RPM Workshop 4: Basic Ratemaking

Introduction to Credibility

Ken Doss, FCAS, MAAA
State Farm Insurance
March 2012

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 intended 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.

What is 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

  - The credibility of data is commonly denoted by the letter Z
    — $0 \leq Z \leq 1$
Why Do We Need Credibility?

- 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 low
    - Unfairly discriminatory: Allocation of overall rate to individuals is based on cost justification

- At various steps in the ratemaking process (state, class, segment, territory, etc), the concept of credibility is introduced to ensure Principle 4 is met

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

- 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 in the rate

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:

  \[
  \text{New Rate} = \text{Credibility} \times \text{Observed Data} + (1-\text{Credibility}) \times \text{Old Rate}
  \]

- Perryman (1932) -- how credible is my data if I have less than required for full credibility?
Methods of Incorporating Credibility

- Limited Fluctuation (Classical credibility)
  - Limit the effect that random fluctuations in the data can have on an estimate
  - Full credibility for frequency, severity, and pure premium
  - Partial credibility

- Least Squares (Greatest Accuracy)
  - Make estimation errors (or squared error) as small as possible
  - Expected value of process variance (EVPV)
  - Variance of hypothetical means (VHM)

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 2.5% of the true value)
- P: Probability that the observation is within k% of the mean. Calculated using the standard Normal distribution (e.g., P = 90% → z = 1.645)
Limited Fluctuation Derivation

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

\[ E_2 = Z^*T + (1-Z)^*E_1 \]

Add and subtract \( Z^*E(T) \)

\[ E_2 = Z^*T + Z^*E(T) - Z^*E(T) + (1-Z)^*E_1 \]

Regroup

\[ E_2 = (1-Z)^*E_1 + Z^*E(T) + Z^*(T-E(T)) \]

Stability, Truth, Random Error

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 \( T \)

Introduce mean and std dev.

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:

\[ Z = \frac{kE(T)}{\sqrt{E(T)}} \]

\[ N = \left( \frac{z_p}{k} \right)^2 \]
Limited Fluctuation – Standards for Full Credibility

- Claim counts required for full credibility based on the previous derivation:
  - Remember, \( N = \left( \frac{z_p}{k} \right)^2 \)

<table>
<thead>
<tr>
<th>Number of Claims</th>
<th>( z_p )</th>
<th>2.5%</th>
<th>5.0%</th>
<th>7.5%</th>
<th>10.0%</th>
</tr>
</thead>
<tbody>
<tr>
<td>90.0%</td>
<td>1.645</td>
<td>4.330</td>
<td>1.082</td>
<td>481</td>
<td>271</td>
</tr>
<tr>
<td>95.0%</td>
<td>1.960</td>
<td>6.147</td>
<td>1.537</td>
<td>683</td>
<td>384</td>
</tr>
<tr>
<td>99.0%</td>
<td>2.576</td>
<td>10.617</td>
<td>2.694</td>
<td>1,180</td>
<td>664</td>
</tr>
<tr>
<td>99.99%</td>
<td>3.891</td>
<td>24.224</td>
<td>6,056</td>
<td>2,692</td>
<td>1,514</td>
</tr>
</tbody>
</table>

Limited Fluctuation – Example 1

- 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 = 400
  - What is the standard for full credibility?
  - Does this data have full credibility?
  - What is the expected loss ratio?

Answer:
- For \( P=95\% \) and \( k=10\% \), the number of claims needed is 384. Since we have 400, the data is considered fully credible.
  
  \[ E_2 = Z \times T + (1-Z) \times E_1 \]
  
  \[ E_2 = 1 \times 67\% + (1 - 1) \times 75\% \]
  
  \[ E_2 = 67\% \]

Limited Fluctuation – Example 1 (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 = 200
  - Assuming \( Z = 0.72 \), what is the expected loss ratio?

Answer:
  
  \[ E_2 = Z \times T + (1-Z) \times E_1 \]
  
  \[ E_2 = 0.72 \times 67\% + (1 - 0.72) \times 75\% \]
  
  \[ E_2 = 69.2\% \]
Limited Fluctuation – Partial Credibility

- Given a full credibility standard based on a number of claims \( N_0 \), what is the partial credibility of data based on a number of claims \( N \) that is less than \( N_0 \)?

- Square root rule

\[ Z = \sqrt{\left( \frac{N}{N_0} \right)} \]

- Calculate credibility (\( Z \)) for \( N_0 = 1,082 \) and \( N = 250, 500, 750, \) and \( 1,000 \).

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

\[ Z = \sqrt{\frac{N}{z_p^2/k}} \]

<table>
<thead>
<tr>
<th>Number of Claims</th>
<th>( k )</th>
<th>2.5%</th>
<th>5.0%</th>
<th>7.5%</th>
<th>10.0%</th>
</tr>
</thead>
<tbody>
<tr>
<td>90%</td>
<td>4,330</td>
<td>3,082</td>
<td>481</td>
<td>271</td>
<td></td>
</tr>
<tr>
<td>95%</td>
<td>6,147</td>
<td>4,537</td>
<td>683</td>
<td>384</td>
<td></td>
</tr>
<tr>
<td>99%</td>
<td>10,817</td>
<td>8,264</td>
<td>1,180</td>
<td>664</td>
<td></td>
</tr>
<tr>
<td>99.99%</td>
<td>24,224</td>
<td>18,456</td>
<td>2,692</td>
<td>1,514</td>
<td></td>
</tr>
</tbody>
</table>

Limited Fluctuation – Example 1 (Revisited)

- 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 = 200
### Limited Fluctuation – Example 2

For the 3 and 5-year periods, calculate the credibility (using the square root rule), credibility-weighted loss ratio and indicated change, given that the expected loss ratio is 75%. Assume \( P = 90\% \) and \( k = 2.5\% \).

<table>
<thead>
<tr>
<th>Year</th>
<th>Loss Ratio</th>
<th>Claim Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td>67%</td>
<td>530</td>
</tr>
<tr>
<td>2008</td>
<td>77%</td>
<td>610</td>
</tr>
<tr>
<td>2009</td>
<td>79%</td>
<td>630</td>
</tr>
<tr>
<td>2010</td>
<td>77%</td>
<td>620</td>
</tr>
<tr>
<td>2011</td>
<td>86%</td>
<td>690</td>
</tr>
</tbody>
</table>

#### Credibility

\[ \sqrt{\frac{1940}{4326}} = 67\% \]

#### Cred-wght Indicated Loss Ratio

\[ 79.0\% = 81\% \times (0.67) + 75\% \times (1 - 0.67) \]

\[ 5.3\% = \frac{79.0\%}{75.0\%} \]

### Limited Fluctuation Formula for Z – Pure Premium

Generalizing to apply to pure premium:

- \( T = \) pure premium = frequency \( \times \) severity = \( N \times S \)
- \( E(T) = E(N) \times E(S) \) and \( Var(T) = E(N) Var(S) + E(S)^2 Var(N) \)

\[ Z = \frac{k E(T)}{\sqrt{Var(T)}} \]

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

\[ N = \left( \frac{z_p}{k} \right)^2 \times \frac{(Var(N)/E(N))}{\frac{Var(S)/E(S)^2}{}} \]

### Degree of confidence multiplier

- \( = 1 \) for Poisson

### Frequency distribution

### Severity distribution

### Coefficient of variation squared

### Limited Fluctuation – Example 3

Given a current territory factor of 1.08, determine the indicated territory factor with 5 years of data. The frequency distribution is Poisson and the severity coefficient of variation of 1.5. Use the square root rule and the limited fluctuation formula for pure premium. Assume that you want to be within 5% of the true value 90% of the time. The statewide frequency is 0.20 and fixed expenses are 15%.

<table>
<thead>
<tr>
<th>Year</th>
<th>Territory Exposure</th>
<th>Territory Claim Count</th>
<th>Territory Loss Ratio</th>
<th>Statewide Loss Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006</td>
<td>3,000</td>
<td>330</td>
<td>125%</td>
<td>78%</td>
</tr>
<tr>
<td>2007</td>
<td>3,020</td>
<td>420</td>
<td>153%</td>
<td>83%</td>
</tr>
<tr>
<td>2008</td>
<td>3,030</td>
<td>630</td>
<td>269%</td>
<td>85%</td>
</tr>
<tr>
<td>2009</td>
<td>3,020</td>
<td>210</td>
<td>122%</td>
<td>79%</td>
</tr>
<tr>
<td>2010</td>
<td>3,050</td>
<td>190</td>
<td>108%</td>
<td>72%</td>
</tr>
<tr>
<td>'06-'10</td>
<td>15,120</td>
<td>1,780</td>
<td>162%</td>
<td>80%</td>
</tr>
</tbody>
</table>
Limited Fluctuation – Example 3 (continued)

If we want to be within 5% of the true value 90% of the time, \( (z_{0.05}/k)^2 \) is 1.082.

Remember, with a Poisson distribution, \( \text{Var}(N) = E(N) \), the second term is 1. The third term is the square of the coefficient of variation, which is 1.52.

\[
N = \frac{(z_{0.05}/k)^2 \times \left( \frac{\text{Var}(N)}{E(N)} + \frac{\text{Var}(S)}{E(S)^2} \right)}{1.082}
\]

Given the 5-year statewide frequency of 0.2:

\[
N_{\text{exposure}} = \frac{3,516.5}{0.2} = 17,582.5
\]

Limited Fluctuation – Example 3 (continued)

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

<table>
<thead>
<tr>
<th>Year</th>
<th>Territory Exposure</th>
<th>Territory Claim Count</th>
<th>Territory Credibility</th>
<th>Exposure Frequency</th>
<th>Credibility Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>’06</td>
<td>3,000</td>
<td>330</td>
<td>41.3%</td>
<td>30.6%</td>
<td>41.3%</td>
</tr>
<tr>
<td>’07</td>
<td>3,020</td>
<td>420</td>
<td>41.4%</td>
<td>34.6%</td>
<td>41.4%</td>
</tr>
<tr>
<td>’08</td>
<td>3,030</td>
<td>630</td>
<td>41.5%</td>
<td>42.3%</td>
<td>41.5%</td>
</tr>
<tr>
<td>’09</td>
<td>3,020</td>
<td>210</td>
<td>41.4%</td>
<td>24.4%</td>
<td>41.4%</td>
</tr>
<tr>
<td>’10</td>
<td>3,050</td>
<td>190</td>
<td>41.6%</td>
<td>23.2%</td>
<td>41.6%</td>
</tr>
<tr>
<td>’06-’10</td>
<td>15,120</td>
<td>1,780</td>
<td>92.7%</td>
<td>71.1%</td>
<td>92.7%</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>Year</th>
<th>Territory Loss Ratio</th>
<th>Territory Credibility</th>
<th>Territory Statewide Loss Ratio</th>
<th>Cred Wgt Statewide Loss Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>’06-’10</td>
<td>162%</td>
<td>92.7%</td>
<td>80%</td>
<td>156.0%</td>
</tr>
</tbody>
</table>

The final indicated territorial factor is \( (156.0% / 80% / 0.85 + 0.15 = 1.81\)

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.
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”

<table>
<thead>
<tr>
<th>If the data analyzed is...</th>
<th>A good complement is...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pure premium for a class</td>
<td>Pure premium for all classes</td>
</tr>
<tr>
<td>Loss ratio for an individual risk</td>
<td>Loss ratio for entire class</td>
</tr>
<tr>
<td>Indicated rate change for a territory</td>
<td>Indicated rate change for the entire state</td>
</tr>
<tr>
<td>Indicated rate change for entire state</td>
<td>Trend in loss ratio or the indication for the country</td>
</tr>
</tbody>
</table>

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.

Least Squares Credibility Illustration

Philbrick’s target shooting example – Round 1
Philbrick's target shooting example – Round 2

Least Squares Credibility Illustration (continued)

What round exhibits more credibility?

Variance between the means = “Variance of Hypothetical Means” or VHM

Higher credibility: less variance within, more variance between

Lower credibility: more variance within, less variance between

Least Squares Illustration (continued)

Variance between the means = “Variance of Hypothetical Means” or VHM

Average within class variance = Expected Value of Process Variance (EVPV)

Higher credibility: less variance within, more variance between

Lower credibility: more variance within, less variance between

Least Squares – EVPV and VHM

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

<table>
<thead>
<tr>
<th>Risk</th>
<th>P(Claim)</th>
<th>P(Risk)</th>
<th>Variance</th>
<th>Mean²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>20%</td>
<td>60%</td>
<td>0.18</td>
<td>0.04</td>
</tr>
<tr>
<td>Medium</td>
<td>30%</td>
<td>25%</td>
<td>0.21</td>
<td>0.09</td>
</tr>
<tr>
<td>High</td>
<td>40%</td>
<td>15%</td>
<td>0.24</td>
<td>0.16</td>
</tr>
<tr>
<td>Total</td>
<td>25.5%</td>
<td>100%</td>
<td>0.1845</td>
<td>0.0705</td>
</tr>
</tbody>
</table>

EVPV: For binomial, variance = P(claim) x P(no claim) 
= (20%)(80%)(60%) + (30%)(70%)(25%) + (40%)(60%)(15%) 
= 0.1845

VHM: Mean² – (Mean)² 
= 0.0705 – (0.255)² 
= 0.0055
**Least Squares Derivation**

- Similar to our limited fluctuation procedure:
  
  \[ E_2 = w \cdot T + (1 - w) \cdot E_1 \]

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

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

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

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

---

**Least Squares Derivation (continued)**

- Now, to simplify:

  \[ w = \frac{n}{(n + K)} \]

  \[ Z = \frac{n}{(n + K)}, \text{ where } K = \frac{EVPV}{VHM} \]

- 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 \( VHM \))

---

**Least Squares – Example**

- Assuming that you have the following book of business, calculate the \( EVPV, \ VHM, \) 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.

<table>
<thead>
<tr>
<th>Risk</th>
<th>P(Claim)</th>
<th>P(Risk)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>40%</td>
<td>20%</td>
</tr>
<tr>
<td>Medium</td>
<td>70%</td>
<td>23%</td>
</tr>
<tr>
<td>High</td>
<td>80%</td>
<td>12%</td>
</tr>
<tr>
<td>Total</td>
<td>51.7%</td>
<td>100%</td>
</tr>
</tbody>
</table>
Least Squares – Example (continued)

- To determine K, we use \( K = \frac{\text{EVPV}}{\text{VHM}} \), which is:
  \[ K = \frac{0.2235}{0.0262} = 8.53 \]

- Since we’re told that we have 4 years of observations, \( n = 4 \). Therefore,
  \[ Z = \frac{n}{n + K} = \frac{4}{4 + 8.53} = 0.319. \]

- 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…

\[ E_2 = Z \times T + (1 - Z) \times E_1 \]

\[ \text{31.9\%} \times 0.75 + 68.1\% \times 0.517 = 0.5913 \]

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 square 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

<table>
<thead>
<tr>
<th>Number of Claims</th>
<th>Credibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.1</td>
</tr>
<tr>
<td>150</td>
<td>0.2</td>
</tr>
<tr>
<td>200</td>
<td>0.3</td>
</tr>
<tr>
<td>250</td>
<td>0.4</td>
</tr>
<tr>
<td>300</td>
<td>0.5</td>
</tr>
<tr>
<td>350</td>
<td>0.6</td>
</tr>
<tr>
<td>400</td>
<td>0.7</td>
</tr>
<tr>
<td>450</td>
<td>0.8</td>
</tr>
<tr>
<td>500</td>
<td>0.9</td>
</tr>
</tbody>
</table>
### Credibility – Bibliography

- Herzog, Thomas. *Introduction to Credibility Theory.*
- Mayerson, Jones, and Bowers. “On the Credibility of the Pure Premium,” PCAS, LV
- Dean, C.G., “Topics in Credibility Theory,” 2004 (SOA Study Note)