The Credibility of the Overall Rate Indication ---Making the Theory Work

By Joseph Boor, FCAS, PhD, CERA
Actuary, Florida Office of Insurance Regulation
Joe.Boor@floir.com
850-413-5330

CAS Annual Meeting, November 13, 2016

CAS Antitrust Notice

- The Casualty Actuarial Society is committed to adhering strictly to the letter and spirit of the to the letter and spirit of the antitrust laws. Seminars conducted under the auspices of the CAS are designed solely to provide a forum for the expression of various points of view on topics described in the programs or agendas for such meetings.
- Under no circumstances shall CAS seminars be used as a means for competing companies or firms to reach any understanding – expressed or implied – that restricts competition or in any way impairs the ability of members to exercise independent business judgment regarding matters affecting competition.
- It is the responsibility of all seminar participants to be aware of antitrust regulations, to prevent any written or verbal discussions that appear to violate these laws, and to adhere in every respect to the CAS antitrust compliance policy.

Overall Rate Indication

Key Features

- Set of annual loss-based data Points, loss ratios or pure premiums for various accident, etc. years
- No outside reference point like grand mean in class ratemaking.
- Complement of credibility is effectively the trended, maybe also adjusted, present pure premium

Two Views of Credibility

- Limited Fluctuation (Square Root) Credibility
- Best Estimate Credibility

Espoused Pros of Square Root Credibility

- Stable rates
- There's not much difference between the results using square root credibility (with the right full credibility "F") and best estimate credibility.
- Easy to compute

Espoused Pros of Best Estimate Credibility

- Best reflection of costs
  - Max competitiveness directed towards classes with lowest losses vs. competitor prices
- Objective – Credibility "Z" just a function of data, no judgment-based "F"

Phone Poll:

See a lot of filings---what % use limited fluctuation Z, what % used best estimate Z in overall indication?

- 50%/50%
- 75%/25%
- 95+%/less than 5%
- 99+%/ less than 1%

Evaluation of Square Root Credibility

- Square root credibility produces stable rates?
  - Doesn't happen every time, but what if
    - Line of business evaluated once every 5 years? 10 years? 15 years?
    - Volatile trend or disagreement on trend
  - Many years of volatile trend for complement is not stable

Evaluation of Square Root Credibility

- Similar results to best estimate credibility?
  - Mahler paper for helping square root credibility match best estimate was for by class credibility

Evaluation of Square Root Credibility

- Similar results to best estimate credibility?- When there is longer period between rate reviews
  - Square root credibility of new data the same as in annual evaluations-function of $E[counts]$
  - Best estimate case-
    - older data further away from time rate in effect -less predictive
    - Then credibility of trended present rates lower than in annual (Boor 1992)
    - So Z, for new data, is higher for longer period between reviews

Evaluation of Square Root Credibility

- Similar results to best estimate credibility?
  - Did not locate formula for square root credibility to mirror best estimate credibility
  - By it's nature, square root credibility cannot consistently mirror best estimate credibility

Evaluation of Square Root Credibility

- Is Square root credibility easier to implement?
  - Yes
  - What formula to use for best estimate credibility for overall indication?

Evaluation of Square Root Credibility

- Does not always produce stable rates
- Likely does not mirror best estimate credibility
- Easy to implement
  - How to implement best estimate credibility anyway

Phone Poll

What should every actuary know that almost no actuaries actually know?

- The phone number 1-800-FixMySpreadsheet
- How to make a CEO happy-consistently-with the 50% increase to reserves you say must be booked.
- The homogeneous Bühlmann-Straub estimate of loss
- The Gerber-Jones formula

A Credibility Formula for the Overall Rate Indication – The Gerber-Jones Formula

First-show the result

- Special case of formula – GBM w/ constant process error multiplier model of data
  - $Z_i \cong \frac{\delta^2+Z_{i-1}\sigma^2}{\delta^2+Z_{i-1}\sigma^2+\sigma^2}$
  - δ^2 reflects $\%^2$ drift variance of GBM (not underlying linear BM), σ^2 is $\%^2$ process-type variance, Z_{i-1} is last year's credibility
  - Formula accommodates other models-with underlying Markov stochastic process governing true costs over time and independent process-type errors

A Credibility Formula for the Overall Rate Indication – The Gerber-Jones Formula

- Full Disclosure
  - Formula accommodates other models of loss experience – Markov process for drift of underlying costs, independent process-type variances affecting each data point, and similar structures
  - GBM approach used because (in stochastic process land) it is simplest.
    - Also consistent with costs being pushed by a large number of multiplicative factors

A Credibility Formula for the Overall Rate Indication – The Gerber-Jones Formula

- Updating - A bonus you didn't realize was a bonus
  - Rate indication doesn't just use Z_i and the most recent data point, it uses Z_{i-1} times the next previous data point, Z_{i-2} times the pint before that, etc.
  - The optimum rate indication might require making better use of prior data, maybe changing the credibility for the old years
  - Because this is an updating formula, <u>each consecutive update, without changing weights, is the true optimum</u>

The Gerber-Jones Formula – Utility Poll

- Do you think the Gerber-Jones approach holds promise for use in overall rate indications?
  - Yes
  - No

The Gerber-Jones Formula

- There's a **big** problem with Gerber-Jones
  - How do you estimate δ^2 and σ^2?
  - Right now, this could be your excuse for not using Gerber-Jones

The Gerber-Jones Formula

- How do you estimate δ^2 and σ^2?
  - This is the focus of the paper

Estimating δ^2 and σ^2 - Poll

- Which types of methods might be effective?
  - Estimate Z via best fit for historical data
  - Estimate $K=\delta^2/\sigma^2$ which is all you need, as best fit using multiple, related datasets
  - Arithmetic formulas using squared differences between loss data points.
  - Structural analysis of process variance
  - Estimating δ^2 with a larger database.

Estimating Z Via Best Fit for Historical Data

- Find the Z that would have worked best in the past
- Start with, say, 10 years of trended, etc. loss ratio, pure premium, etc. estimates L_i
- Pick provisional Z
- Use $L_1,\ldots, L_5$ with Z to estimate L_6
  - Compute error^2 in estimating the L_6 you know

Estimating Z Via Best Fit for Historical Data Continuation of Process

- Similarly,
  - use $L_2, \ldots, L_6$ with Z to estimate L_7;
  - $L_3, \ldots, L_7$ with Z to estimate L_8;
  - $L_4, \ldots, L_8$ with Z to estimate L_9;
  - $L_5, \ldots, L_9$ with Z to estimate L_{10}.
- Compute error^2 in estimating the L_7, etc. you know

Estimating Z Via Best Fit for Historical Data Continuation of Process

- Sum up the squared errors
- Vary the Z using solution routine – in most spreadsheet software – to find least squared error.
- Result is optimum Z
  - Note : only for steady-state, but steady-state may be good enough

Estimating Z Via Best Fit for Historical Data
A Quibble

- "This method used one year forward estimate, but there is actually a two year gap when I make rates"
- Unlike gaps between reviews, the two or three, etc. year forward indications use the same credibility as the one year.

Sample Calculation of Z from Initial Reported Data and Final Cost of Ten Years of Data

Input/Output for Solution Function			
Value to minimize =	Target=	0.046	
Value to vary to minimize Target is	Z=		0.366

Part 1. Data and Estimation of Older Years

Accident Year	(1) Data Initial Data Values	(2) Data Final Ultimate Value	(3) Z((1-Z)^k) All Estimating Weights	(4) [5 Later (3)] Weights for Estimating 1995	(5) (1)*(4) 1995 Estimate	(6) [4 Later (3)] Weights for Estimating 1996	(7) (1)*(5) 1996 Estimate
1991	1.023	1.070	0.010	0.093	0.095	0.059	0.061
1992	0.991	1.107	0.015	0.147	0.146	0.093	0.092
1993	1.209	1.022	0.024	0.232	0.280	0.147	0.178
1994	0.576	0.923	0.038	0.366	0.211	0.232	0.134
1995	0.886	0.769	0.059		0.000	0.366	0.324
1996	0.858	0.907	0.093		0.000		0.000
1997	0.810	0.880	0.147		0.000		0.000
1998	1.061	0.871	0.232		0.000		0.000
1999	0.891	0.767	0.366		0.000		0.000
2000	0.967	0.826	0.000		0.000		0.000
A. Column Sums				0.838	0.732	0.897	0.788
B. (A./[A. in Prev. col.] Loss Ratio Est.					0.874		0.879
C. (from (1)) Actual Loss Ratio Values					0.769		0.907
D. (B-C.)^2 Squared Error in Estimate					0.011		0.001

Part 2. Estimation of Remaining Years and Total Prediction Error (Target)

Accident Year	(8) [3 Later(3)]*(1) 1997 Estimate	(9) [2 Later (3)]*(1) 1998 Weights	(10) [Next Row(3)]*(1) 1999 Estimate	(11) (3)*(1) 2000 Estimate	
1991	0.038	0.024	0.015	0.010	
1992	0.059	0.037	0.024	0.015	
1993	0.113	0.072	0.045	0.029	
1994	0.085	0.054	0.034	0.022	
1995	0.206	0.130	0.083	0.052	
1996	0.314	0.199	0.126	0.080	
1997	0.000	0.296	0.188	0.119	
1998	0.000	0.000	0.388	0.246	
1999	0.000	0.000	0.000	0.326	
2000	0.000	0.000	0.000	0.000	
A. (as above)	0.814	0.812	0.903	0.899	**Sum of Est. Errors =Target**
B. (as above)	0.871	0.847	0.928	0.914	
C. (as above)	0.880	0.871	0.767	0.826	
D. (as above)	0.000	0.001	0.026	0.008	0.046

Estimating Z Via Best Fit for Historical Data Conditioning

- Method is ill-conditioned when all the prior values are about the same
- Of course, that is when credibility probably doesn't matter

Fitting K and B Across a Large Number of Similar Datasets

- What are K and B?
  - K is similar to "K" in class ratemaking,
    - except instead of process variance over parameter variance have process variance over drift variance ${\sigma^2}/{\delta^2}$

Fitting K and B Across a Large Number of Similar Datasets

- What are K and B?
  - B is symbol used in class ratemaking---"$nK + B$"
  - B is λ^2/δ^2, where λ^2 is loss development uncertainty, independent but of equal size among all points
  - Not in original model, but consistent w/Gerber-Jones

Fitting K and B Across a Large Number of Similar Datasets

- Issues with K and B
  - May increase conditioning problem if all the datasets are of same size. Can only distinquish between K and B via $nK + B$ if the "n" takes different values
    - Consider just using K
  - The λ^2/δ^2 and B also enhance Bühlmann-Straub

Fitting K and B Across a Large Number of Similar Datasets

- Fitting K and B
  - Start with provisional values for K and B
  - Use a largish (say, 12+) group of trended, on level, etc. loss ratios, etc. over time, from a group of states, classes, that would be expected to have about the same process variance (up to exposure differences), drift variance, and development variance - $\sigma^2, \delta^2, \lambda^2$
  - Need exposure/on-level premium too

Fitting K and B Across a Large Number of Similar Datasets

- Given, say 6 years for each "state"
- Use common K, B In formula derived from Gerber-Jones
  - $Z_{i,s} \cong \frac{U_{i,s}+Z_{i-1,s}(K+BU_{i,s})}{U_{i,s}+(1+Z_{i-1,s})(K+BU_{i,s})}$,
  - $Z_{i,s}$ is updating credibility for state 's', year 'i'
  - $U_{i,s}$ is premium/exposure for state 's', year 'i'

Fitting K and B Across a Large Number of Similar Datasets

- Variation from traditional Gerber-Jones
  - $Z_{1,s} \cong \frac{U_{1,s}}{U_{i,s}+K+BU_{1,s}},$
  - No assumption of zero process error at beginning

Fitting K and B Across a Large Number of Similar Datasets

- Use weighted average of ratios
  - $Z_{1,s} \cong \frac{U_{1,s}}{U_{i,s}+K+BU_{1,s}}$,
  - $Z_{i,s}$ generated consecutively: $\boldsymbol{Z_{i,s}} = \frac{[\boldsymbol{U_{i,s}}+\boldsymbol{Z_{i-1,s}}(\boldsymbol{K}+\boldsymbol{BU_{i,s}})]}{[\boldsymbol{U_{i,s}}+(\boldsymbol{1}+\boldsymbol{Z_{i-1,s}})(\boldsymbol{K}+\boldsymbol{BU_{i,s}})]}$
  - $W_{i,s}$ = $Z_{i,s}(1\text{-}Z_{i+1,s})$ … $(1-Z_{n,s})$
  - No assumption of zero process error at beginning means non-Gerber-Jones $Z_{1,s}$- best to match needed $Z_{2,s}$

Fitting K and B Across a Large Number of Similar Datasets

- Fitting step
  - Compute estimate of next loss ratio, pure premium,etc. using weights times historical loss ratios, pure premiums, etc.
  - Compute (estimate-actual)^2 for each s, and sum the results
  - Modify K and B till sum of squats is minimized. May then use them with the same set of s 's.
  - Per earlier $Z_{i,s} \cong \frac{U_{i,s}+Z_{i-1,s}(K+BU_{i,s})}{U_{i,s}+(1+Z_{i-1,s})(K+BU_{i,s})}$

Algebraic Differences of Squares

- Basic concept
  - Difference between first and last point S_n - S_1 is mostly due to drift variance δ^2
  - Difference between adjacent points S_{i+1} - S_i is mostly due to process variance σ^2
  - Use those relationships

Algebraic Differences of Squares

- Formulas
  - Convert to linear Brownian motion by taking logs of S ‘s.
  - Continuing to use S , δ^2, σ^2 notation, although these now refer to values in the linear space – following the paper
  - $E\left[\frac{(n-1)(S_n-S_1)^2-\sum_{i=1}^{n-1}(S_{i+1}-S_i)^2}{(n-1)(n-2)}\right] = \delta^2$
  - $\frac{E\left[\sum_{i=1}^{n-1}(S_{i+1}-S_i)^2 \quad -(S_n-S_1)^2\right]}{2(n-2)} = \sigma^2$

Algebraic Differences of Squares

- Then convert back to exponential version
- Will move freely between geometric and linear Brownian motions in this and remaining sections– should be clear in context, especially per paper

Algebraic Differences of Squares

- Issue
  - Certain patterns may be ill-conditioned
  - Example, medium variations in most of data with big spike at end
    - Is it large drift?
    - Large process variance at the one point?

Estimating σ^2 Structurally

- Recall collective risk equation for process variance
  - $\alpha^2 = \frac{E[\#claims] \times Var[severity] + Var[\#claims] \times E[severity].}{(premium\ or\ exposures)^2}$
  - β^2 = uncorrelated loss development variance
- In the linear version
  - $\sigma^2 = \log\left(\beta^2 + \frac{\alpha^2}{(expected\ loss)^2} + \frac{\alpha^2\beta^2}{(expected\ loss)^2} + 1\right)$
- Formula for linear version of δ^2
  - $\delta^2 \cong \frac{(S_n - S_1)^2 - 2\sigma^2}{n-1}$

Estimating σ^2 Structurally

- Key issue to watch out for
  - $(S_n - S_1)^2$ and σ^2 of about the same size.

Estimating δ^2 From Larger Database

- May be able to locate large database,
  - with minimal process error
  - Very similar character to the business generating the losses
    - Countrywide vs. state? Possibly.

Estimating δ^2 From Larger Database

- Estimate δ^2 (in the linear space) using larger dataset and algebraic formula for δ^2
- Then, in the specific pricing dataset you are working with
  - $\frac{\sum_{i=1}^{n-1}(S_{i+1}-S_i)^2}{2(n-1)} - \frac{\delta^2}{2} \cong \sigma^2$

Overall Concern – Handling Ill-Conditioned Data

- Suggest you use multiple methods and assess strengths and weaknesses of each when selecting
- More like reserving than ratemaking.

Should the Transition to Gerber-Jones be Difficult?- Poll

- Which challenges might it present?
  - Too computationally difficult
  - Estimating key constants
  - Different data used

The Fine Print

- For Gerber-Jones to work, data used in each iteration can't have been used in a previous iteration
  - Making rates every two years using the latest five years of data is not covered by Gerber-Jones
  - Recall that Gerber-Jones is an updating formula
  - Testing shows that if formula has overlapping years between iterations it can't be an updating model.

Resolving The Fine Print

- Main advantage of multiple years in ratemaking – recognizing loss development
  - Suggest correcting last few years in complement of credibility (data receiving complement of credibility) for changes in ultimate loss.
- Getting more credibility – an illusion

Recent Improvements

- Class Ratemaking
  - Credit scoring
  - GLMs
- Cat Models

- What about overall rate level?

The Credibility of the Overall Rate Indication ---Making the Theory Work

???