North American Actuarial Journal ›
Volume 23, 2019 - Issue 2

Feature Articles

It's About Time: An Examination of Loss Reserve Development Time Horizons

Michael M. Barth, Evan M. Eastman & David L. Eckles

Pages 143-168 | Published online: 08 Feb 2019

Cite this article https://doi.org/10.1080/10920277.2018.1538804

Abstract

A rich body of academic research has addressed the question of earnings management in the property-casualty insurance industry via manipulation of loss reserve estimates. This study analyzes the variability of reserve estimates at different development horizons to determine whether the predominant practice of relying on five years of development is appropriate. We examine two common measures of reserve estimation error, calendar year development and accident year development, and compare and contrast the two approaches. We also consider the appropriateness of the common practice of aggregating lines of business. After examining reserve development patterns for each of the major lines of business, we conclude that the appropriate development horizon to adequately establish ultimate liability may be longer than the current maximum reported horizon of 10 years found in Schedule P for most lines of business, including the aggregate reserves. Although longer-term development

horizons are necessary to establish insurers' ultimate liability, relatively short-term development horizons may be more appropriate when attempting to identify deliberate manipulations or to assess solvency risk, where the short-term variations are the primary object of interest. Ultimately, this article investigates the degree to which methodology originally developed for estimating loss reserve errors is appropriate today, in particular, relative to current data availability.

ACKNOWLEDGMENTS

The authors would like to thank two anonymous referees for their helpful comments that greatly improved the article.

Notes

1 An insurer's earnings are based (at least partially) on the losses the insurer reports. Increases (decreases) in loss reserves, therefore, decrease (increase) an insurer's earnings during the reporting period. We would consider any intentional altering of reserves to affect insurer earnings to be "earnings management."

2 Anderson (1971) referred to these as the loss reserve adjustment method and the loss reserve development method, respectively. We consider accident year development (AYD) and calendar year development (CYD) more expressive.

3 We appreciate an anonymous reviewer for pointing out these issues.

4 A reviewer pointed out that the use of five-year reserve errors also facilitates comparisons across studies.

5 For most of the lines, 10 years of loss history is provided for each individual accident year (AY) from AYt to AYt + 9. An aggregate line reports development for any remaining reserves from prior accidents years in Line 01 of the various parts of Schedule P. For some short-tailed lines (specialty property, auto physical damage, fidelity/surety, accident/health, financial/mortgage guaranty, and warranty) only two accident years of

history are provided, though a 10-year history of the aggregate short-tail lines can be derived by subtracting the other long-tail lines from the summary total for all lines.

6 Researchers have often used different scaling variables (premiums, reserves, incurred losses) to measure the relative amount of error between insurers. In our multivariate tests, we will scale by the developed reserve for our tests on the individual lines of business and by total assets for our tests on aggregate reserves.

7 Anderson used a one-year development, though, subsequently, most research has utilized a longer development horizon.

8 Schedule P has been completely redesigned since the time that Anderson conducted his study and now includes much greater reporting detail, more lines of business, and a full 10 years of development history (originally, Schedule P included only five years of development). Even with all the reporting changes, these two basic approaches—the AYD and the CYD—continue to appear in the literature today. Moreover, the one-year and two-year CYD measures are now incorporated into Schedule P Part 2 for each of the major lines of business.

9 Subsequent researchers have often labeled development from AY to AY + 4 as “five-year development” because there is some loss development during the accident year itself. However, Forbes specifically defined loss reserve development in the statement year as “no development” and referred to development from AY to AY + 4 as four- year development (Forbes 1970, 531). We consistently use the terminology of Forbes and Anderson throughout this article for consistency and refer to the development horizon as the number of years beyond the original accident year. Therefore, we refer to the development horizon from AY to AY + 4 as four-year development.

10 Schedule P was revised significantly in 1969 and again in 1989 to the current format. At the time Forbes was working with the data, the rudimentary Schedule P was all that was available.

11 It is also important to note that unlike Forbes (1970), KFS examined aggregate reserves and not the specific lines of business.

12 These papers primarily used four- or five-year reserve errors and relied on either the AYD or CYD method of reserve error calculation.

13 Petroni (1992) was using data published in AM Best's Insurance Reports, and that was the maximum development period available. The CYD and AYD approaches may reveal earnings management issues differently. We leave the mechanism out of this article for parsimony, but interested readers can contact the authors for an illustrative case.

14 The hypothesis is written such that rejecting the null would be in support of the Forbes Standard.

15 For the CYD calculation, the reserves are summed for each accident year within a calendar year; that is, the t subscript represents the sum of all reserves set in year t.

16 There are data limitations to the KFS study, particularly related to the lack of consistency in defining the ultimate losses; that is, instead of using AYt + 9 for ultimate losses, KFS were forced to use the latest year available. That means that for certain accident years, the "ultimate" developed reserves were established as early as AYt + 3. The KFS data set was also relatively small, with between 48 and 67 observations in each of the 10 available accident years from 1977 to 1987. Additionally, only three of those accident years (1977, 1978, and 1979) had a full 10 years of loss history, and hence we see the relaxation of the definition of "ultimate" in their research. Therefore, each of the other accident years in their study (1980–1987) had fewer than 10 years of development, and the ultimate losses were probably misstated. By contrast, we examine thousands of observations (depending on the line of business), and each of our 10 accident years is fully developed to AYt + 9.

17 KFS only consider the aggregate reserves. We consider the aggregate line as well as all individual lines of business.

18 Though some studies exclude reinsurers (e.g., Beaver, McNichols, and Nelson 2003; Grace and Leverty 2012), we have no reason to believe their behavior is any different and include reinsurers in our analysis. Our results are qualitatively unchanged when included firms with more than 25% of premiums written in reinsurance. Our results are also qualitatively unchanged when not excluding firms with reserves less than $1 million.

19 We note that in all tables "SUM" refers to the combined aggregation of all lines of business. The number of observations for this line will not be the sum of the observations from the other lines because not all insurers will write all lines.

20 These lines are also significantly greater than 97% at the 1% level.

21 Tables 2a and 2b report the mean paid to incurred ratios (and conduct statistical test on their difference relative to 97%). The results presented here are similar when conducting similar tests on the median paid to incurred ratio in that all lines except four take more than five years to develop to 97% or greater of “ultimate.” The four exceptions include three previously discussed (homeowners, private passenger automobile liability, and miscellaneous short tailed lines) as well as the combined line. In this case, the median paid to incurred ratio for the combined lines is statistically greater than 97% after five years of development. This last result will give some solace to those using a five-year reserve error period; however, we note that this relies on the median, and consistent differences still exist across lines.

22 Again, most of the extant literature examines reserve errors on an aggregated basis.

23 As an alternative measure of accuracy, we computed the proportion of insurer-year observations in each development year that reported developed accident year reserves that fell within 1% (or within 5%) of the ultimate developed reserve. In untabulated results (available from the authors), very few insurer-year observations achieve the 1% standard of accuracy for any given line of business or for their aggregate results. In fact, only 36% of insurer-year observations show a reserve within 1% of the final reported reserve by year 5. Even with the looser standard of having the estimate fall within 5% of the ultimate requires a significant development horizon, and that the accuracy differs greatly from one line of business to the next.

24 With the CYD measure, the measure of “ultimate” losses is more cumbersome since 10 years of development is the maximum amount available for individual accident years, and even that is not sufficient to fully develop most lines of business.

25 However, as noted earlier, the differences may be statistically significantly different from zero but still be materially insignificant. For example, in Table 4 the mean change in the aggregate reserve reestimation from $CYt + 3$ to $CYt + 5$ is 0.013, which is significantly different than zero, but only represents about a 1.3% change in reserves. Therefore, depending on the issue being studied, a justifiable conclusion may be that three years of development of the calendar year reserves is sufficient to get the reserves to “close enough” to ultimate value, with “close enough” being a purely subjective assessment that might satisfy some, but not all, researchers. Additionally, if the purpose of the research is to detect deliberate over and underreserving, the bulk of

the errors show up relatively quickly. Arguably, a one-, two-, or three-year horizon is sufficient to detect deliberate manipulations.

26 While we do not use the reserve error metric in Grace and Leverty (2012), we use their model as an overview of commonly cited reserve error-based incentives. See Barth and Eckles (2018) for a further discussion on the reserve error used in Grace and Leverty (2012).

27 We define the following lines as long-tailed: farm multiperil, homeowners, commercial, medical malpractice, workers compensation, products liability, auto liability, and other liability. This is consistent with the definition found in Hoyt and McCullough (2010), Eckles and Halek (2010), and Carson, Eastman, and Eckles (2018).

28 The Wald test allows testing the significance of coefficients across models. See Judge et al. (1985) for further discussion.

29 We appreciate an anonymous reviewer for this observation.