

# Credibility — Counting the Cost

By Joel Atkins

## Abstract

This paper explores examples where common approaches may assign too much credibility to our estimates. It also discusses issues with how we select complements. In some cases, the paper offers better solutions. In other cases, the paper only provides awareness of the issues.

## Introduction

Insurance data is noisy. Most lines have low loss frequencies and highly skewed severities. This is why it makes sense for insured to pool their risks, but they also make it difficult to estimate appropriate rates. To address this, actuaries learn early in their careers to look for a similar cohort to use as a complement and then blend portfolio loss experience with that of the complement. Many helpful books and papers share the derivations behind two common approaches, limited fluctuation credibility and Bühlmann Credibility. Many books and papers explain them well, such as the textbook Basic Ratemaking [Werner] and a paper by Mahler [Mahler].

Credibility formulas discussed in these sources are built on some shaky assumptions. We often overestimate the credibility of our data and use complements that are not as helpful as we'd like. This gives us a false sense of security in our work. Below, we will discuss issues with the credibility standard and then with the complement. While we can't solve these issues, we think it is important to raise the issues and be aware of them. In some cases, we will suggest ways to mitigate them. In other cases, we hope that others will offer suggestions in the future.

## Credibility Standards

There are a number of issues with how we determine the credibility of our data.

### Observations are not independent

Consider this actual auto data from the Minnesota Department of Public Safety [Minnesota DPS]:

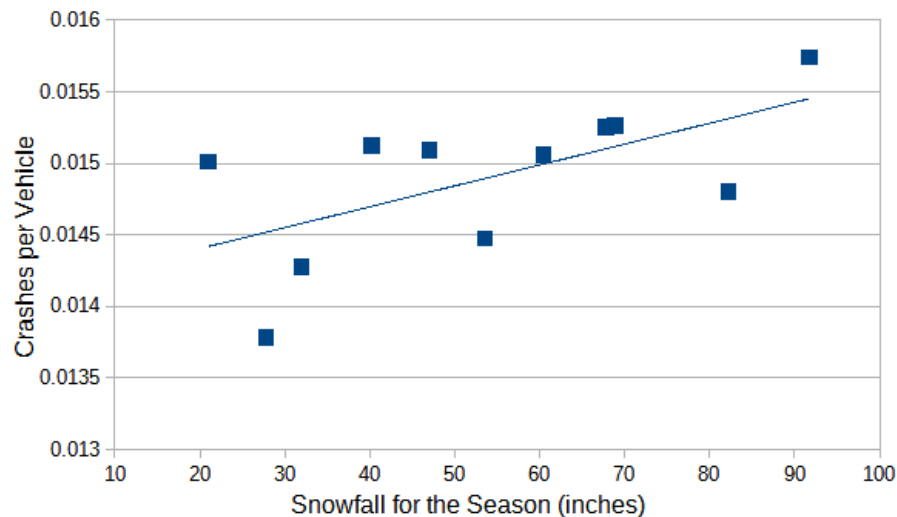
Year	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Registered Vehicles (in millions)	4.87	4.92	4.98	5.02	5.09	5.14	5.24	5.23	5.23	5.35	5.12
Traffic Crashes (in thousands)	73.5	74.1	72.1	69.2	77.7	78.4	74.8	79.1	78.5	79.2	80.6
Traffic Crashes per vehicle	0.0151	0.0151	0.0145	0.0138	0.0153	0.0153	0.0143	0.0151	0.0150	0.0148	0.0157
Change from Prior Year		0.0000	-0.0006	-0.0007	0.0015	0.0000	-0.0010	0.0008	-0.0001	-0.0002	0.0009
% Change from Prior Year		-0.21%	-3.87%	-4.79%	10.74%	-0.08%	-6.41%	5.95%	-0.76%	-1.37%	6.33%

Under any credibility standard, each year would be almost fully credible with 69,200 crashes in the best year. However, we are seeing significant change in the crash frequency. It changes more than 5% in four out of ten years. This is disappointing for all of us who have lots of data.

When we think about Minnesota, one thing that comes to mind is snow. The Twin Cities area had over 55% of Minnesota’s population in both 2010 and 2020 [Census], so we consider their snowfall. (We considered Ramsey, Hennepin, Washington, Anoka, Dakota, Scott, Carver, and Wright Counties to be the Twin City area.) We got the snowfall for the twin cities from the Minnesota Department of Natural Resources [Minnesota DNR].

Year	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Days with more than 1" of snow	14	14	18	7	18	18	10	10	6	21	23
Snowfall for season	47	60.4	53.5	27.8	68.9	67.8	32	40.2	21	82.2	91.7

Making a graph with the above data confirms our prior belief that there is a relationship between snowfall and car accidents.



Given the thin data and lack of an obvious curve, we fit the line shown to this data. Our fitted line is expected crashes per vehicle = 0.0141 + 0.0000146 \* snowfall in inches. The standard deviation for our slope is 0.0000062, so our estimate is 2.34 standard deviations above 0. We get similar, but slightly less significant results using the number of days with more than one inch of snow.

We see in this case that the loss experience of every auto on the road in a given year is correlated with the loss experience of every other auto on the road through the weather. Since our observations are not independent, we will assign too much credibility to our data. One way to mitigate this is to use data from several winters as the data from different winters will be less correlated than data from the same winter. For example, we might use three years of data for April through December and nine years of

data for January through March. We would need to give the January, February, and March data 1/3 the weight of other months since we have three times as much data.

We also lose our assumption of independence when we have multiple observations from the same insured, such as vehicles from a large fleet in commercial auto. Using bootstrap simulations [and sampling insured and not individual automobiles] will give a better sense of how credible our data is.

## Using observed data to determine credibility

When loss experience data has more than the expected number of claims, credibility will be larger than when it has fewer than the expected number of claims. This will cause our estimate to be biased, even when our complement is the true mean of our data.

Consider the simple, illustrative case, in which we have a 0.5 probability of 20 claims and a 0.5 probability of 30 claims. Our complement is 25 claims – the correct expected values. If we define our credibility as  $(\text{number of claims} / 100)^{1/2}$ , we would have the following result:

number of claims	probability	credibility	complement	estimate
20	0.5	0.447	25	22.764
30	0.5	0.548	25	27.739

And the expected value of our estimate would be 25.25. So even though the actual number of claims in both the experience and the complement are unbiased, our estimate is biased because we base our credibility on the actual number of claims. Korn [Korn] suggests a correction, but it requires assumptions about the frequency and severity distributions. While this seems like an extreme example to get a 1% bias, it's not hard if observations are correlated, like the previous case.

## Nothing is fully credible

Any time the environment changes, our past experience cannot be a fully credible estimate of the future. This includes cases such as automobiles becoming safer each year, employers making their workplace safer, and social inflation. Knowing history perfectly will not tell us what to expect next year. Based on this, we may want a credibility measure that asymptotically approaches some number less than one, as data volume increases, rather than assigning “full credibility” to a large enough amount of historical data, as some credibility measures do. (We also want to rely on our colleagues in underwriting and risk control to understand how insureds change over time.)

## Trending old years is imprecise

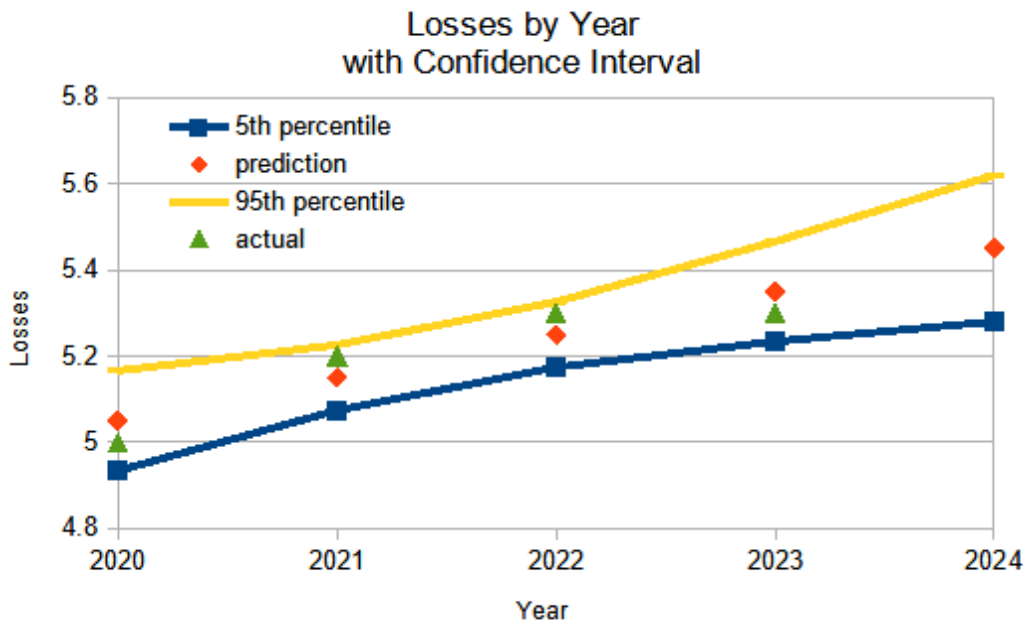
If we think that the older years are different based on some trend, such as inflation or cars becoming safer, we might trend the historical years to some common point in the future. These additional years will give us more data, but trending data is imprecise and understates the noise in our estimate of the trend. Consider the following example. (For illustrative purposes, we'll use a linear trend.)

year	2020	2021	2022	2023
loss	5	5.2	5.3	5.3

This data has a slope of 0.1 per year over the experience period. Trending each observation to 2024 would give us 5.4, 5.5, 5.5, and 5.4.

A naive approach would be to estimate the mean as 5.45. The residuals would be -0.05, 0.05, 0.05, and -0.05. We would then estimate  $\sigma^2$  as  $( (-0.05)^2 + 0.05^2 + 0.05^2 + (-0.05)^2 )^{1/2} / (n - p - 1)$  [this and later estimates based on linear regression use Formula 3.8 on page 47 of Hastie]. Since we applied a linear trend earlier, we will say that p is 1 and our estimate of  $\sigma^2$  is 0.005. We will then say that the variance of our estimate is  $\sigma^2 / n$  or 0.00125. Thus, the standard deviation is 0.035 and our estimate of the average loss is 5.45 +/- 0.035.

This approach ignores the uncertainty in the slope. Considering that uncertainty, our confidence intervals will widen as we move away from the years of our data, and our estimate of the loss in 2024 will be 5.45 +/- 0.087. The graph below shows how the confidence interval widens. The 0.087 standard deviation is derived in the appendix. This is more than twice the standard deviation we found using the naive approach.



There is also the possibility that our trend is picking up a change in mix of business. In light of this, using external benchmarks or industry trends based on bureau data may be better than using our own data.

## Complements may not be appropriate

Often, the cases where our data has the least credibility are also the cases where it's hard to find a good complement.

- In General Liability and Workers Compensation, there are many thin classes which are fairly unique. It's not clear what other classes would make a good complement. In these cases, it may be helpful to use external data, such as injury rates from the Bureau of Labor Statistics.
- If we have a small state, like Rhode Island (very densely populated) or Wyoming (very sparsely populated), it's not clear that countrywide averages are a great complement. We might look for similar states as complements or try to build a countrywide model that incorporates differences in geography, possible through weather, infrastructure, and demographics.
- From 1995-1999, Montana's speed limit was "reasonable and prudent", according to law enforcement interpretation, making it difficult to find a complement for automobile data.

## Shrinkage methods may not be the answer

Shrinkage methods (such as Lasso, Ridge Regression, and Elastic Net) also require us to tune hyper-parameters to determine how much to shrink our parameters. They also make the complement for our parameters 0, which may not make sense if we have some reason to believe that the effect is real, such as seeing it in competitor filings or underwriting intuition.

Other statistical methods have their own issues. Mixed models and hierarchical Bayesian methods give us elegant solutions, but it is not clear how appropriate they are. If we want to think of Wyoming and Rhode Island as samples from some distribution, it's not clear what that distribution is or how we would parameterize it. We have the same question with other applications such as claims models where we have injuries varying from minor lacerations to traumatic brain injury. In these methods, we replace subjective decisions on hyper-parameters with subjective decisions on distributions.

## Conclusion

The use of credibility reduces the variability of our estimate, but we do pay a price. We usually accept some amount of bias and there is rarely a good way to estimate that bias or correct for it. We should be more conscious of how we use credibility. When we assign credibility, we should consider both how stable our data is and how good our complement is. We should also look for external proxies that might be more stable.

We may want to blend multiple estimates which each include some amount of modeling and/or trending. This will help mitigate any changes in our mix of business. We will also need to think about how to weight those estimates. Using cross validation to back test different complements and weights may help.

## Appendix

In the formula we cited in footnote 8,  $\text{Var}(\beta) = (X^T X)^{-1} \sigma^2$ . The variance of our estimate at  $y$  would be  $y^T \text{Var}(\beta) y$  or  $y^T (X^T X)^{-1} \sigma^2 y$ .

In the naive case,  $y = [1]$  and  $X = [1 \ 1 \ 1 \ 1]$  giving us a variance of  $[1] * [4]^{-1} * 0.005 * [1] = 0.00125$  and a standard deviation of 0.035 for our estimate.

When we consider the slope, we have  $y = [1 \ 2024]$  and  $X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2020 & 2021 & 2022 & 2023 \end{bmatrix}$ .

This gives  $X^T X = \begin{bmatrix} 4 & 8086 \\ 8086 & 16345854 \end{bmatrix}$

and

$(X^T X)^{-1} = \begin{bmatrix} 817292.7 & -404.3 \\ -404.3 & 0.2 \end{bmatrix}$

and  $y^T (X^T X)^{-1} \sigma^2 y = 0.0075$ , giving us a standard deviation of 0.087.

## References

Census Bureau <https://www.census.gov/quickfacts/fact/table/scottcountyminnesota,MN/PST045222> and similar links for the other counties.

Hastie, T., Tibshirani, R., and Friedman, J. (2018) *The Elements of Statistical Learning, Second Edition*, Springer.

Korn, U. (2020) An Alternative Approach to Credibility for Large Account Pricing. *Variance* **13**, 13--30, <https://www.casact.org/sites/default/files/2021-07/Alternative-Approach-Credibility-Korn.pdf>

Mahler, H. C. (1990) An Example of Credibility and Shifting Risk Parameters. *Proceedings of the Casualty Actuarial Society* **LXXVII**, 225--282. [https://www.casact.org/sites/default/files/2021-03/8\\_Mahler.pdf](https://www.casact.org/sites/default/files/2021-03/8_Mahler.pdf)

Minnesota Department of Natural Resources

[https://www.dnr.state.mn.us/climate/historical/acid\\_stn\\_data\\_monthly\\_table.html?sid=mspthr&sname=Twin%20Cities%20Area&sdate=1884-07-01&edate=por&element=snow&span=season&counts=cnt\\_ge\\_1](https://www.dnr.state.mn.us/climate/historical/acid_stn_data_monthly_table.html?sid=mspthr&sname=Twin%20Cities%20Area&sdate=1884-07-01&edate=por&element=snow&span=season&counts=cnt_ge_1) and [https://www.dnr.state.mn.us/climate/historical/acid\\_stn\\_data\\_monthly\\_table.html?sid=mspthr&sname=Twin](https://www.dnr.state.mn.us/climate/historical/acid_stn_data_monthly_table.html?sid=mspthr&sname=Twin)

Minnesota Department of Public Safety <https://dps.mn.gov/divisions/ots/reports-statistics/Documents/2018-crash-facts.pdf> and <https://dps.mn.gov/divisions/ots/reports-statistics/Documents/2019-crash-facts.pdf>

Werner, G. and Modlin, C. (2016) *Basic Ratemaking, Fifth Edition*, Casualty Actuarial Society.  
[https://www.casact.org/sites/default/files/2021-03/5\\_Werner\\_Modlin.pdf](https://www.casact.org/sites/default/files/2021-03/5_Werner_Modlin.pdf)