RPM Workshop 4: Basic Ratemaking

Introduction to Credibility

Ken Doss, FCAS, MAAA
State Farm Insurance

March 2011

Antitrust Notice

- The Casualty Actuarial Society is committed to adhering strictly 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.

General Concept

- Principle 4 of the Statement of Principles Regarding Property and Casualty Ratemaking:
  - A rate cannot be "excessive, inadequate, or unfairly discriminatory"
    - — Excessive: Too high
    - — Inadequate: Too high
    - — Unfairly discriminatory: Allocation of overall rate to individuals is based on cost justification
- At various steps in the ratemaking process, the concept of credibility is introduced (state, class, segment, territory, etc)
- The credibility of data is commonly denoted by the letter "Z"
  - **$0 \le Z \le 1$**

Definitions of Credibility

- Common usage:
  - "Credibility" = the quality of being believed or trusted
  - Implies you are either credible or you are not
- In actuarial science:
  - Credibility is "a measure of the credence that…should be attached to a particular body of experience"

    -- L.H. Longley-Cook
  - Refers to the degree of believability of the data under analysis

    — A relative concept, not an absolute

Why Do We Need Credibility?

- Property / casualty insurance losses are inherently stochastic
  - Losses are fortuitous events

    — Any given insured may or may not have a claim in a given year

    — The size of the claim can vary significantly
- So how much can we believe our data? What other data can be used to aid in calculating the rate for an insured?
- Credibility is a balance of stability and responsiveness

History of Credibility in Ratemaking

- The CAS was founded in 1914, in part to help make rates for a new line of insurance – Workers Compensation – and credibility was born out the problem of how to blend new experience with initial pricing
- Early pioneers:
  - Mowbray (1914) -- how many trials/results need to be observed before I can believe my data?
  - Albert Whitney (1918) -- focus was on combining existing estimates and new data to derive new estimates:

    New Rate = Credibility x Observed Data + (1-Credibility) x Old Rate
  - Perryman (1932) -- how credible is my data if I have less than required for full credibility?

Methods of Incorporating Credibility

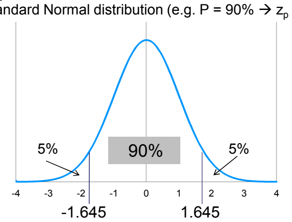
- Limited Fluctuation
  - Limit the effect that random fluctuations in the data can have on an estimate
    - — "Classical credibility"
- Least Squares
  - Make estimation errors as small as possible
    - — Greatest Accuracy
    - — Empirical Bayesian
    - — Bühlmann

Limited Fluctuation Credibility Description

- Goal: Determine how much data one needs before assigning it with full credibility (Z = 1)
  - Standard for full credibility
- Concepts:
  - Full credibility for estimating frequency
  - Full credibility for estimating severity
  - Full credibility for estimating pure premium
  - Amount of partial credibility when data is not fully credible
- Alternatively, the credibility (Z) of an estimate (T) is defined by the probability (P) that it is within a tolerance (k), of the true value

Limited Fluctuation – Meet the Variables

- T: Estimate → the data that we want to test for credibility (e.g. loss ratio)
- Z: Credibility, which is between 0 and 1
- k: Tolerance for error (e.g. the observation is within k = 2.5% of the mean)
- P: Probability that the observation is within k% of the mean. Calculated using the standard Normal distribution (e.g. P = 90% → z_p = 1.645)



Limited Fluctuation Derivation

- New estimate = (Credibility)*(Data) + (1-Credibility)*(Prior Estimate)

$$E2 = Z*T + (1-Z)*E1$$

Add and subtract Z*E[T] → $E2 = Z*T + Z*E[T] - Z*E[T] + (1-Z)*E1$

Regroup → $E2 = (1-Z)*E1 + Z*E(T) + Z*(T-E(T))$

Stability | Truth | Random Error

10

Limited Fluctuation Formula for Z

- Probability that "Random Error" is "small" is P
  - For example, the probability {random error is less than 5%} is 90%

$$P[Z\,(T-E(T)) < kE(T)] = P$$

Isolate T → $P[T < E(T) + kE(T)/Z] = P$

Assuming T is Normally distributed, then…

Introduce mean and std dev. → $E(T) + kE(T)/Z = E(T) + z_p\sqrt{Var(T)}$

$$kE(T)/Z = z_p\sqrt{Var(T)}$$

$$Z = (kE(T)) / (z_p\sqrt{Var(T)})$$

11

Limited Fluctuation Formula for Z – Frequency

- Assuming the insurance frequency process has a Poisson distribution, and ignoring severity:
  - Then E(T) = number of claims (N) and E(T) = Var(T), so:

$Z = (kE(T)) / (z_p\sqrt{Var(T)})$ becomes

$$Z = (kE(T)) / (z_p\sqrt{E(T)})$$

$$Z = (k\sqrt{E(T)}) / (z_p)$$

$$Z = (k\sqrt{N}) / (z_p)$$

Solving for N = Number of claims for full credibility (Z=1)

$$N = (z_p / k)^2$$

12

Limited Fluctuation – Standards for Full Credibility

- Claim counts required for full credibility based on the previous derivation:
  - Remember, $N = (z_p / k)^2$

Number of Claims		k			
P	**z_p**	**2.5%**	**5.0%**	**7.5%**	**10.0%**
90.0%	1.645	4,326	**1,082**	481	291
95.0%	1.960	6,147	1,537	683	584
99.0%	2.576	10,623	2,656	1,180	664
99.99%	3.891	24,219	6,055	2,691	1,514

Limited Fluctuation – Example

- Calculate the expected loss ratio, given that the prior estimated loss ratio is 75%. Assume P=95% and k=10%.
  - Scenario 1:
    Data: Observed loss ratio = 67%, Claim count = 600
    - What is the standard for full credibility?
    - Does this data have full credibility?
    - What is the expected loss ratio?

Limited Fluctuation – Example (continued)

- Calculate the loss ratio, given that the prior estimated loss ratio is 75%. Assume P=95% and k=10%.
  - Scenario 2:
    Data: Observed loss ratio = 67%, Claim count = 400
    - Assuming Z = 0.72, what is the expected loss ratio?

Limited Fluctuation Formula for Z – Pure Premium

- Generalizing to apply to pure premium:
  - T = pure premium = frequency * severity = N * S
  - E(T) = E(N)*E(S) and Var(T) = E(N)*Var(S) + E(S)²*Var(N)

$Z = (kE(T)) / (z_p\sqrt{Var(T)})$

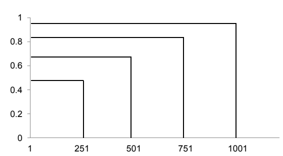
Reduces to, when solving for N = Number of claims for full credibility (Z=1)

$N = (z_p / k)^2 \times (Var(N)/E(N) + Var(S)/E(S)^2)$

Degree of confidence multiplier	Frequency distribution: tends to be close to 1 (equals 1 for Poisson)	Severity distribution: square of coefficient of variation (can be significant)

Limited Fluctuation – Partial Credibility

- Given a full credibility standard based on a number of claims N_F, what is the partial credibility of data based on a number of claims N that is less than N_F?
- Square root rule
  - **$Z = \sqrt{N / N_F}$**
- Calculate Z for N_F = 1,082 and N = 250, 500, 750, and 1,000. What do you notice?
- Exposures vs. Claims



Limited Fluctuation – Increasing Credibility

- Under the square root rule, credibility Z can be increased by
  - Getting more data (increasing N)
  - Accepting a greater margin of error (increasing k)
  - Conceding to smaller P = being less certain (decreasing z_p)

- Based on the formula

$Z = \sqrt{N / N_F}$

$Z = \sqrt{N/(z_p/k)^2}$

$Z = k^*\sqrt{N}/z_p$

Number of Claims	k			
P	**2.5%**	**5.0%**	**7.5%**	**10.0%**
90%	4,326	**1,082**	481	291
95%	6,147	1,537	683	584
99%	10,623	2,656	1,180	664
99.99%	24,219	6,055	2,691	1,514

Limited Fluctuation – Complement of Credibility

- Once the partial credibility Z has been determined, the complement (1-Z) must be applied to something else – the "complement of credibility"

If the data analyzed is…	*A good complement is...*
Pure premium for a class	**Pure premium for all classes**
Loss ratio for an individual risk	**Loss ratio for entire class**
Indicated rate change for a territory	**Indicated rate change for the entire state**
Indicated rate change for entire state	**Trend in loss ratio or the indication for the country**

Limited Fluctuation – Major Strength & Weaknesses

- The strength of limited fluctuation credibility is its simplicity
  - Thus its general acceptance and use
- Establishing a full credibility standard requires subjective selections regarding P and k
- Typical use of the formula based on the Poisson model is inappropriate for most applications
- Partial credibility formula – the square root rule – only holds for a normal approximation of the underlying distribution of the data. Insurance data tends to be skewed.
- Treats credibility as an intrinsic property of the data.

Limited Fluctuation – Example 2

- Calculate the credibility-weighted loss ratio and indicated change, given that the expected loss ratio is 75%. Use the square root rule and when P= 90% and k = 2.5%.

Year	Loss Ratio	Claim Count
2006	67%	530
2007	77%	610
2008	79%	630
2009	77%	620
2010	86%	690

			Credibility	Cred-Wght Loss Ratio	Indicated Rate Chg
'08-'10	81%	1,940			
'06-'10	77%	3,080			

Limited Fluctuation – Example 3

- Given a current territory factor of 1.08, determine the indicated territory factor with 5 years of data. Use the square root rule and the limited fluctuation formula for pure premium. Assume a Poisson frequency distribution and severity coefficient of variation of 1.5.

Year	Territory Exposure	Territory Claim Count	Territory Loss Ratio	Statewide Loss Ratio
2006	3,000	330	125%	78%
2007	3,020	420	153%	83%
2008	3,030	630	269%	85%
2009	3,020	210	122%	79%
2010	3,050	190	108%	72%
'06-'10	15,120	1,780	162%	80%

Limited Fluctuation – Example 3 (continued)

$$N = (z_p / k)^2 * (Var(N)/E(N) + Var(S)/E(S)^2)$$

- Remember, with a Poisson distribution, Var(N) = E(N), so the second term is 1. The third term is the square of the coefficient of variation, which is 1.5^2. Now we just need to select the confidence levels.

- If we want to be within 5% of the true value 90% of the time, the value for $(z_p / k)^2$ is 1,082. Plugging into the formula:

$$N_{claims} = 1{,}082 * (1 + 1.5^2) = 3{,}516.5$$

- Assuming the 5-year statewide frequency is 0.2:

$$N_{exposures} = 3{,}516.5 / 0.2 = 17{,}582.5$$

Limited Fluctuation – Example 3 (continued)

- To show the impact of our selection of an exposure standard instead of a claims standard.

Year	Territory Exposure	Territory Claim Count	Exposure Credibility	Claim Credibility
2006	3,000	330	41.3%	30.6%
2007	3,020	420	41.4%	34.6%
2008	3,030	630	41.5%	42.3%
2009	3,020	210	41.4%	24.4%
2010	3,050	190	41.6%	23.2%
'06-'10	15,120	1,780	92.7%	71.1%

Using a claims standard of 3,517 and an exposure standard of 17,583

Limited Fluctuation – Example 3 (continued)

- Determine what the indicated territorial factor, assuming 15% for fixed expenses.

Year	Territory Loss Ratio	Territory Credibility	Statewide Loss Ratio	Cred Wght Loss Ratio
'06-'10	162%	92.7%	80%	156.0%

156.0% = 92.7% x 162% + 7.3% x 80%

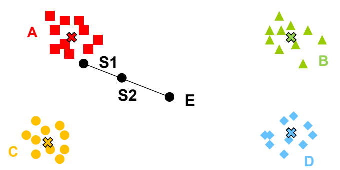
The final indicated territorial factor is (156% / 80%)*0.85 + 0.15 = **1.81**

An alternative approach would be to calculate the indicated factor prior to applying credibility, and then credibility weight the current factor with the indicated factor.

25

Least Squares Credibility Illustration

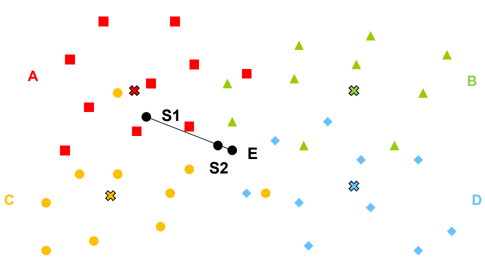
Steve Philbrick's target shooting example...



26

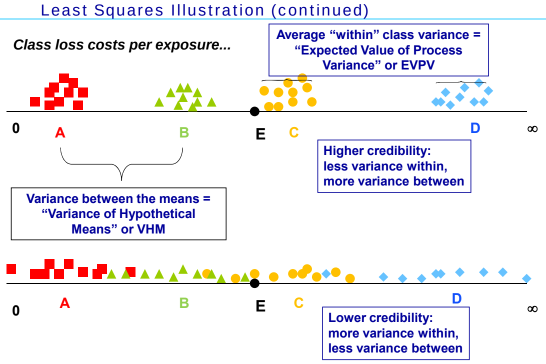
Least Squares Illustration (continued)

Which data exhibits more credibility?



27

Least Squares Illustration (continued)



Least Squares – EVPV and VHM

- Assume we have 3 types of risk: low, medium, and high, which associated probabilities. Calculate the EVPV and VHM.

Risk	P(Claim)	P(Risk)	Variance	Mean^2
Low	20%	60%	0.16	0.04
Medium	30%	25%	0.21	0.09
High	40%	15%	0.24	0.16
Total	25.5%	100%	0.1845	0.0705

- EVPV: For binomial, variance = P(claim) x P(no claim)
  = (20%)(80%)(60%) + (30%)(70%)(25%) + (40%)(60%)(15%)
  = 0.1845
- VHM: Mean^2 – (Mean)^2
  = 0.0705 – $(0.255)^2$
  = 0.0055

Least Squares Derivation

- Similar to our limited fluctuation procedure:

E2 = w * T + (1 – w) * E1, where w = weight

- One method of weighting estimators is to have w be proportional to the reciprocal of the respective variances. So,

$$w = \frac{\frac{1}{(EVPV/n)}}{\frac{1}{(EVPV/n)} + \frac{1}{VHM}} \quad \text{and} \quad 1 - w = \frac{\frac{1}{VHM}}{\frac{1}{(EVPV/n)} + \frac{1}{VHM}}$$

- The denominator chosen to the weights add to 1. Next,

$$w = \frac{n}{(n + EVPV/VHM)} \quad \text{and} \quad 1 - w = 1 - \frac{n}{(n + EVPV/VHM)}$$

Least Squares Derivation (continued)

- Now, to simplify:

  w = n / (n + K)

  Z = n / (n + K), where K = EVPV / VHM

- This results in the minimum of squared errors
- Credibility Z can be increased by:
  - Getting more data (increasing n)
  - Getting less variance within classes (e.g., refining data categories) (decreasing EVPV)
  - Getting more variance between classes (increasing VHM)

31

Least Squares – Example

- Assuming that you have the following book of business, calculate the EVPV, VHM, K, and Z. The prior estimate of the frequency is 0.517. With 4 years of observations and an observed frequency of 0.75, what is the estimated future frequency? Assume the claims are binomially distributed.

Risk	P(Claim)	P(Risk)
Low	40%	65%
Medium	70%	23%
High	80%	12%
Total	51.7%	100%

32

Least Squares – Example (continued)

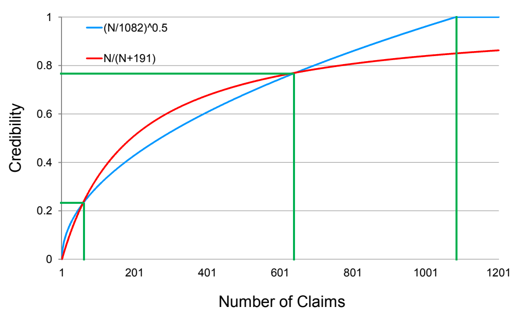
- To determine K, we use K = EVPV/VHM, which is
- Since we're told that we have 4 years of observations, n = 4. Therefore,
- The prior estimate of frequency is the same as the mean calculated before, 0.517, and the observed data results in a frequency of 0.75. This observed data as 31.9% credibility, so…

33

Least Squares – Strengths and Weaknesses

- The least squares credibility result is more intuitively appealing.
  - It is a relative concept
  - It is based on relative variances or volatility of the data
  - There is no such thing as full credibility
- Issues
  - Least squares credibility can be more difficult to apply. Practitioner needs to be able to identify variances.
  - The Credibility Parameter K, is a property of the *entire* set of data. So, for example, if a data set has a small, volatile class and a large, stable class, the credibility parameter of the two classes would be the same.
  - Assumes the complement of credibility is given to the overall mean, which may not be valid in real-world applications.

Comparing Limited Fluctuation and Least Squares



Credibility – Bibliography

- Herzog, Thomas. Introduction to Credibility Theory.
- Longley-Cook, L.H. “An Introduction to Credibility Theory,” PCAS, 1962
- Mayerson, Jones, and Bowers. “On the Credibility of the Pure Premium,” PCAS, LV
- Philbrick, Steve. “An Examination of Credibility Concepts,” PCAS, 1981
- Venter, Gary and Charles Hewitt. “Chapter 7: Credibility,” Foundations of Casualty Actuarial Science.
- Mahler, H.C. and Dean, C.G., “Credibility,” Foundations of Casualty Actuarial Science (Fourth Edition), 2001, Casualty Actuarial Society, Chapter 8.
- Dean, C.G., “Topics in Credibility Theory,” 2004 (SOA Study Note)