Credibility Methods for Individual Life Insurance

Maxwell Gong*

AQR Capital Management, Greenwich, CT 06830 maxwell.gong@aqr.com

Zhuangdi Li*

Deloitte, San Francisco, CA 94105, zhuali@deloitte.com

Maria Milazzo, FSA

Unum, Chattanooga, TN 37402, mmilazzo@unum.com

Kristen Moore, PhD, ASA

University of Michigan, Ann Arbor, MI 48109-1043, ksmoore@umich.edu

Matthew Provencher, FSA

Mutual of Omaha Insurance Co, Omaha, NE 68175-1004 Matt.Provencher@mutualofomaha.com

Credibility theory is used widely in group health and casualty insurance; however, it is generally not used in individual life and annuity business. With the introduction of Principle Based Reserving (PBR), which relies more heavily on company-specific experience, credibility theory is becoming increasingly important for life actuaries. In this paper, we review the two most commonly used credibility methods: Limited Fluctuation and Greatest Accuracy (Bühlmann) credibility. Following the approach of Klugman et al. (2009), we apply the Limited Fluctuation Method to M Financial Group's experience data and describe some general, qualitative observations. In addition, we use simulation to generate a universe of data and compute Limited Fluctuation and Greatest Accuracy credibility factors for actual-to-expected (A/E) mortality ratios following the approach of Klugman et al. (2009). We also compare the two credibility factors to an intuitive benchmark credibility measure. We see that for our simulated data set, as well as for the real data in Klugman et al. (2009), the Limited Fluctuation factors are significantly lower than the Greatest Accuracy factors, particularly for low numbers of claims. Thus, the Limited Fluctuation method may understate the credibility for companies with favorable mortality experience.

---

1. Introduction

1.1. Background

Insurance is priced based on assumptions regarding the insured population. For example, in life and health insurance, actuaries use assumptions about a group's mortality or morbidity, respectively. In auto insurance, actuaries make assumptions about a group of drivers' propensity toward accidents, damage, theft, etc.

The credibility ratemaking problem is the following: suppose that an individual risk has better (or worse) experience than the other members of the risk class. Note that the individual risk might actually be a group; for example, a group of auto insurance policyholders or an employer with group health coverage for its employees.

To what extent is the experience credible? How much of the experience difference can be attributed to random variation and how much is due to the fact that the individual is actually a better or worse risk than the rest of the population? To what extent should that experience be used in setting future premiums?

We can formulate the problem as follows.

- Denote losses by X_j and assume that we have observed independent losses $\mathbf{X} = (X_1, X_2, \ldots, X_n)$. Note that X_j might be the annual loss amount from policyholder j, or the loss in the jth period, depending on the context.

- Let $\xi = E[X_j]$ and $\sigma^2 = \mathrm{Var}(X_j)$.

- Let $S = \sum_{j=1}^{n} X_j$ and let $\overline{X} = \frac{S}{n}$ be the sample mean.
  — Observe that $E[S] = n\xi, E[\overline{X}] = \xi, \mathrm{Var}(S) = n\sigma^2$, and $\mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n}$.

- Let M be some other estimate of the mean for this group. M might be based on industry data or large experience studies on groups similar to the risk class in question.

Credibility theory provides actuaries with a method for combining $\overline{X}$ and M for pricing. The resulting credibility estimate is

$$P = Z\overline{X} + (1 - Z)M; \tag{1}$$

Z is called the credibility factor.

1.2. Significance

Credibility theory is important for actuaries as it provides a means for using company- or group-specific experience in pricing and risk assessment. While credibility theory is used widely in health and casualty insurance, it is generally not used in life and annuity business. The 2008 Practice Note of the American Academy of Actuaries' (AAA's) Life Valuation Subcommittee observes, "For some time, credibility theory has been applied within the property and casualty industry in order to

solve business problems. This has not been the case within the life, annuity and health industries. Therefore, examples of the use of credibility theory and related practices are somewhat difficult to find and somewhat simplistic in their content," AAA (2008). Similarly, the 2009 report Klugman et al. (2009) notes, "The major conclusion from this survey of 190 US insurers is that credibility theory is not widely adopted among surveyed actuaries at United States life and annuity carriers to date in managing mortality-, lapse- and expense- related risks."

Actuarial Standard of Practice 25 (ASOP 25) recommends that credibility theory be used and provides guidance on credibility procedures for health, casualty, and other coverages. In 2013, the Actuarial Standards Board revised ASOP 25 to include the individual life practice area; thus, it will be important for life actuaries to start to use credibility methodology, ASB (2013).

Moreover, credibility theory is increasingly important for life actuaries as the Standard Valuation Law (SVL) is changing to require that Principle Based Reserving (PBR) be used in conjunction with the traditional, formulaic approaches prescribed by state insurance regulations. PBR relies more heavily on company-specific experience; thus, it will be important for actuaries to have sound credibility methodology, NAIC (2013). There is a proposed ASOP for PBR that places significant emphasis on credibility procedures, ASB (2014).

1.3. Overview of Paper

Our paper is structured as follows. In Section 2, we provide a brief overview of the two most common credibility methods: Limited Fluctuation (LF) and Greatest Accuracy (GA) or Bühlmann Credibility. We will see that the LF method is easy to apply, but has several significant shortcomings. On the other hand, GA has a stronger mathematical foundation, but it generally can not be applied in practice because of data constraints. In Section 3, we summarize some of the results of Klugman et al. (2009), in which the authors illustrate an application of both the LF and GA methods to mortality experience from the Society of Actuaries (SOA) 2004-2005 Experience Study. In Section 4, we apply the LF method to M Financial's experience data and share some qualitative observations about our results. In Section 5, we use simulation to generate a "universe" of data. We apply the LF and GA credibility methods to the data and compare the results to an intuitive (though not mathematically grounded) benchmark "credibility factor." Based on the results of the qualitative comparison of the methods in Section 5, we document our conclusions and recommendations in Section 6.

2. Brief Overview of Credibility Methods

The two most common methods for computing the credibility factor Z are Limited Fluctuation (LF) credibility and Greatest Accuracy (GA) or Bühlmann credibility.

2.1. Limited Fluctuation Credibility

2.1.1. Full Credibility
The Limited Fluctuation method is most commonly used in practice. To apply LF, one computes the minimum sample size so that $\overline{X}$ will be within distance r of the true mean ξ with probability p; in other words, we seek n such that

$$\text{pr}\left(-r \leq \frac{\overline{X}-\xi}{\xi} \leq r\right) \geq p. \tag{2}$$

Here, the sample size can be expressed in terms of number of claims, number of exposures (for example, person-years), or aggregate claims. If the sample size meets or exceeds the minimum, then full credibility ($Z = 1$) can be assigned to the experience data. Otherwise, partial credibility may be assigned based on the ratio of the actual sample size to the size required for full credibility.

Observe that (2) yields the following equivalent conditions:

$$\text{pr}(-r\xi \leq \overline{X} - \xi \leq r\xi) \geq p \Longleftrightarrow$$
$$\text{pr}\left(\frac{-r\xi}{\sigma/\sqrt{n}} \leq \frac{\overline{X}-\xi}{\sigma/\sqrt{n}} \leq \frac{r\xi}{\sigma/\sqrt{n}}\right) \geq p.$$

By the Central Limit Theorem, we conclude that the random variable $\frac{\overline{X}-\xi}{\sigma/\sqrt{n}}$ has standard normal distribution. Denote this random variable by Y. Then we seek n so that

$$\text{pr}\left(\frac{-r\xi}{\sigma/\sqrt{n}} \leq Y \leq \frac{r\xi}{\sigma/\sqrt{n}}\right) \geq p.$$

This condition holds if and only if

$$2\Phi\left(\frac{r\xi}{\sigma/\sqrt{n}}\right) - 1 \geq p,$$

where $\Phi(u)$ is the standard normal cumulative distribution function (CDF) evaluated at u.

Finally, we see that condition (2) holds if and only if

$$\frac{r\xi}{\sigma/\sqrt{n}} \geq y_p,$$

or, equivalently,

$$n \geq \left(\frac{y_p}{r}\right)^2 \left(\frac{\sigma}{\xi}\right)^2 = n_f, \tag{3}$$

where y_p is the $\frac{1+p}{2}$-percentile of the standard normal distribution. We denote this critical value of n for which full credibility is awarded by n_f.

From (3), we have the following:

$$Z = 1 \Longleftrightarrow n \geq n_f \Longleftrightarrow \text{Var}(\overline{X}) = \frac{\sigma^2}{n} \leq \xi^2 \left(\frac{r}{y_p}\right)^2. \tag{4}$$

This condition has intuitive appeal; full credibility is awarded if the observations are not too variable.

For example, suppose $r = 0.05$ and $p = 0.90$. Then we seek n so that the probability is at least 90% that the relative error in $\overline{X}$ is smaller than 5%. We have that

- $\frac{1+p}{2} = 0.95$,
- $y_p = 1.645$,
- $\left(\frac{y_p}{r}\right)^2 \approx 1082$,

and for full credibility, we require that

$$n \geq 1082 \left(\frac{\sigma}{\xi}\right)^2.$$

Similarly, if we choose $r = 0.03$ and $p = 0.90$, $\left(\frac{y_p}{r}\right)^2 \approx 3007$, and the standard for full credibility is

$$n \geq 3007 \left(\frac{\sigma}{\xi}\right)^2.$$

Remark: We computed the standard for full credibility to control the deviation of $\overline{X}$ from its mean ξ. Alternatively, we could compute n to control the error in S relative to its mean $n\xi$. Following the derivation above, the same criterion as in (3) results. This is not surprising as S is a scalar multiple of $\overline{X}$.

2.1.2. Application to Life Insurance Suppose now that the X_j are Bernoulli random variables that assume the values 1 or 0 with probabilities q and $1 - q$, respectively. In the life insurance context, the random variable S counts the number of deaths. $S = \sum_{j=1}^{n} X_j$ has binomial distribution, and under appropriate conditions, we can approximate the distribution of S with the Poisson distribution with mean and variance $\lambda = nq$. Note that λ is the expected number of deaths in a group of n lives. Applying (3) with $\lambda = n\xi = nq$, the standard for full credibility is

$$\lambda = nq = \text{expected number of claims} \geq \left(\frac{y_p}{r}\right)^2. \tag{5}$$

In practice, we replace the expected number of claims by the observed number of claims, applying full credibility if at least 1,082 (or 3,007), for example, are observed.

Remark: In some derivations, λ is used to represent the expected number of claims per policy; thus, the credibility standard is written as $n\lambda \geq \left(\frac{y_p}{r}\right)^2$; see Klugman et al. (2012), for example. In our derivation above, we used λ to represent the expected number of claims for the group of n lives; thus, the standard is expressed as in (5).

2.1.3. Partial Credibility Suppose now that full credibility is not justified; i.e., that $n < n_f$. What value of $Z < 1$ should we assign in computing the credibility estimate (1)? In Klugman et al. (2012), the authors note, "A variety of arguments have been used for developing the value of Z, many of which lead to the same answer. All of them are flawed in one way or another." They present the following derivation. We choose the value of Z in order to control the variance of the credibility estimate P in (1). Observe first that

$$\text{Var}(P) = Z^2 \text{Var}(\overline{X}) = Z^2 \frac{\sigma^2}{n}.$$

From (4), we see that when $Z < 1$, we can not ensure that $\text{Var}(\overline{X})$ is small. Thus, we choose the value of $Z < 1$ so that $\text{Var}(P)$ is fixed at its upper bound when $Z = 1$; in other words, we choose Z so that

$$\text{Var}(P) = Z^2 \text{Var}(\overline{X}) = Z^2 \frac{\sigma^2}{n} = \xi^2 \left(\frac{r}{y_p} \right)^2;$$

thus, we set

$$Z = \sqrt{n} \left(\frac{\xi}{\sigma} \right) \left(\frac{r}{y_p} \right) = \sqrt{\frac{n}{n_f}}. \tag{6}$$

Remarks

1. If full credibility is not justified, i.e., if $n < n_f$, the partial credibility factor Z is the square root of the ratio of the number of observations n to the number of observations n_f required for full credibility.

2. Observe that as σ increases, Z decreases; thus, lower credibility is awarded when the observations are more variable. Again, this is consistent with our intuition.

3. In (6), the term

$$\sqrt{n} \left(\frac{\xi}{\sigma} \right) = \frac{\xi}{(\sigma/\sqrt{n})} = \frac{E[\overline{X}]}{\sqrt{\text{Var}(\overline{X})}}$$

is the mean of the estimator divided by the standard deviation of the estimator $\overline{X}$ of the unknown mean ξ.

4. We can write the formula (6) succinctly to include both the full and partial credibility cases by writing

$$
\begin{aligned}
Z &= \min \left\{ 1, \sqrt{n} \left(\frac{\xi}{\sigma} \right) \left(\frac{r}{y_p} \right) \right\} = \min \left\{ 1, \frac{E[\text{estimator}]}{\sqrt{\text{Var(estimator)}}} \left(\frac{r}{y_p} \right) \right\} \\
&= \min \left\{ 1, \sqrt{\frac{n}{n_f}} \right\}
\end{aligned}
\tag{7}
$$

2.1.4. Strengths and Weaknesses of the Limited Fluctuation Approach The LF method is simple to apply and, unlike GA Credibility, it relies only on company-specific data; thus, LF is used widely in practice. However, it has numerous shortcomings. For example:

- There is no justification for choosing an estimate of the form (1).

- There is no guarantee of the reliability of the estimate M, and the method does not account for the relative soundness of M versus $\overline{X}$.

- The choices of p and r are completely arbitrary. Note that as $r \to 0$ or $p \to 1$, $n_f \to \infty$; thus, given any credibility standard n_f, one can select a value of r and p to justify it!

2.2. Greatest Accuracy (Bühlmann) Credibility

Another common approach is Greatest Accuracy (GA) or Bühlmann credibility. In this approach, we assume that the risk level of the members of the risk class are described by a parameter Θ, which

varies by policyholder. Note that we assume here that the risk class has been determined by the usual underwriting process; i.e., that all of the "standard" underwriting criteria (e.g., smoker status, health history, driving record, etc.) have already been considered. Thus, Θ represents the residual risk heterogeneity within the risk class. We assume that Θ and its distribution are unobservable. An obvious choice for the premium would be $\mu_{n+1}(\theta) := E[X_{n+1}|\Theta = \theta]$.

Suppose we restrict ourselves to estimators that are linear combinations of the past observations; i.e., estimators of the form

$$\alpha_0 + \sum_{j=1}^{n} \alpha_j X_j.$$

Define the expected value of the hypothetical means μ by

$$\mu = E[E[X_j|\theta]].$$

One can show that, under certain conditions, the credibility premium $P = Z\overline{X} + (1-Z)\mu$ minimizes the squared error loss

$$Q = E\{[\mu_{n+1}(\Theta) - \alpha_0 - \sum_{j=1}^{n} \alpha_j X_j]^2\},$$

where the credibility factor is

$$Z = \frac{n}{n+k}.$$

Here n is the sample size and

$$k = \frac{\nu}{a} = \frac{E[\mathrm{Var}(X_j|\theta)]}{\mathrm{Var}(E[X_j|\theta])} = \frac{\text{expected process variance}}{\text{variance of the hypothetical means}}.$$

It turns out that P is also the best linear approximation to the Bayesian premium $E[X_{n+1}|\mathbf{X}]$.

We observe that under the GA method, the following intuitive results hold:

- $Z \to 1$ as $n \to \infty$.

- For more homogeneous risk classes (i.e., those whose value of a is small relative to ν), Z will be closer to 0. In other words, the value of μ is a more valuable predictor for a more homogenous population. However, for a more heterogeneous group (i.e., those whose vale of a is large relative to ν), Z will be closer to 1. This result is appealing. If risk classes are very similar to each other (a is small relative to ν), the population mean μ should be weighted more heavily. If the risk classes are very different from each other (a is large relative to ν), the experience data should get more weight.

2.2.1. Strengths and Weaknesses of the Greatest Accuracy Method The GA method has a more sound mathematical foundation in the sense that we do not arbitrarily choose an estimator of the form (1). Rather, we choose the best estimator (in the sense of mean-squared error) among the class of estimators that are linear in the past observations. Moreover, unlike LF, GA credibility takes into account how distinctive a group or risk class is from the rest of the population. However, in practice, companies do not have access to the data required to compute the expected process variance a or the variance in the hypothetical means ν; these quantities rely on (proprietary) data from other companies. As a result, GA credibility is rarely used in practice.

2.3. Other Credibility Methods

The credibility ratemaking problem can be framed in the context of Generalized Linear Mixed Models (GLMMs). We refer the reader to Frees et al. (1999) and Nelder and Verrall (1997) for more information. In fact, the GA method presented in Section 2.2 is a special case of the GLMM; see Frees et al. (1999) and Klinker (2011) for details. Expressing credibility models in the framework of GLMMs is advantageous as they allow for more generality and flexibility. Moreover, one can use standard statistical software packages for data analysis. However, for our purposes, and for our simulated data set, the additional generality of GLMMs was not required. Other methods include Mixed Effects Models, Hierarchical Models, and Evolutionary Models; we refer the reader to Buhlmann and Gisler (2005), Dannenbburg et al. (1996), and Goovaerts and Hoogstad (1987).

3. Previous Literature: Application of Credibility to Company Mortality Experience Data

In Klugman et al. (2009), the authors apply both the LF and GA methods to determine credibility factors for companies' actual-to-expected (A/E) mortality ratio in terms of claim counts and amounts paid. Expected mortality is based on the 2001 VBT and actual mortality is from 10 companies that participated in the Society of Actuaries (SOA) 2004-2005 Experience Study. The authors develop the formulae for the A/E ratios and the credibility factors and include an Excel spreadsheet for concreteness.

We apply the methods and formulae of Klugman et al. (2009) in the work that follows in Sections 4 and 5. For completeness and readability, we briefly summarize the notation and formulae.

3.1. Notation

Assume that there are n lives.
- f_i is the fraction of the year for which the ith life was observed.
- $d_i = 1$ if life i died during the year; otherwise, $d_i = 0$.

- q_i is the observed mortality rate.

- q_i^s is the standard table mortality rate.

- We assume that $q_i = m_c q_i^s$; i.e., that the actual mortality is a constant multiple of the table.

- $A_c = \sum_{i=1}^{n} d_i =$ the actual number of deaths.

- $E_c = \sum_{i=1}^{n} f_i q_i{}^s =$ the expected number of deaths.

- $\hat{m}_c = \frac{A_c}{E_c} =$ the estimated actual-to-expected (A/E) mortality ratio based on claim *counts*.

Observe that A_c and E_c give the actual and expected *number* of deaths. We define similar quantities for the actual and expected *dollar amounts paid* as follows. Let b_i be the benefit amount for policy i. Then we define

- $A_d = \sum_{i=1}^{n} b_i d_i =$ the actual amount paid.

- $E_d = \sum_{i=1}^{n} b_i f_i q_i{}^s =$ the expected amount paid.

- $\hat{m}_d = \frac{A_d}{E_d} =$ the estimated actual-to-expected (A/E) mortality ratio based on claim *amounts*.

3.2. Limited Fluctuation Formulae

In order to compute the credibility factor, we need the mean and variance of the estimators $\hat{m}_c$ and $\hat{m}_d$. We present the results for $\hat{m}_c$; the results for $\hat{m}_d$ are similar. One can show that

$$E[\hat{m}_c] = \frac{\sum_{i=1}^{n} E[d_i]}{E_c} = \frac{\sum_{i=1}^{n} f_i q_i}{E_c} = \frac{\sum_{i=1}^{n} f_i m_c q_i^s}{E_c} = m_c$$

and

$$\mathrm{Var}(\hat{m}_c) = \frac{\sum_{i=1}^{n} f_i q_i (1 - f_i q_i)}{E_c^2} = \frac{\sum_{i=1}^{n} f_i m_c q_i^s (1 - f_i m_c q_i^s)}{E_c^2}. \tag{8}$$

If q_i^s is sufficiently small, we can assume that $(1 - f_i m_c q_i^s)$ is approximately 1, and the expression above for the variance simplifies:

$$\mathrm{Var}(\hat{m}_c) \approx \frac{\sum_{i=1}^{n} f_i m_c q_i^s}{E_c^2} = \frac{m_c}{E_c}. \tag{9}$$

Now, combining the expressions for the mean and variance of the estimator $\hat{m}_c$ with the expression for the credibility factor given in (7), we have that

$$Z = \min \left\{ 1, \left(\frac{r}{y_p} \right) \left(\frac{\sum_{i=1}^{n} f_i m_c q_i^s}{\sqrt{\sum_{i=1}^{n} f_i m_c q_i^s (1 - f_i m_c q_i^s)}} \right) \right\}.$$

If we use the approximation for the variance given in (9), we have

$$Z = \min \left\{ 1, \left(\frac{r}{y_p} \right) \sqrt{\sum_{i=1}^{n} f_i m_c q_i^s} \right\}.$$

Finally, replacing the unknown quantity m_c with its estimate from the observed data $\hat{m}_c = \frac{A_c}{E_c}$, the expressions simplify as

$$
\begin{aligned}
Z &= \min\left\{1, \left(\frac{r}{y_p}\right)\left(\frac{\sum_{i=1}^{n} f_i q_i^s}{\sqrt{\sum_{i=1}^{n} f_i q_i^s (1 - f_i \hat{m}_c q_i^s)}}\right)\sqrt{\frac{A_c}{E_c}}\right\} \\
&\approx \min\left\{1, \left(\frac{r}{y_p}\right)\sqrt{A_c}\right\}.
\end{aligned}
\tag{10}
$$

Observe that the final expression in (10) is equivalent to the expression in (6). Recall that if $r = 0.05$ and $p = 0.90$, the approximate expression in (10) becomes

$$
Z = \min\left\{1, \sqrt{\frac{A_c}{1,082}}\right\}.
$$

Similarly, if $r = 0.03$ and $p = 0.90$, the approximate expression in (10) becomes

$$
Z = \min\left\{1, \sqrt{\frac{A_c}{3,007}}\right\}.
$$

We remark that these parameters and the resulting requirement of 3,007 claims for full credibility are prescribed by the Canadian Committee on Life Insurance Financial Reporting, Canadian Institute of Actuaries (2002).

3.3. Greatest Accuracy Formulae

We define the notation as in Section 3.1. Following Klugman et al. (2009), we suppress the subscripts c and d for count and dollar, respectively. We add the subscript h to emphasize that we are computing the credibility factor for company h, $h = 1, \ldots, r$. Thus, for example, f_{hi} the fraction of the year for which the ith life from company h was observed. We denote the number of lives in company h by n_h. We present the formulas for the credibility factors based on dollar amounts and remark that one can compute the credibility factors based on claim counts by setting the benefit amounts b_{hi} equal to 1.

In Klugman et al. (2009), the authors show that

$$
Z = \frac{E_h}{E_h + \frac{\mu}{\sigma^2 E_h} B_h - \frac{\mu^2 + \sigma^2}{\sigma^2 E_h} C_h}
$$

minimizes the mean squared error

$$
E[(m_h - (Z\hat{m}_h + (1 - Z)\mu))^2],
$$

where

$$
B_h = \sum_{i=1}^{n_h} b_{hi}^2 f_{hi} q_{hi}^s \text{ and } C_h = \sum_{i=1}^{n_h} b_{hi}^2 f_{hi}^2 (q_{hi}^s)^2.
$$

The authors derive the following estimators for the mean and variance:

$$\hat{\mu} = \frac{\sum_{h=1}^{r} A_h}{\sum_{h=1}^{r} E_h} = \frac{A}{T} = \frac{\text{total actual deaths over all companies}}{\text{total expected deaths over all companies}}$$

and

$$\hat{\sigma}^2 = \frac{\sum_{h=1}^{r} E_h(\hat{m}_h - \hat{\mu})^2 - \hat{\mu}\left(\sum_{h=1}^{r} \frac{B_h}{E_h} - \frac{1}{T}\sum_{h=1}^{r} B_h\right) + \hat{\mu}^2\left(\sum_{h=1}^{r} \frac{C_h}{E_h} - \frac{1}{T}\sum_{h=1}^{r} C_h\right)}{T - \frac{1}{T}\sum_{h=1}^{r} E_h^2 - \sum_{h=1}^{r} \frac{C_h}{E_h} + \frac{1}{T}\sum_{h=1}^{r} C_h}.$$

4. LF Analysis Applied to M Financial's Data: Qualitative Results

We applied the LF method to M Financial's data to compute credibility factors for the A/E ratios based on M's experience data. In this section, we describe our calculations and share some qualitative observations about our results.

We computed the credibility factors using four different methods. For each method, we computed an aggregate credibility factor as well as specific factors for sex and smoker status. Thus, for each of the four methods, we computed five credibility factors: aggregate, male nonsmoker, female nonsmoker, etc. In each case, the actual mortality is based on M Financial's 2012 Preliminary Experience Study and the expected mortality is based on the 2008 VBT.

First, we used the methods described in Klugman et al. (2009) and Section 3.2 to compute the LF credibility factors for M Financial's observed A/E ratios based on claim counts. We computed the credibility factors using both the "exact" and approximate expressions given in (10). We denote the resulting credibility factors by Z_c^e and Z_c^a, respectively. We also compute the credibility factors for the *overall* mortality rate

$$\hat{q} = \frac{\text{number of claims}}{\text{total exposures}}.$$

We denote the resulting credibility factor by Z_q. Finally, we used the methods described in Klugman et al. (2009) and Section 3.2 to compute the LF credibility factors for M Financial's observed A/E ratios based on retained net amount at risk (NAR). We denote the resulting credibility factor by Z_{NAR}.

The values of Z_c^a, Z_c^e and Z_q were remarkably close for the aggregate credibility factor and for each sex / smoker status combination. More specifically, the maximum relative difference among the factors was 3%. Thus, while computing a credibility factor for for the overall mortality rate $\hat{q}$ is too simplistic, in this case, the resulting credibility factors were remarkably close to the credibility factors based on claim counts.

The credibility factors Z_{NAR} for the A/E ratio based on retained NAR were significantly lower than the credibility factors based on claim counts; the relative difference ranged from 47% to 64%, depending on the sex / smoker status. This is not surprising. As we observed in Remark 2 of Section

2.1.3, the credibility factor should decrease as the variance increases. When we compute the A/E ratio for NAR, there is an additional source of randomness, namely, whether claims occurred for high- or low-value policies.

This raises the question of whether to use claim counts or NAR as the basis for the credibility factors. According to Klugman (2011), "If there is no difference in mortality by amount, there really is no good statistical reason to use amounts. They add noise, not accuracy. If there is a difference, then the mix of sales has an impact. As long as the future mix will be similar, this will provide a more accurate estimate of future mortality."

5. Qualitative Comparison of Credibility Methods Using a Simulated Data Set

As we described in Section 2, the LF method is easy to apply, as it relies only on company-specific data; however, it has several significant shortcomings. The GA method addresses these shortcomings, but requires data from other companies. Because experience data is proprietary, GA is rarely used in practice.

In this section, we examine the performance of the LF and GA credibility methods on a simulated data set, and we compare the resulting credibility factors with an intuitive, though not mathematically grounded "credibility factor."

5.1. Overview

In the simulation, we created a dataset consisting of 1 million individuals. More specifically, we created 20 risk classes or populations of 50,000 individuals each and computed the A/E ratio for each of the risk classes. Expected mortality is based on the 2008 VBT, and actual mortality is based on simulation from the table, or a multiple of it. Thus, the hypothetical means - the A/E ratio for each risk class - are known. To generate the experience data, we sampled from the risk classes and computed the observed A/E ratios.

Then, given the observed A/E ratios, the "industry average" (expected value of the hypothetical means or overall A/E ratio for the universe), and the known hypothetical means, we computed credibility factors three different ways: GA, LF, and an intuitively pleasing, though not mathematically grounded, benchmark credibility factor. We applied the GA and LF methods as in Klugman et al. (2009).

We contrast the credibility results in a series of figures. The most notable result is that LF with the "standard" range and probability parameters yields a significantly lower credibility factor than the other methods when the number of claims is small. Thus, the LF method might understate the credibility for companies with good mortality experience.

5.2. Generating the "Universe"

We generated 20 risk classes, or populations, of 50,000 people each; thus, the universe consists of 1 million individuals. Each of the 20 populations had a different age distribution and had an A/E ratio prescribed *a priori*. More specifically, the A/E ratio for each risk class was prescribed by scaling the 1-year 2008 VBT probabilities q_x by a multiplier $\alpha_h \in \{0.73, 0.76, \ldots, 1.30\}, h = 1, \ldots, 20$.

Using the scaled 2008 VBT table, we computed $_tq_x$ for $t \leq 20$ and for various values of x. These values give the cumulative distribution function (CDF) of the random variable $T(x)$, the remaining future lifetime of (x). Then, we generated the outcomes of $T(x)$ in the usual way; namely, we generated outcomes from the uniform distribution on $(0, 1)$ and inverted the CDF to determine the outcome of the random variable $T(x)$. If $T(x) < 20$, then a claim occurred during the 20-year period; otherwise, no claim occurred.

We then calculated the following:

• the ratio m_h of actual deaths to expected deaths over the 20-year time period, $h = 1, \ldots, 20$; in other words, we computed the hypothetical mean for each of the 20 risk classes. These values ranged from 0.71 to 1.28.

• the overall A/E ratio $\mu = 0.95$ for the universe of 1 million individuals.

5.3. Generating the Experience Data and Computing the LF and GA Credibility Factors

We then generated experience data for each of the 20 risk classes. We view the experience data as the experience of a particular company, as in Klugman et al. (2009). We fixed the number n of policyholders (e.g., $n = 500, n = 1500, \ldots$) and randomly selected n "lives" from from each of the risk classes. We computed the observed A/E ratio $\hat{m}_h$ and we computed the credibility factors $\tilde{Z}_h^{LF}$ and $\tilde{Z}_h^{GA}$ using the LF and GA methods, respectively, as described in Klugman et al. (2009) and in Section 3.
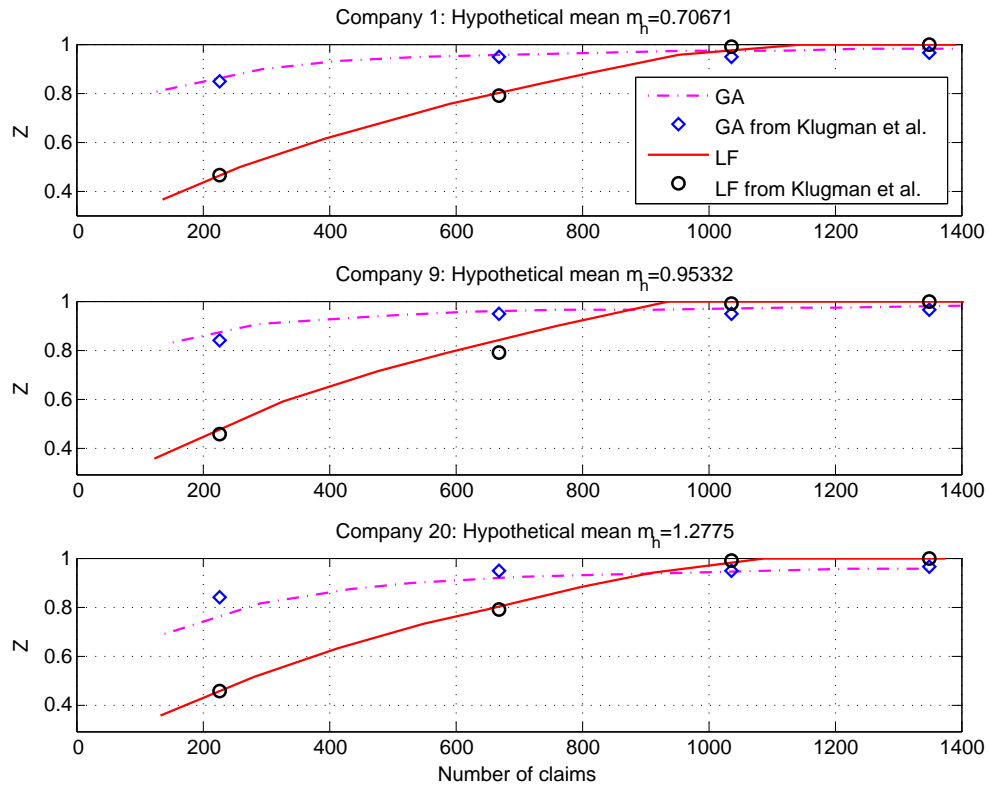
In Figure 1, we show the LF and GA credibility factors for three "companies" (risk classes) from our simulated data set and contrast our results with the results from Klugman et al. (2009), which are based on real data. The results are remarkably consistent.

We observe that, without exception, the LF factors are considerably lower than the GA factors. It is not surprising that the factors differ significantly, as the underlying methods are so different.

5.4. An Intuitively Appealing Benchmark "Credibility Factor"

We wanted to compare the LF and GA credibility factors with an intuitively appealing benchmark. To achieve this, we posed the question, "In repeated draws of experience data from the risk classes, how frequently is the observed A/E ratio $\hat{m}_h$ closer than the "industry average" μ to the true value

Figure 1   We show the LF and GA credibility factors for three "companies" (risk classes) from our simulated data set and contrast our results with the results from Klugman et al. (2009), which are based on real data. The results are remarkably consistent.

(hypothetical mean) m_h?" Thus, we generated 2,000 trials of experience data for n policyholders from each risk class.

We introduce the following notation.

- Let h refer to the risk class or population; $h = 1, \ldots, 20$.
- Let $n =$ number of policyholders in the company's experience data; $n = 500, 1500, \ldots$.
- Let $i =$ trial; $i = 1, \ldots, 2,000$.
- Let $\mu =$ the A/E ratio for the universe of 1 million individuals. In the simulation, we had $\mu \approx 0.95$.
- Let $\hat{m}_{hni} =$ the observed A/E ratio for company h, trial i, when there are n policyholders in the group.
- Let m_h be the true A/E ratio for population h. Recall that this was prescribed *a priori* when we generated the 20 populations.
- Let Z_{hn} be the credibility factor for company h when there are n policyholders in the group.

We computed an intuitively appealing benchmark credibility factor Z_{hn} as follows. For each of the 2,000 trials, define

$$I_{hni} = \begin{cases} 1 & |\hat{m}_{hni} - m_h| < |\mu - m_h| \\ 0 & \text{otherwise.} \end{cases}$$

Thus, I_{hni} indicates whether the observed A/E ratio $\hat{m}_{hni}$ or the "industry average" $\mu = 0.95$ is more representative of the true mean m_h.

Then, we define $Z_{hn} = \frac{1}{2000} \sum_{i=1}^{2000} I_{hni}$. Thus, Z_{hn} computed this way is the proportion of the 2,000 trials for which the experience data was more representative of the risk class's A/E ratio than the universe's A/E ratio μ.

This definition of Z has intuitive appeal; the weight given to $\hat{m}_h$ should be the proportion of the time that it is a better representation of the true mean m_h than μ. However, we emphasize that there is no mathematical underpinning for this choice. It seems that we could just as easily use the square root of the proportion, or the proportion squared, for example. Moreover, in practice, one could not compute the benchmark factor. The benchmark factor simply allows us to compare, in our simulated universe, the LF and GA results with an intuitively appealing measure.

We wish to compare the benchmark factor to the LF and GA factors computed in Section 5.3. To make the comparison meaningful, we must express Z_{hn}^B as a function of the number of claims. Thus we introduce the following notation:

- Let $d_{hni} =$ actual number of deaths for company h, trial i, when there are n policyholders.
- Let $d_{hn} = \frac{1}{2000} \sum_{i=1}^{2000} d_{hni}$; i.e., d_{hn} is the average actual number of deaths over the 2000 trials for company h when there are n policyholders.
- Denote d_{hn} by d_h; we will suppress the index n. Observe that d_h is analogous to A_h, the actual number of claims for company h, from Klugman et al. (2009).

• Let $\tilde{Z}_{hd}^B$ be the benchmark credibility factor for company h when there are d claims. Thus we express the credibility factor $\tilde{Z}$ as a function of $d_h = d_{hn}$. This will make the comparison with the LF and GA results meaningful.

• Let $\tilde{Z}_{hd}^{LF}$ and $\tilde{Z}_{hd}^{GA}$ be the credibility factors for company h when there are d claims computed via the LF and GA methods, respectively. We computed these factors in Section 5.3.

5.5. Qualitative Comparison of the Credibility Methods

We computed the benchmark factor in Section 5.4. We also computed credibility factors for the simulated data set using the GA and LF methods as in Klugman et al. (2009). For LF, we used the common parameter choices $r = 0.05$ and $p = 0.9$ and the Canadian standard $r = 0.03$ and $p = 0.9$, Canadian Institute of Actuaries (2002). In other words, under the LF method, we assign full credibility if the probability that the relative error in the A/E ratio is less than 0.05 (or 0.03) is at least 90%. Recall that from the approximate expression in (10), these parameter choices yield 1,082 and 3,007 claims, respectively, as the standards for full credibility.

We summarize our observations for the 20 companies (risk classes) below, and show the credibility factors for 3 of the 20.

1. Generally speaking, the benchmark and GA factors are similar and considerably higher than the LF factors when the number of claims is small; see Figures 2 and 4.

2. The exception to Observation 1 occurs when the hypothetical mean m_h is close to the overall population mean $\mu \approx 0.95$; see Figure 3. This is not surprising; the benchmark factor is the relative frequency in 2,000 trials that the observed A/E ratio $\hat{m}_h$ is closer to the true hypothetical mean m_h than the overall population mean μ. When m_h is very close to μ, it is unlikely that $\hat{m}_h$ will land closer; i.e., the event $|\hat{m}_h - m_h| < |\mu - m_h|$ is unlikely, resulting in a smaller benchmark Z.

3. For our simulated data set, and for the real data in Klugman et al. (2009), the GA method produces significantly higher credibility factors at low numbers of claims than the LF method with $r = 0.05$. Of course, the difference is even more pronounced when we choose the Canadian standard $r = 0.03$.

6. Conclusion and Recommendations

Without exception, the LF factors based on the full credibility requirement of 1,082 claims were significantly lower than the GA factors when the number of claims is small. Of course, the disparity is greater when we use 3,007 as the standard for full credibility.

Our analysis suggests that the LF method may significantly understate the appropriate credibility factor for populations with exceptionally good mortality experience, such as M Financial's
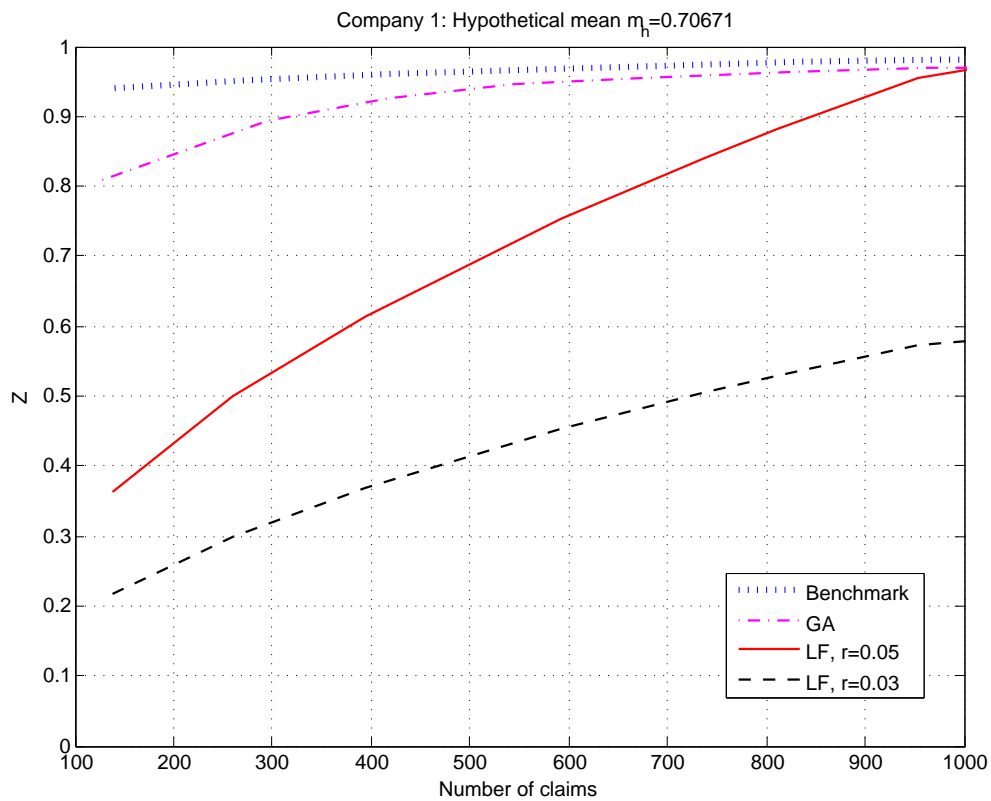
Figure 2 We contrast the Benchmark, GA, and LF credibility factors for "company" (risk class) 1, whose hypothetical mean is $m_1 \approx 0.707$.
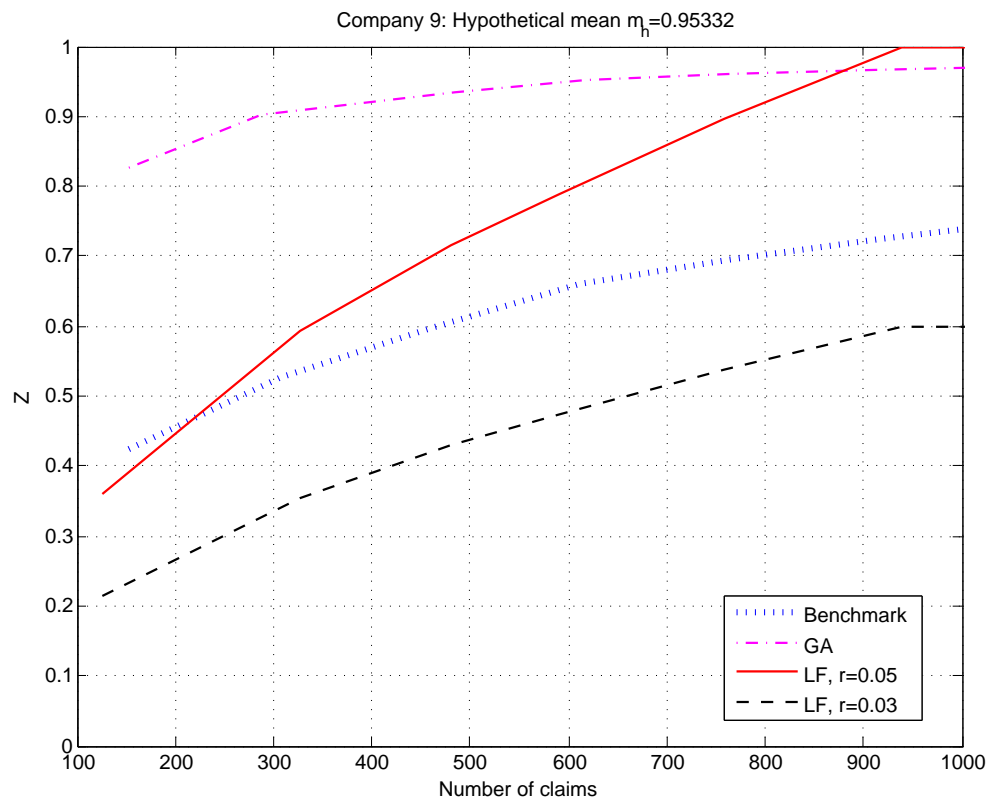
Figure 3   We contrast the Benchmark, GA, and LF credibility factors for "company" (risk class) 9, whose hypothetical mean is $m_9 \approx 0.953$.
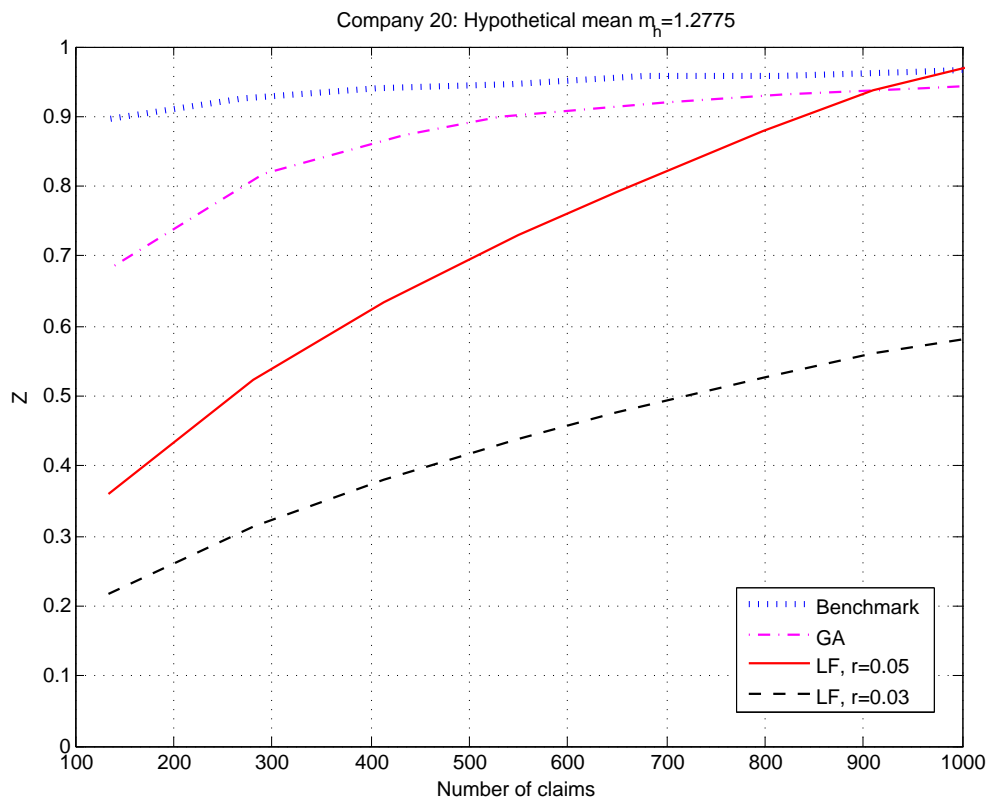
Figure 4   We contrast the Benchmark, GA, and LF credibility factors for "company" (risk class) 20, whose hypothetical mean is $m_{20} \approx 1.278$.

clientele. Thus, our analysis provides support for awarding higher credibility factors than the LF method yields when the number of claims is low.

The NAIC recognizes that there is a need for life insurance experience data in support of PBR. The PBR Implementation Task Force are developing an Experience Reporting Framework for collecting, warehousing, and analyzing experience data. This is ongoing work; see AAA (2016) and NAIC (2016), for example.

Acknowledgments

References

Actuarial Standards Board. 2013. Actuarial Standard of Practice 25.

Actuarial Standards Board. 2014. Proposed Actuarial Standard of Practice: Principle-Based Reserves for Life Products, Second Exposure Draft. http://www.actuarialstandardsboard.org/wp-content/uploads/2014/10/PBR_second_exposure_draft_August2014.pdf.

American Academy of Actuaries. 2008. Credibility Practice Note.

American Academy of Actuaries. 2016. PBA Perspectives. http://www.actuary.org/email/2016/pba/PBA_April_2016.html

Bühlmann, H. and A. Gisler. 2005. *A Course in Credibility Theory and its Applications.* Berlin: Springer.

Canadian Institute of Actuaries. 2002. Expected Morality: Fully Underwritten Canadian Individual Life Insurance Policies. http://www.actuaries.ca/members/publications/2002/202037e.pdf.

Dannenburg, D., R. Kaas, and M. Goovaerts. 1996. Practical Actuarial Credibility Models. Institute of Actuarial Science and Econometrics, University of Amsterdam, Amsterdam, The Netherlands.

Frees, E., V. Young, and Y. Luo. 1999. A Longitudinal Data Analysis Interpretation of Credibility Models. *Insurance: Mathematics and Economics* 24:229-247.

Goovaerts, M. and W. Hoogstad. 1987. *Credibility Theory, Surveys of Actuarial Studies.* Rotterdam: Nationale-Nedelanden N.V.

Klinker, F. 2011. Generalized Linear Mixed Models for Ratemaking: A Means of Introducing Credibility into Generalized Linear Model Setting. Casualty Actuarial Society e-Forum. http://www.casualtyactuarialsociety.org/pubs/forum/11wforumpt2/Klinker.pdf.

Klugman, S. 2011. Mortality Table Construction and Forecasting. Society of Actuaries Symposium: Information, Insight, and Inspiration. https://www.soa.org/Professional-Development/Presentations-Archive/2011/2011-SOA-Symposium—Information,-Insight—Inspiration-(Shanghai,-China).aspx.

Klugman, S., H. Panjer, and G. Wilmot. 2012. *Loss Models: From Data to Decisions, Fourth Edition.* Hoboken: John Wiley and Sons.

Klugman, S., T. Rhodes, M. Purushotham, S. Gill, and MIB Solutions. 2009. Credibility Theory Practices. Society of Actuaries.

National Association of Insurance Commissioners (NAIC) and the Center for Insurance Policy and Research. (2013) Principle-Based Reserve (PBR) Educational Brief. www.naic.org/documents/committees_ex_pbr_implementation_tf_130621_educational_brief.pdf.

National Association of Insurance Commissioners. 2016. Principle-Based Reserving (PBR) Implementation Plan, 3/7/2016. http://www.naic.org/documents/committees_ex_pbr_implementation_tf_150323_pbr_implementation_pla

Nelder, J. and R. Verrall. 1997. Credibility Theory and Generalized Linear Models. *ASTIN Bulletin* 27:71-82.