The Ignorance of Crowds: Understanding Loss Reserving Errors in the Liability Catastrophe of 1997-2001

Eren Cifci* Qianlong Liu† Lars Powell‡ Kenny Wunder§ George Zanjani¶

– PRELIMINARY DRAFT –

PLEASE DO NOT CITE OR DISTRIBUTE

Abstract

We estimate first-report company-level ultimate loss ratios by accident year and line using Chain Ladder and Bornhuetter-Ferguson techniques. We find that posted first-report loss ratios track closely with Bornhuetter-Ferguson estimates, and that these estimates are more precise than Chain Ladder estimates at the company level. However, when the estimates are rolled up to the industry level, the Chain Ladder roll-ups are significantly more accurate than both the Bornhuetter-Ferguson roll-ups and the roll-ups of the results actually posted. We interpret this as an application of the "ignorance of crowds," driven in this case by the reliance of companies on reserving methods that are essentially Bayesian updates keyed to highly correlated priors.

Keywords: Loss Reserves · Liability Insurance · Catastrophe

JEL Classifications: D22 · D25 · G22

*Department of Economics, Finance, and Legal Studies, Culverhouse College of Business, University of Alabama, ecifci@crimson.ua.edu.

†Institute of Insurance and Risk Management, Lingnan University, Hong Kong, qlliu@ln.edu.hk.

‡Alabama Center for Insurance Information and Research, Culverhouse College of Business, University of Alabama, lars.powell@cba.ua.edu.

§Department of Economics, Finance, and Legal Studies, Culverhouse College of Business, University of Alabama, kwunder@cba.ua.edu.

¶Department of Economics, Finance, and Legal Studies, Culverhouse College of Business, University of Alabama, ghzanjani@cba.ua.edu.

1 Introduction

Academics have focused much attention on the topic of property-casualty loss reserving over the past decades. Much of the interest has been centered on different aspects of reserving at the micro-level. Actuarial research, for example, features a rich tradition of innovation in reserving methods. Scholars in economics and accounting, on the other hand, have focused on reserving errors, particularly on the issue of the various incentives that individual companies face to misstate reserves.

In this paper, we examine the micro-foundations of reserving errors but with a different target in mind. Our primary interest is not in dissecting the individual errors of the companies themselves, but instead in understanding the aggregation of those errors into a collective error of the industry.

This collective error varies significantly across lines of business and over time. To illustrate the importance of this variation, Figure 1 shows the aggregate industry accident year loss ratio from Schedule P at first report and final report for several major commercial casualty lines – commercial auto, other liability, and workers' compensation. In general, the aggregate incurred industry loss ratios estimated at first report show much less time variation than the final result 10 years later.

Of particular interest to this paper are the 1997-2001 accident years, when the aggregated industry loss ratios in these lines at first report proved to be well short of the final result. The understatement reached its nadir in 1999, where the commercial auto, other liability, and workers' compensation loss ratios at first report were, respectively, 17%, 27%, and 24% short of the revised estimates 9 years later. In these three lines alone, the tally of the initial understatement of loss reserves for the 1997-2001 accident years exceeded $30 billion.

Why did the industry as a whole miss so badly? And is it possible to detect these sorts of errors as they are happening?

We argue that the roots of the problem lie in the reserving methods used by actuaries, particularly the reliance on methods, such as the Bornhuetter-Ferguson method (Bornhuetter and Ferguson, 1972), that are relatively unresponsive to changing conditions. There are compelling reasons for using such methods at the company level. Alternative reserving methods that are highly responsive to emerging experience, such as the Chain Ladder method, produce estimates that are sensitive to random fluctuations in the timing of claims and thus lead to unwelcome volatility in reported results. Bornhuetter-Ferguson estimates, on the other hand, essentially incorporate emerged experience into an ex ante expectation of outcomes to produce a Bayesian update of a prior forecast.

When these ex ante expectations are constructed from past experience at a company or industry level, however, the information contained within the current experience, i.e. as losses related to the accident year are paid over time, is suppressed in the revised estimate. This suppression of current private information at the company level can cause the typically assumed properties of

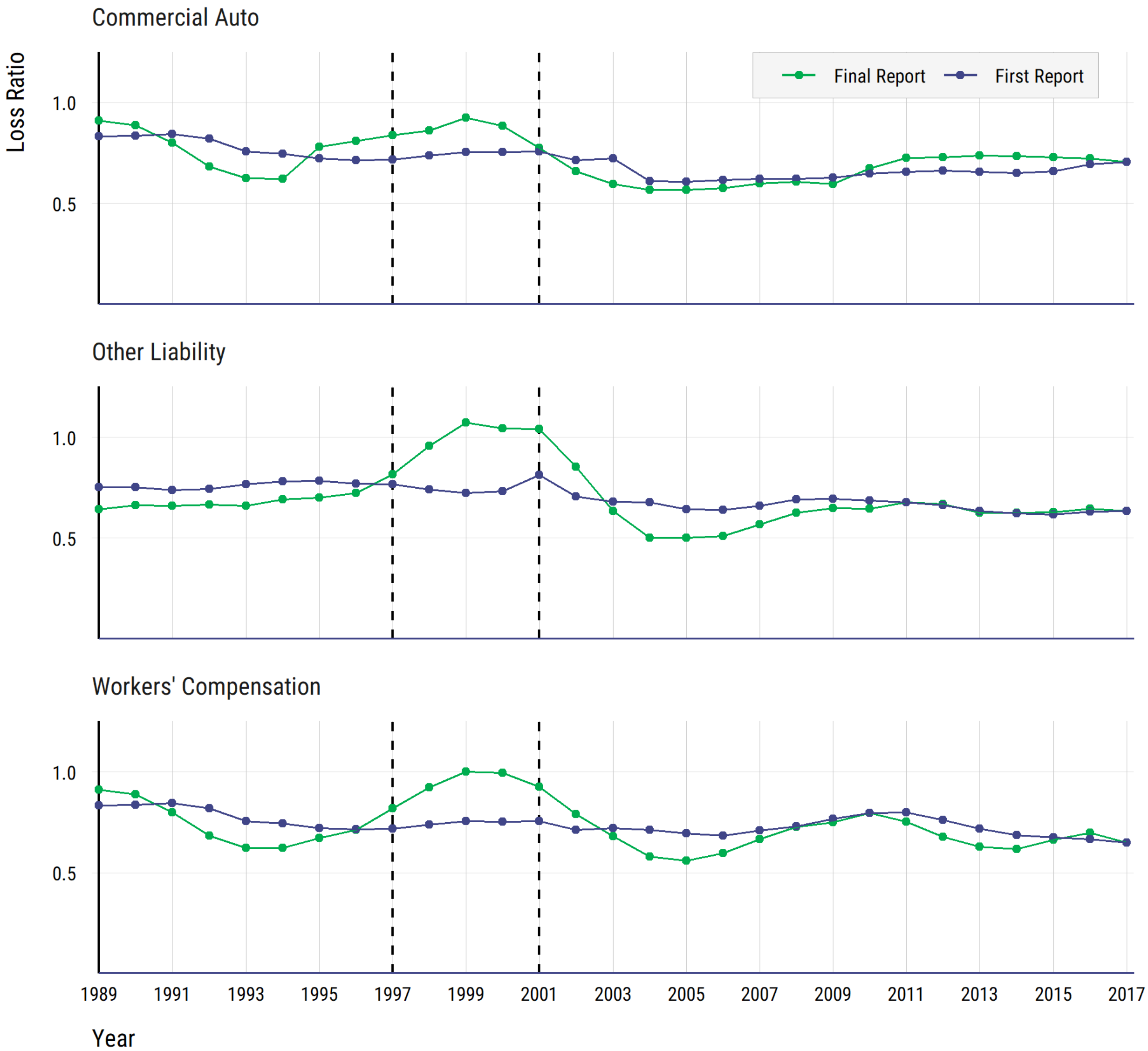

Figure 1: Comparison of first and final reports for the aggregated industry.

statistical aggregation to fail. Specifically, the "wisdom of crowds" depends on aggregating private signals and the averaging out of errors from independent individual forecasts. If the individuals are already weighting their forecasts together with highly correlated prior expectations, the crowd may not be so wise. In the extreme, the crowd's "wisdom" could boil down to a commonly held prior expectation which is entirely blind to emerging loss development.

In this paper, we use actual company Schedule P loss triangles to develop two types of loss reserve estimates based on observed losses occurring within the initial accident year. The first is a Chain Ladder estimate based on link ratios derived from the company's loss triangle. The second is a Bornhuetter-Ferguson estimate, where the company's emerged experience is blended with an expected loss ratio (in Bayesian style) to produce an estimate. We experiment with two types of expected loss ratios: (1) the industry loss ratio for the previous accident year and (2) the average loss ratio for the company over the three accident years preceding the year of estimation.

We show that the posted company reserves tend to track with Bornhuetter-Ferguson estimates and, furthermore, that these Bornhuetter-Ferguson estimates forecast ultimate results at the company level with greater precision (in the sense of lower mean squared error) than Chain Ladder estimates. However, this difference in precision reverses when the company-level data are aggregated into collective loss triangles at the line level within the industry. When applied to line-level aggregate triangles, Chain Ladder estimates tend to be more informative than Bornhuetter-Ferguson estimates. In particular, Chain Ladder estimates are better at picking up the deterioration in results in long-tailed lines during the late 1990's and early 2000's.

The rest of this paper is organized as follows. In Section 2, we review related literature, particularly the literatures on reserving errors in property-casualty insurance and statistical aggregation. In Section 3, we lay out the basic theoretical argument regarding the impact of Bayesian updating on the wisdom of crowds. In Section 4, we explain the basic actuarial reserving methods, highlighting the role of prior information embedded within them. We present our empirical analysis in Section 5. Section 6 concludes.

2 Related literature

Unlike participants in other industries where the cost of goods sold is known at the time of sale, insurers must estimate the ultimate liabilities arising from policies written well in advance of cost observation. As insurance contracts are sold to policyholders, insurers create capital reserves to ensure that sufficient capital will be available to pay losses as they emerge. The difference between the initial estimation of the reserve and the final–or ultimate–payment for losses is the reserve error. A large literature studies reserving errors in the property-casualty insurance industry, examining a variety of motivations for manipulating the initial reserve value.

Early foundations for this work were laid by Forbes (1970) and Anderson (1971), who defined reserve error metrics that were used and refined in the literature that would follow. Income smoothing, as well as tax incentives in some cases, were stressed as motives for reserve misstatement in Smith (1980), Weiss (1985), and Grace (1990). Clarification on and best practices for using reserve error metrics were discussed by Kazenski et al. (1992), with a later contribution in this line being Grace and Leverty (2012). Hsu et al. (2019) connect earnings management incentives to aspects of corporate governance.

An accounting literature starting with Petroni (1992) focused attention on balance sheet management (also see Gaver and Paterson (2004)) as a motivation for reserve manipulation. Beaver et al. (2003) stress GAAP earnings management. Nelson (2000) pointed to strategic positioning with respect to rate regulation. Given the focus on the bottom line of the balance sheet or GAAP income statement, the accounting literature tends to focus on calendar year reserve development (Barth and Eckles, 2018), while other papers focus on accident year development for a particular line of business, such as commercial auto.

The literature on statistical aggregation has shown that the wisdom of crowds can be reduced when individuals are making use of public information. For examples, Goldstein and Yang (2019) and Da and Huang (2020) both show that crowd-based measures (price, in the case of Goldstein and Yang; average earnings estimates in the case of Da and Huang) can become less informative after public information disclosures. Our mechanism is similar to those envisioned in these papers. We consider the historic statutory filings of insurers to be public information.

Our paper's application of reserving methods also connects to a vast actuarial literature on loss reserving techniques. We will not attempt to enumerate this literature in detail here, but instead note the availability of survey articles. For examples, Schmidt and Zocher (2007) offers a comprehensive survey of traditional methods as well as more recent extensions, while Taylor (2019) surveys recent developments in loss reserving approaches using machine learning.

3 Statistical aggregation and Bayesian updating

When information (particularly stable, credible information) regarding the underlying distribution of individual outcomes within a group is available, individuals may benefit by learning from the group's collective experience. This additional information allows for increased precision in outcome estimation, but it comes at a cost. If an individual is an unwitting outlier within the group, the shared information may mislead them, creating a biased though less noisy expectation. The ignorance of crowds can occur when individuals within a group share a commonly held or highly correlated prior belief that does not accurately reflect the distribution of individual outcomes. In such a scenario, we can imagine all participants to be outliers, in which case all expectations are biased and most (if not all) individuals may be worse off than if the shared information had not been observed.

To demonstrate this outcome, we consider a simple setting where each individual i generates an observation e_i of a population parameter μ:

$$e_i = \mu + \epsilon_i, \tag{1}$$

where ϵ_i is an error term. If we assume that the error term has mean zero and finite variance σ^2, the aggregation of n independent observations will produce a wisdom of crowds in the sense that

$$\frac{1}{n}\sum_{i=1}^{n} e_i = \mu, \tag{2}$$

wherein the individual error ϵ_i has disappeared as a result of the aggregation.

The situation changes, however, if individuals are endowed with a prior distribution of the population parameter. For simplicity, we assume that individuals share a common prior belief that μ is distributed with mean μ_0 and variance σ_0^2. If both the prior distribution and the error term in the individual estimate are assumed normal, the individual will update their estimate of the population parameter as a weighted average of the individual observation and the mean of the prior, as in

$$\hat{\mu}_i = \left(\frac{\sigma^2}{\sigma^2+\sigma_0^2}\right)\mu_0 + \left(\frac{\sigma_0^2}{\sigma^2+\sigma_0^2}\right)e_i. \tag{3}$$

Several insights can be gleaned from Equation 3. First, if the true value of the population parameter μ does not align with the mean of the prior distribution μ_0, the updated estimate will be biased. That is, the error in the initial expectation carries through to the estimation. Even with an infinitely large sample, there will be no convergence to the true population parameter, i.e.

$$\lim_{n\to\infty}\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\mu}_i\right) \neq \mu. \tag{4}$$

Second, if individual observations are noisy in relation to the prior distribution (i.e., σ^2 is much larger than σ_0^2), then the Bayesian update will be heavily weighted toward the mean of the prior distribution, μ_0. In that case, the individual observation will be de-emphasized when forming the Bayesian update. Importantly, when noisy individual observations collide with an incorrect belief about the true underlying parameter, as is the case when an unexpected (and uncertain) loss shock occurs, individuals are driven confidently toward a flawed expectation.

4 Background on loss reserving methods

There is an extraordinarily wide variety of reserving methods in use by actuaries, and the typical approach to reserving at the company level will not rely entirely on any one single method. Nevertheless, two methods in particular stand out among the crowd in terms of actuarial practice. A global survey of insurers revealed that Chain Ladder and Bornhuetter-Ferguson are the two dominant methods in use, with all respondents in the United States particularly indicating that both were "main" methods used in their reserving.[1] Brief descriptions of these two methods follow.

[1]2016 Report – ASTIN Working Party on Non-Life Reserving Practices

The Chain Ladder technique relies on applying a factor-to-ultimate (FTU) to an emerged, i.e. observed, paid (or paid-plus-case reserve) loss figure to project that emerged figure to an expected ultimate value. The difference between the projected ultimate loss value (the expected final loss) and the emerged figure (the observed loss amount at the time of projection) is the reserve, which reflects bulk and case reserves (in the case of an emerged paid figure) and bulk reserves only (in the case of an emerged paid-plus-case figure). The FTU is derived by analyzing a loss triangle that depicts the development of emerged losses for a set of accident years over time. For each accident year, development factors are calculated by dividing the emerged loss at each report date by the emerged loss at the previous report date.[2] The FTU is then calculated by multiplying all of the factors together (i.e., "chaining" successive development factors together), which, when applied to an emerged loss at an appropriate stage of development, produces an estimate of the ultimate loss for that accident year.

In mathematical terms,

$$U_k^{CL} = L_{k,t} \cdot F(t), \tag{5}$$

where U_k^{CL} is the Chain Ladder projection of ultimate losses attributable to accident year k, $L_{k,t}$ is the t^{th} period emerged loss value for accident year k, and $F(t)$ is the projection factor (i.e. the FTU) from development period t to ultimate.

The Chain Ladder method thus leverages emerged experience fully, which can result in volatile estimates of ultimate losses if there is significant randomness in the timing of loss payments in the emerging experience. That is, if there is any measurement error in the emerged loss payment, that error is amplified by the FTU in the projection of ultimate loss. To demonstrate, consider an insurer that observes emerged loss $L_{k,t} = e_i$. Combining Equation 1 and Equation 5 yields

$$\begin{aligned} U_k^{CL} &= (\mu + \epsilon_i) \cdot F(t) \\ &= \mu F(t) + \epsilon_i F(t), \end{aligned} \tag{6}$$

where ϵ_i is the randomness in timing. The final term in Equation 6 is essentially amplified noise.

The polar opposite of the Chain Ladder method is the Expected Loss Ratio (ELR) method, in which an expected (ultimate) loss ratio is applied to the premium in the line to produce an estimate of the ultimate loss. In contrast to Chain Ladder, the ELR method is completely insensitive to emerged experience, with the underlying assumption being that the ultimate loss is implicitly known ex ante (by virtue of knowing the expected ultimate loss ratio and the premium). Specifically,

[2] Modifications to this simplified explanation of Chain Ladder development factors exist and are considered in our analyses. In particular, we consider premium weighted averages of development factors as a robustness check.

$$U_k^{ELR} = \pi_k \cdot P_k, \tag{7}$$

where U_k^{ELR} is the ELR projection of ultimate losses attributable to accident year k, π_k is the ex ante expected loss ratio for accident year k, and P_k is the premium for year k. Note that Equation 7 contains no reference to development period t, in general, and no reference to observed loss $L_{k,t}$, in particular.

The Bornhuetter-Ferguson (BF) method can be interpreted as a hybrid of the Chain Ladder and ELR methods. The BF method is similar to the Chain Ladder method in that it uses historical loss data to determine appropriate FTUs and fully accepts emerged losses at face value. However, unlike the Chain Ladder method—and similar to the ELR method—the BF method incorporates an ex ante expectation of ultimate losses. The BF method diverges from the Chain Ladder method further in that it does not assume that the pattern of deviation in emerged losses relative to expectation contains any information about future loss payments; on the contrary, future loss payments are assumed to follow according to the ex ante expected loss ratio. That is,

$$U_k^{BF} = L_{k,t} + \left(1 - \frac{1}{F(t)}\right) \cdot \pi_k \cdot P_k. \tag{8}$$

Conceptually, the BF method begins with a prior belief about the ultimate loss outcome and revises that belief as losses emerge. This allows a recognition of deviation from expectation while avoiding the noise amplification in the Chain Ladder method (Equation 6). If the observed loss in development period t differs from the loss which, when projected to ultimate via the FTU would yield the prior expected ultimate outcome, the expectation of ultimate loss is revised by the amount of the difference. This back-projection of expected outcome takes the place of μ in Equation 6. Deviation from this expectation in any single period is regarded as the noise term, ϵ_i, and is not considered indicative of deviations in future development periods. Thus, the BF method is easily interpreted as a form of Bayesian updating, with the expected loss ratio being the mean of the Bayesian prior.

In practice, the appeal of the BF method comes from its reduced sensitivity (relative to the Chain Ladder method) to randomness in the timing of emerging loss payments. On the other hand, the resulting reduction in variance is accompanied by the introduction of a bias in the case that $\mu \neq \mu_0$. For this reason, the ability of the Bornhuetter-Ferguson method to provide meaningful estimation of ultimate loss outcomes is critically dependent on an accurate value of π_k.

5 Empirical analysis

Applying the theory of Section 3 to the case at hand, Chain Ladder estimates are the analog of the individual-level observations, while the BF estimates are the equivalent of Bayesian updates given a prior expectation. This suggests three questions of empirical inquiry:

1. Are company reserve estimates best explained by Bornhuetter-Ferguson or Chain Ladder methodologies?

   If companies are in fact following a Bayesian approach to reserving, the posted reserves should track more closely with the method rooted in Bayesian analysis, which is Bornhuetter-Ferguson, assuming that we can identify a reasonable proxy for the expected loss ratio.

2. Are final results at the company level (i.e. fully emerged losses observed 10 years after the accident year) best forecast by Chain Ladder estimates or Bornhuetter-Ferguson estimates?

   It is important to understand that the Bayesian update, in theory, is rational from the individual company's perspective. That is, it provides the best estimate given the assumed information. As a Bayesian methodology, we would expect BF to provide a better forecast than the Chain Ladder method at the company level. The reserves companies actually post, given the additional information available to them beyond what we are able to reconstruct using triangles pulled from their statutory filings, should align with ultimate outcomes even more closely.

3. Are final results at the industry level best forecast by the roll-up of Chain Ladder estimates, the roll-up of Bornhuetter-Ferguson estimates, or the roll-up of the reserves companies actually posted?

   Despite the predicted superiority of Bornhuetter-Ferguson and/or actual posted company results, the ignorance of crowds under Bayesian updating, as described in Section 3, should lead Chain Ladder estimates – an unbiased though noisy representation of the signal that a company receives – to provide a better estimate of the industry average when aggregated across all companies.

5.1 Industry-level analysis

In contrast to aggregating individual company estimations, we begin by analyzing aggregated company data at the industry level. The basic objectives here are twofold. First, this exercise provides a sense of the usefulness of formulaic reserve estimation when applied only to the limited data available through insurers' individual statutory filings. Second, because aggregated data is inherently less noisy, it provides a conceptual starting point from which we can assess the value of applying a Bayesian methodology. That is, Equation 2 holds in the aggregate.

The sources for the information are various editions of A.M. Best Aggregates & Averages – Property-Casualty going back to 1990. Each year, A.M. Best aggregates the various pages of the Statutory Annual Statements filed by all companies, including all parts of Schedule P. As a result, we have, or are able to create, paid, paid-plus-case-reserve, and incurred loss triangles for all of the Schedule P lines. We also have information on the aggregated industry estimate of the ultimate loss for each accident year for each development year ranging from the first evaluation of the accident year to the tenth evaluation of the accident year (for those accident years for which a tenth evaluation has been released).

We calculate Chain Ladder and BF final loss estimates for each accident year-development year combination as follows. For a given line and evaluation year, we take the paid triangle for that evaluation year from Part 3 and extend it by adding 4 additional rows of accident years from prior evaluation periods. For example, the triangle from the 2008 evaluation year would cover accident years 1999-2008. To this triangle we add the 2007 evaluation of the 1998 accident year, the 2006 evaluation of the 1997 accident year, the 2005 evaluation of the 1996 accident year, and the 2004 evaluation of the 1995 accident year. With this extended triangle, we calculate development factors and have sufficient data to calculate a five year average for each period-to-period development factor and also a five-year average for a "tail" factor, which develops the cumulative paid amount at 10^{th} evaluation to ultimate. The latter factor is calculated by taking the corresponding entries for the incurred amounts from Part 2, and dividing by the paid amounts from Part 3 at the 10^{th} evaluation. Chaining together the development factors thus yields FTUs that can be applied to paid losses from the Part 3 triangle to produce estimates of ultimate loss via the Chain Ladder method. The same FTUs are used in the application of the BF method. For the expected loss ratios, i.e. the Bayesian priors, we use the ultimate incurred loss ratio from the previous evaluation for every accident year except the most recent (which, of course, does not have a previous evaluation). For the expected loss ratio to be used with the first evaluation of an accident year, we use the ultimate incurred loss ratio from the previous evaluation of the previous accident year. For example, for the 2004 evaluation of the 2004 accident year, we would use the ultimate loss ratio for the 2003 accident year estimated as of 2003 for the expected loss ratio.

We present the results of this exercise in Figure 2 for the 3 lines discussed in Section 1: workers'

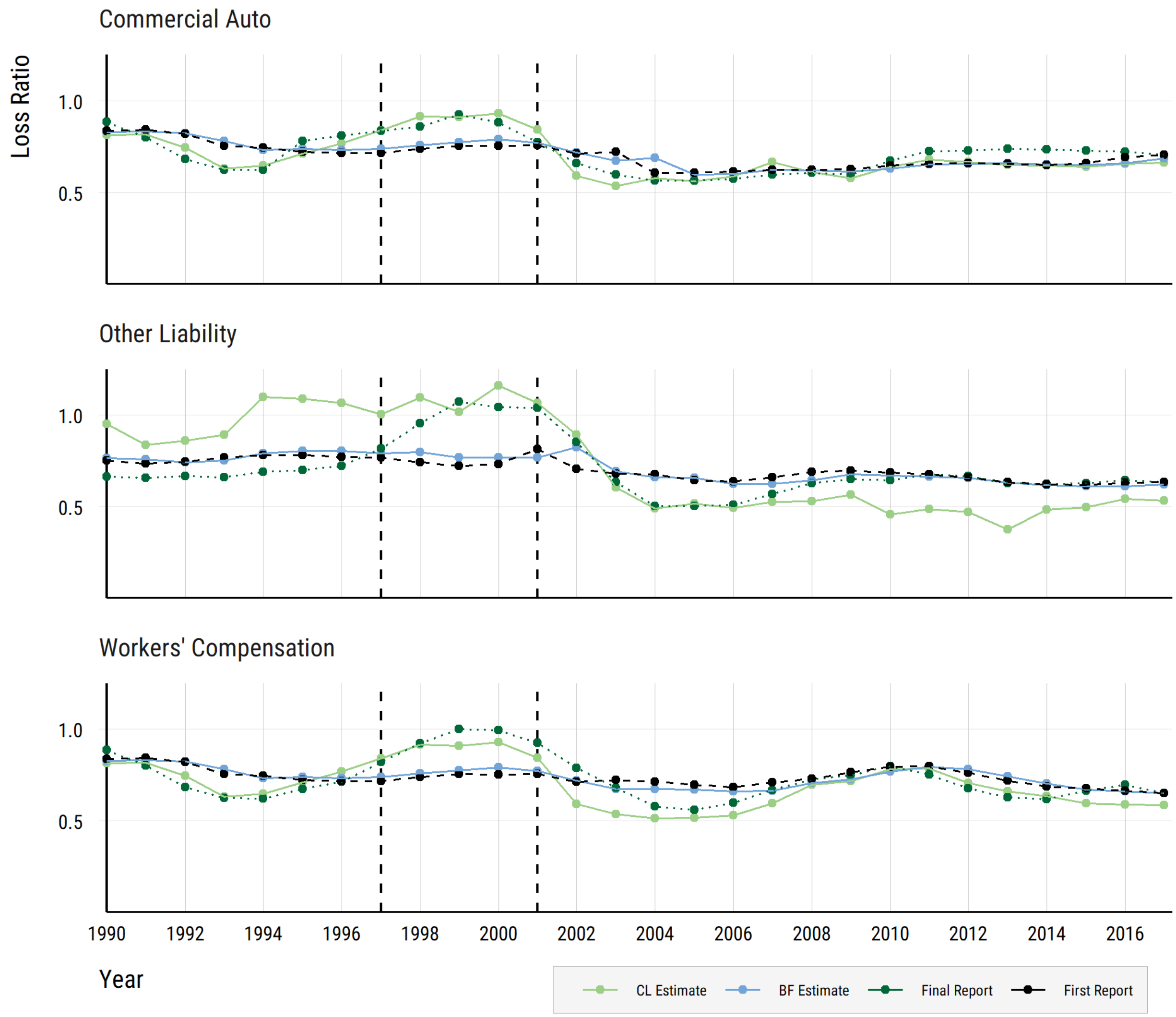

Figure 2: Comparison of industry ultimate loss ratios to posted and estimated (BF and CL).

compensation, commercial auto, and other liability. For workers' compensation and commercial auto, the story is clear: the initially posted ultimate estimates at first report track very closely with the BF ultimate estimates, while the 10^{th} report (Final) ultimate estimates track very closely with Chain Ladder (CL). The story for other liability is somewhat similar, in that the initially posted estimates of ultimate are fairly stable over time and track closely with BF estimates. The final estimates are much more volatile, as are the Chain Ladder estimates – which, in some years but not others, track closely with the final results. Some of the difference in other liability could be due to its changing composition in terms of claims made versus occurrence coverages over time.

5.2 Company-level and roll-up analysis

We now turn to the question of whether the patterns suggested by the analysis at the industry level mirror the patterns at the company level. On the one hand, the industry-level analysis is revealing in the sense that a simple Chain Ladder methodology applied to the first report of an accident year often gives an accurate forecast of the final ultimate estimate. It is also suggestive that estimates produced by a simple BF methodology, adapted with a plausible (albeit naïve) method for choosing the expected loss ratio, would match closely to the initially posted ultimate estimates. It is important, however, to recognize that actual reserves are produced at a company (or group) level, not at the industry level. So it is necessary to investigate whether these simple methodologies hold any explanatory power at the micro-level.

For the company-level analyses, we use Schedule P data directly from the National Association of Insurance Commissioners, which contains the statutory filings for all insurers from years 1993-2018. We consider affiliated and unaffiliated single insurers. We limit our data to stock and mutual companies that write more direct business than reinsurance. We further limit the data by requiring that each included insurer reports complete and consistent triangles in all parts of Schedule P. As was the case with the earlier aggregate analysis, we are able to create, paid, paid-plus-case-reserve, and incurred loss triangles for all of the Schedule P lines. The final requirement is that each insurer-line analyzed must also be present in the data 9 years later (at the 10^{th} development period) so that we are able to observe the ultimate loss ratio and determine the error in the initial reserve estimation.

Consistent with our two year extension of the Schedule P triangles, we use three year averages to develop FTUs. The FTUs are based on paid data from Part 3. For the Bornhuetter-Ferguson expected loss ratios, we use averages of the previous ultimate loss ratio estimates from the previous three accident years rather than just the previous ultimate loss ratio estimate from the previous accident year (as was the method in the industry-level analysis). The reasoning for this change is that the loss ratios at the company-line level are likely to be much more volatile.

The reserving methodology is similar to that used for the industry level with an exception to account for the lower volume of data at the company level. We extend our Schedule P triangles historically by three years, and we use the three year average of ultimate incurred to 10^{th} development paid in order to determine appropriate company-level tail factors. This adaptation requires that we observe company-level data from the three years directly preceding the year of estimation. As a result, we are unable to estimate reserves for the first three years of our time frame. For the BF expected loss ratios, we again use averages of the ultimate loss ratio estimates from the previous three accident years, as was the case in the aggregate analysis.

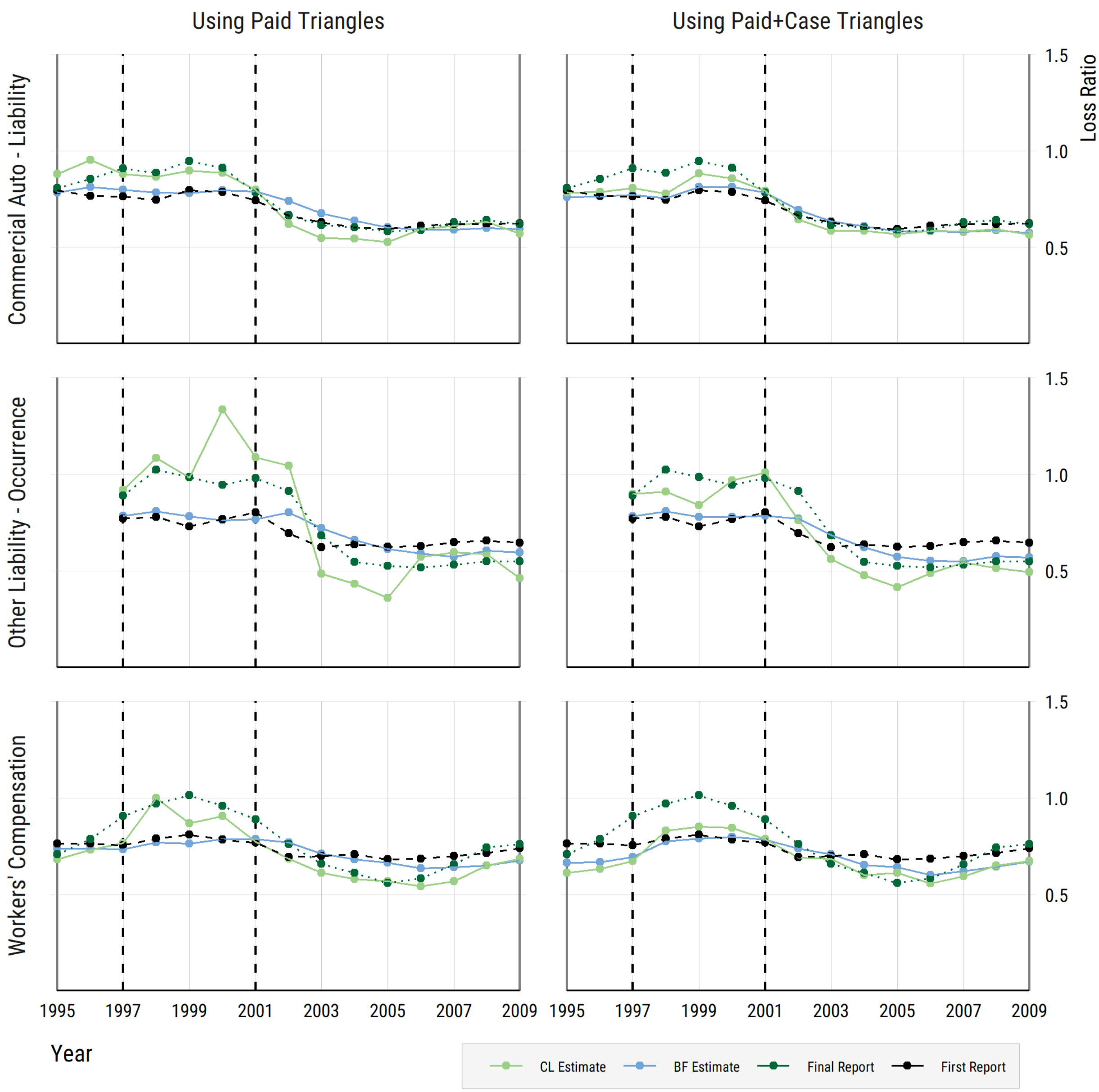

Figure 3: Comparison of industry roll-up ultimate loss ratios to posted and estimated (BF and CL).

After making all company-level estimations, we aggregate the company-level outcomes into a single industry roll-up estimate for each line of business within each year. This differs from the industry aggregation described in Section 5.1, in that the roll-up is an aggregation of estimates, while the earlier analysis is an estimation based on aggregated data.

Figure 3 shows the industry roll-up loss ratios at first and final reports as well as the final estimations determined under both Chain Ladder and BF methods. The story is similar to what was seen in the industry-level figures.

The final ultimate outcomes appear to track more closely with Chain Ladder estimates than with

BF estimates in all lines, especially in workers' compensation and commercial auto. Furthermore, at least in the cases of these two lines, the roll-up of the Chain Ladder estimates even outperforms the roll-up of the initial estimates actually posted across all insurers. The roll-up of estimates actually posted for the industry track very closely with the BF roll-ups in all three lines.

The forecast errors associated with the BF, Chain Ladder, and company reported estimates in relation to the final (10^{th} report) loss ratio, as well as the forecast errors between our BF and Chain Ladder loss ratio estimates and the actual posted loss ratio are shown in Table 1. The top panel reports the results at the individual company level, and the bottom panel reports the forecast errors associated with the industry level roll-ups.

	Company Level					
	Workers Compensation		Commercial Auto - Liability		Other Liability - Occurrence	
	Paid	Paid+Case	Paid	Paid+Case	Paid	Paid+Case
Posted vs Final	0.387	0.387	0.302	0.302	0.502	0.502
BF vs Final	0.394	0.393	0.303	0.296	0.497	0.493
CL vs Final	0.495	0.403	0.327	0.303	0.708	0.549
BF vs Posted	0.106	0.128	0.1	0.104	0.13	0.137
CL vs Posted	0.349	0.183	0.205	0.145	0.584	0.326
	Industry Roll-up					
	Workers Compensation		Commercial Auto - Liability		Other Liability - Occurrence	
	Paid	Paid+Case	Paid	Paid+Case	Paid	Paid+Case
Posted vs Final	0.113	0.113	0.078	0.078	0.157	0.157
BF vs Final	0.119	0.12	0.073	0.075	0.13	0.121
CL vs Final	0.079	0.111	0.049	0.055	0.146	0.09
BF vs Posted	0.041	0.059	0.035	0.027	0.06	0.061
CL vs Posted	0.1	0.084	0.086	0.043	0.262	0.149

Table 1: Root Mean Squared Errors (RMSEs) for ultimate loss ratio comparisons.

The table illustrates the prevalence of smoothing at the company level. As shown in the last two lines of both panels, our BF estimates provide a much closer approximation to the estimates actually posted by companies, both at the company level and at the roll-up level. The evidence here gives a clear indication that company reserve estimates follow the Bayesian-like approach of Bornhuetter-Ferguson.

In terms of accuracy in relation to the final loss ratio result at 10^{th} report, the picture is more nuanced. The BF method results in estimates that are more precise in all cases at the company level save for one (Workers Comp, Paid+Case). The loss ratio estimate actually posted by the company is the most accurate in all cases but one (Commercial Auto, Paid+Case). These results are generally supportive of the individual rationality of the Bayesian approach to reserving explored in Question 2. That is, the reserves actually posted by companies are closer to BF than to Chain Ladder, and those estimates also tend to align more closely with the ultimate company level outcomes.

The bottom panel, however, clearly shows the paradoxical ignorance of crowds from Question 3. The Chain Ladder roll-ups outperform not only the BF roll-ups, but also the roll-ups of the actual posted company estimates. As suggested in Section 2, we can expect the ignorance of crowds to be most pronounced in cases where (1) the error in individual observations (i.e. σ^2 in Equation 3) is significant in relation to perceived variance of the Bayesian prior and (2) where the difference between the "truth" and the prior expectation (i.e. the difference between μ and μ_0, the bias) is significant. On the second point at least, we can measure the difference between the "true" outcome for an accident year (at 10^{th} report) and a plausible proxy for the prior expectation—the posted industry loss ratio from the preceding accident year.

Figure 4 plots the ratio of the Chain Ladder RMSE to the Posted RMSE, with each point representing the ratio for a given accident year-line combination. Ratios below 1 imply that the roll-up of our Chain Ladder estimates was more accurate than the roll-up of the posted estimates for that line and accident year.

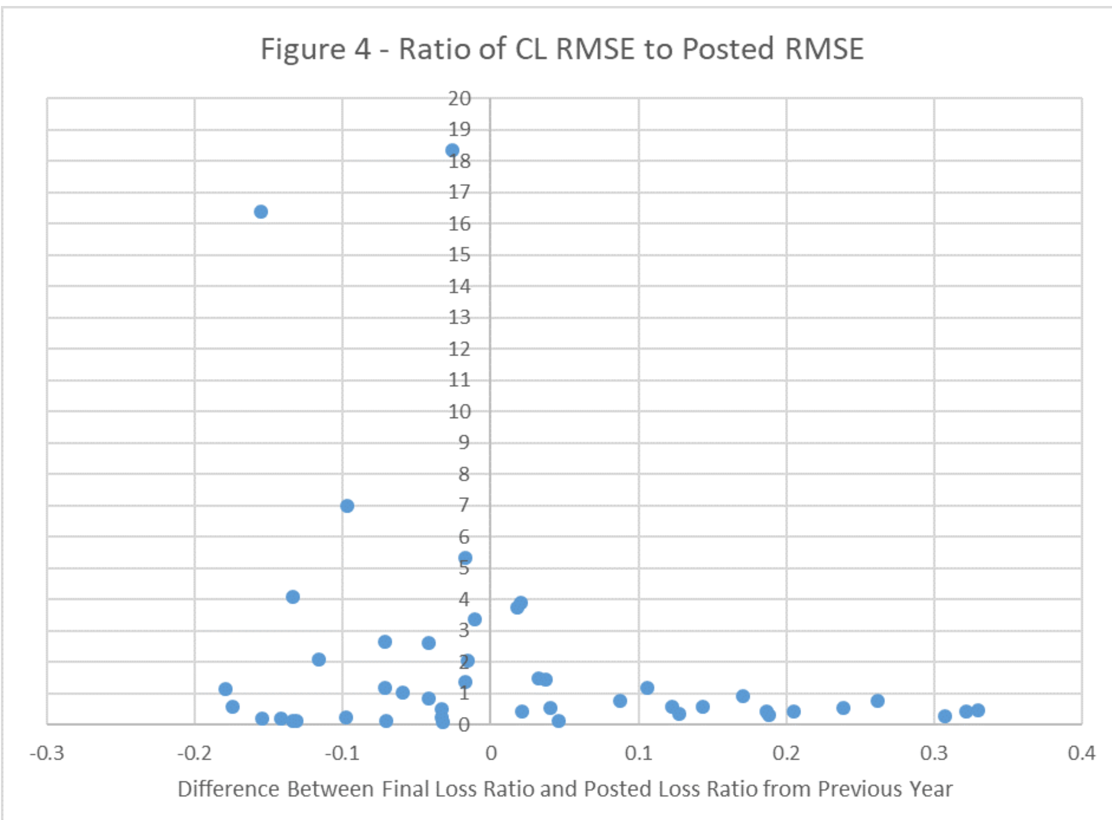

Figure 4: Ratio of CL RMSE to Posted RMSE.

As can be seen, larger differences, especially in cases where the "truth" was significantly worse than the expectation (i.e., large positive differences between the final loss ratio and the posted loss ratio from the previous year), were associated with RMSE ratios below 1—so that the Chain Ladder estimate roll-ups were superior to those actually posted. Put differently, the ignorance of crowds was most consistent and most pronounced when the truth was much worse than expected—which was, of course, the case during 1997-2001.

6 Conclusion

Consistent with the findings in other financial settings (Goldstein and Yang, 2019; Da and Huang, 2020), crowds can become more ignorant when public information is being referenced by individual when making forecasts. The information is undoubtedly valuable to the ones making the estimates–the individual estimates are improved by the incorporation of public information, and the Bayesian update is the best forecast for the individual company. Unfortunately, a consequence of this process–when viewed from the perspective of the collective industry–is that a great deal of information is lost.

The consequence of this is potentially severe, as industry aggregates are important inputs for assessing reserves at both the company and industry levels. To the extent that reported industry aggregates are relied upon for assessing reserve adequacy, regulators and the companies themselves may be handicapped in recognizing the potential for adverse development.

Fortunately, we have shown that this problem can be addressed through methods that are both conceptually simple and easily implemented. Chain Ladder, implemented at the industry level, can provide regulators with a simple and useful indicator of reserve adequacy at the accident year level. For companies, the challenge will be to find ways to integrate industry-level feedback into meaningful and practical adjustments to company-level reserve estimates.

References

Anderson, D. R. (1971). Effects of under and Overevaluations in Loss Reserves. *The Journal of Risk and Insurance*, 38(4):585.

Barth, M. M. and Eckles, D. L. (2018). Anatomy of an Error: The Case of Loss Reserve Development Error Research. *Journal of Insurance Issues*, 41(2):193–214.

Beaver, W. H., McNichols, M. F., and Nelson, K. K. (2003). Management of the loss reserve accrual and the distribution of earnings in the property-casualty insurance industry. *Journal of Accounting and Economics*, 35(3):347–376.

Bornhuetter, R. L. and Ferguson, R. E. (1972). The actuary and IBNR. *Proceedings of the casualty actuarial society*, 59(112):15.

Da, Z. and Huang, X. (2020). Harnessing the Wisdom of Crowds. *Management Science*, 66(5):1847–1867.

Forbes, S. W. (1970). Loss Reserving Performance within the Regulatory Framework. *The Journal of Risk and Insurance*, 37(4):527.

Gaver, J. J. and Paterson, J. S. (2004). Do insurers manipulate loss reserves to mask solvency problems? *Journal of Accounting and Economics*, 37(3):393–416.

Goldstein, I. and Yang, L. (2019). Good disclosure, bad disclosure. *Journal of Financial Economics*, 131(1):118–138.

Grace, E. V. (1990). Property-Liability Insurer Reserve Errors: A Theoretical and Empirical Analysis. *The Journal of Risk and Insurance*, 57(1):28.

Grace, M. F. and Leverty, J. T. (2012). Property-Liability Insurer Reserve Error: Motive, Manipulation, or Mistake. *Journal of Risk and Insurance*, 79(2):351–380.

Hsu, W., Huang, Y. R., and Lai, G. (2019). Reserve Management and Audit Committee Characteristics: Evidence from U.S. Property–Liability Insurance Companies. *Journal of Risk and Insurance*, 86(4):1019–1043.

Kazenski, P. M., Feldhaus, W. R., and Schneider, H. C. (1992). Empirical Evidence for Alternative Loss Development Horizons and the Measurement of Reserve Error. *The Journal of Risk and Insurance*, 59(4):668.

Mack, T. (1993). Distribution-free calculation of the standard error of chain ladder reserve estimates. *ASTIN Bulletin*, 23(2):14.

Murphy, K. P. (2007). Conjugate Bayesian analysis of the Gaussian distribution. *def 1.2σ2*, 1(2):29.

Nelson, K. K. (2000). Rate Regulation, Competition, and Loss Reserve Discounting by Property-Casualty Insurers. *The Accounting Review*, 75(1):25.

Petroni, K. R. (1992). Optimistic reporting in the property-casualty insurance industry. *Journal of Accounting and Economics*, 15(4):485–508.

Schmidt, K. and Zocher, M. (2007). The Bornhuetter-Ferguson Principle. *Variance*, 2(1):26.

Smith, B. D. (1980). An Analysis of Auto Liability Loss Reserves and Underwriting Results. *The Journal of Risk and Insurance*, 47(2):305.

Taylor (2019). Loss Reserving Models: Granular and Machine Learning Forms. *Risks*, 7(3):82.

Weiss, M. (1985). A Multivariate Analysis of Loss Reserving Estimates in Property-Liability Insurers. *The Journal of Risk and Insurance*, 52(2):199.