recent year as a year needing more analysis and explanation. Depending upon the line of business, and the context, other decision rules can be selected. For example, the decision rules should vary by the size of the database used in the calculation.

Of course, other distributions may be more appropriate than the Normal distribution. The user may elect a different distribution but the Normal distribution will usually be satisfactory for identifying outliers in a consistent manner.

I hope that the CAS or one of the paper's readers will take on the task of developing calculations and cut off points for selected distributions. It would be helpful and feasible to use the last 10 years for each Schedule P line that has 10 years of industry data to construct a table each year to be used in flagging outliers. The availability of such a table would advance the science of evaluating the reasonableness of loss reserves by setting scientific standards for the expected ratios and assuring that ratios outside the specified range would be investigated.

In conclusion, Mr. Khury has made a valuable addition to the loss reserve literature. These techniques and tools should provide some important benchmarks on the way to more accurate reserves. The real usefulness of this paper is in pointing the way to a different approach to the reasonableness review process. If we follow the lead provided by the paper, we can look forward to significant improvements in the ability to detect loss reserve accuracy problems in an organized and systematic way.