Applying Credibility Concepts to Develop Weights for Ultimate Claim Estimators

Rajesh Sahasrabuddhe, FCAS, MAAA

June 19, 2014 *

*LaTeXed on July 19, 2014

In estimating ultimate claim and claim expense amounts, actuaries often rely on estimates developed using multiple actuarial methods. Combining these estimates is often left to the actuary's professional judgment. That judgment generally reflects informed but subjective opinion on the relative stability and responsiveness of various methods and the reasonableness of the results of those methods.

This paper proposes a more quantitative approach. The approach is based on credibility concepts which are often used in ratemaking contexts but have yet to find their way into this particular aspect of estimating unpaid claims and claim expenses. As with the ratemaking context, credibility is based on the variance of estimators. However the application to unpaid claim estimates requires a different approach. That approach is the subject of this paper.

Keywords Credibility, Reserving

1. Introduction

1.1. Research Context

The author was not able to locate any prior research on this specific topic. During the course of preparing this paper, the author identified Rehman & Klugman[RK10] as having some similarity in underlying concepts - though that paper has a much different application.

The CAS Taxonomy for this paper is as follows:

- Actuarial Applications and Methodologies > Reserving >Management Best Estimate
- Actuarial Applications and Methodologies > Reserving >Reserving Methods

1.2. Objective

Typically[1], actuarial estimates of ultimate claims are based on a review of multiple actuarial indications (such as those based on the chain-ladder and Bornhuetter-Ferguson methods) and the actuary's professional judgment. We can think of the final (selected) actuarial estimate as a weighted average of actuarial indications and expert opinion and/or prior knowledge. The mathematical description of that process would likely be similar to the following:

$$\hat{C} = \hat{\boldsymbol{I}} \times \boldsymbol{Z} + P \times (1 - Z) \tag{1}$$

which we can recognize as being similar to credibility-weighted averages commonly used in ratemaking contexts.

In the context of estimating ultimate claims[2]:

$C =$ ultimate claims

$\hat{C} =$ an estimator of ultimate claims

$\hat{\boldsymbol{I}} =$ a vector of actuarial projections (indications) of C

$P =$ the actuary's prior estimate of ultimate claims; possibly equal to $\hat{C}_{t-1}$

$\boldsymbol{Z} =$ a vector of credibility factors

$Z =$ the sum of the elements of $\boldsymbol{Z}$

There are several important considerations with respect to these variables:

[1] This section is based strictly on the author's observation of common practices and does not imply that all actuaries use the approach described.

[2] In this paper, the term "claim" is used rather than "loss" to be consistent with Actuarial Standard of Practice No. 43, *Property/Casualty Unpaid Claim Estimates*

- In Equation (1), we assume that $\hat{I}$, P and, by consequence, $\hat{C}$, are all unbiased[3] estimators of C. The author is aware that certain papers challenge this assumption. Measuring and correcting for bias of actuarial methods is outside the scope of this paper.
- "Prior" in this paper is not used in the rigorous statistical sense of that term. Rather, the term refers to an estimate that management or the actuary may be "targeting" and is based on the author's observation that, in practice, reasonableness of an estimate is often evaluated relative to some other (benchmark) estimate[4]. For example, the implied loss ratios of estimates based on the first evaluation of an experience period are assessed through reconciliation of those estimates to the loss ratio used in establishing the premium.

In the usual judgment-based model, "credibility" for method i is then usually assigned based on the distance between $\hat{I}_i$ and P. That is:

$$Z_i \propto \frac{1}{d(\hat{I}_i, P)} \tag{2}$$

$$Z_i = f_j\left(\frac{1}{d(\hat{I}_i, P)}\right) \tag{3}$$

where d represents a generic difference function such as absolute or squared difference. As commonly applied, f_j might be termed the "actuarial judgment function." In Figure 1, we present a visualization of that "actuarial judgment function."

The objective of this paper is to offer an approach to *calculate* Z rather than use an "actuarial judgment function."

1.3. Outline

The remainder of the paper proceeds as follows:

- Section 2 is a discussion of approaches that are commonly used and the approach proposed by this paper.
- Section 3 provides the theory and practice of the credibility weighted approach. A workbook accompanies this report to supplement the reader's understanding of the application in practice.
- Section 4 provides a generalized model and summarizes the findings of this paper.

[3] Or, more precisely, that we are aware of and can adjust for any biases in $\hat{I}$ or P.

[4] As with all generalizations, this of course is not universally true. For example, if there were a known shock loss, the benchmark may be disregarded.

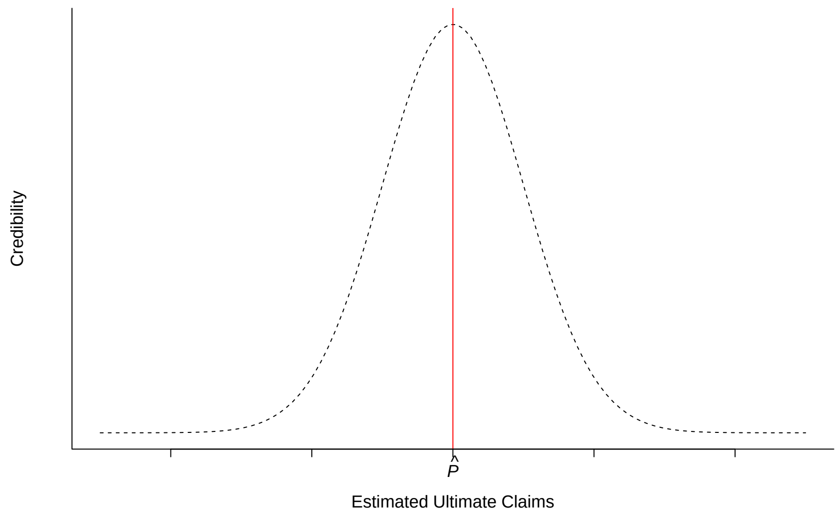

Figure 1: The Actuarial Judgment Function

2. Background and Methods

2.1. Credibility in a Ratemaking Context

Actuaries are familiar with the use of credibility in a ratemaking context. In that context, the goal is to assign predictive value to the experience of a class relative to the predictive value of the experience of an aggregation of classes. Although there are various models to estimate this predictive value, credibility, generally, is proportional to the variance between classes (also referred to as the *variance of hypothetical means, VHM*) and inversely proportional to the average variance within classes (commonly referred to as the *expected value of process variance* (*EVPV*))[5].

2.2. Credibility in a Reserving Context

The extension of credibility to a reserving context may not be immediately clear until we consider the *general* definition of the term "credibility." That is, we need to consider credibility as a measure of the predictive value, possibly measured on a relative basis, of an estimator. We are not referring to credibility as calculated under a *specific* model. In the approach presented in this paper, credibility for each available actuarial method is

[5]Though these terms are used in several papers, to the best of the author's knowledge, they were first used (or at least popularized) by Philbrick [Phi81].

developed based on the variance of that method *relative to the variance of each of the other methods.*

For the moment, assume that we have two competing estimates, I_1 and I_2, for C. Then, the credibility model becomes:

$$\hat{C} = \hat{I}_1 \times Z_1 + \hat{I}_2 \times Z_2 \tag{4}$$

There is an important difference between Equation (1) and Equation (4). The latter does not include the *prior* estimate. In ratemaking, the prior estimate is considered to the extent that we cannot assign credibility to an indication(s). In unpaid claim estimation, we have models (such as the loss ratio or Bornhuetter-Ferguson models) that allow for the consideration of a "prior" estimate so we need not consider the estimate separately and explicitly. This paper presents an approach where we assign 100% of the credibility to available estimates based on the *relative* variance.

In evaluating variance, we consider the residual, or the difference between the observed prediction and the "best" prediction. Furthermore, rather than consider the variance of I_i, we consider the variance of the underlying actuarial method or model i at maturity j. We denote the observed indication $\hat{M}_{i,j}$.

Then, the credibility weighted average of estimators can be written as:

$$M_{1,j} \times Z_{1,j} + \ldots + M_{n,j} \times Z_{n,j} \tag{5}$$

Finally, we define credibility in this context as the probability that the error (residual) of $M_{i,j}$ is smaller than that of $n-1$ competing estimates $M_{1,j} \ldots M_{i-1,j}$, $M_{i+1,j} \ldots M_{n,j}$. where $\sum Z_{1,j} \ldots Z_{n,j} = 1$.

2.3. Residual Errors

The proposed approach is based on an analysis of the distribution of residuals (ε) of each method, M_i, at a particular maturity j. As an illustration of the concept of the variance of residuals, we consider chain-ladder estimates based on an analysis of a development triangle. In that situation, we would have a series of observed predictions at 12 months which are the product of the 12-month claim development factor and previously observed claims at 12 months for prior experience periods. After normalizing the triangle for differences between experience periods[6], we would then calculate the residuals as the difference between those predictions and the "best" values. This is only one method for developing an estimate of the variance of the residuals. We explore the issue further in Section 3.3.1.

[6] For example, such difference may include those attributable to changes in costs level or exposure volume.

3. Results and Discussion

Rather than separating theory and practice/example, I have elected to present the theoretical underpinnings of the proposed credibility model in the context of a minimally specified example. I intend for this presentation to demonstrate the practicality of this approach. This presentation also allows for easier identification of assumptions underlying the credibility model.

3.1. Proposed Credibility Model

For our minimally specified example, we will assume the following:

- We have two competing methods, the paid chain-ladder (Method 1, M_1) and the reported incurred chain-ladder (Method 2, M_2). We are concerned with the estimate at 12 months maturity. (The maturity is unimportant to this example but it is helpful to define the context.)
- Assume that ultimate claims (C) are 1,000. This estimate provides a sense of scale though it is not necessary for our minimally specified example as we are provided the distribution of residuals. (In a different circumstance, we may have the coefficient of variation of the residuals. In this case, the estimate of ultimate claims would be necessary.)
- We assume that the methods are unbiased. Therefore the means of the residuals for both models are assumed to be 0 ($\mu(\varepsilon_1) = \mu(\varepsilon_2) = 0$).
- In our example, the paid chain-ladder has more variability in predictions than the reported incurred chain-ladder. The residual errors are assumed to be normally distributed with standard deviations of 200 and 300 for the reported incurred chain-ladder and paid chain-ladder methods, respectively, ($s(\varepsilon_1) = 300; s(\varepsilon_2) = 200; E[s] = \sigma$) as presented in Figure 2.

Under our definition, the credibility of the reported incurred chain-ladder is the probability that the error of M_2 (random variable denoted X_2) is less than or equal to the error of M_1 (random variable denoted X_1).

So for any $X_2 = x_2$ (where x_2 is an observation of X_2), we have the following possibilities:

1. $|X_1| < |x_2|$ (Credibility to Method 1)
2. $|X_1| > |x_2|$ (Credibility to Method 2)

Given the symmetric distribution centered around 0, for simplicity, we use only the positive domain of x and consider both tails of the distribution of x_1. We use F and f to represent the distribution and density functions, respectively, of the residuals. We then have the

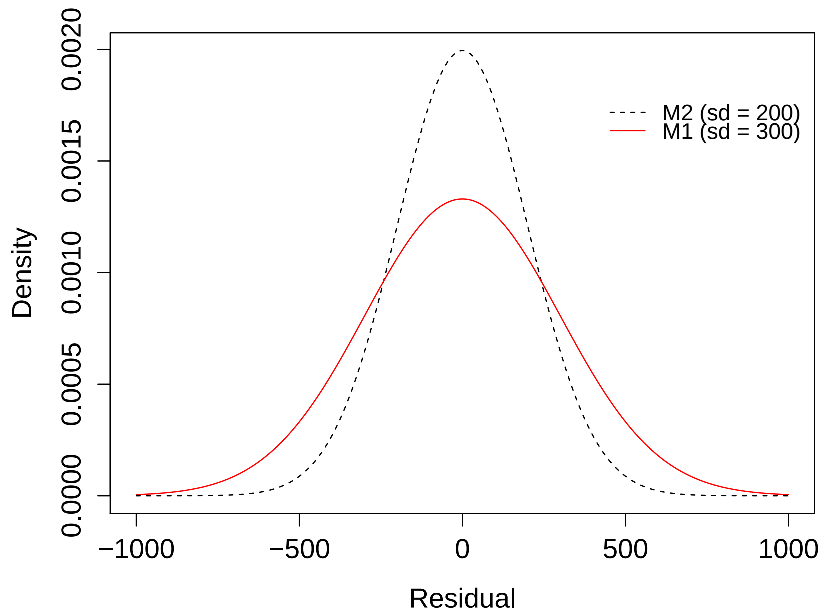

Figure 2: Distribution of Residual Errors

following credibility calculation:

$$Z_2/2 = \int_0^\infty 2\,[1 - F_1(x)]\, f_2(x)\, dx \tag{6}$$

In words, Equation (6) states:

> "Over the domain of positive values of x (i.e. $\int_0^\infty$), the credibility assigned to Method 2 (i.e. Z_2) is the probability that the error of Method 1 is greater than x (i.e. $(1 - F_1(x))$) or less than $-x$ (which by symmetry is also equal to $1 - F_1(x)$) given that $X_2 = x$ (i.e. $f_2(x)\,dx$)."
>
> The *2* inside the integral provides consideration for both values of X_1 less than $-x_2$ and greater than $+x_2$. For example, if $x_2 = 100$, we would assign credibility to Method 2 for X_1 probabilistically greater than 100 or X_1 probabilistically less than -100.
>
> The *2* on the left-side of Equation (6) is necessary as our limits of integration only consider one-half the domain of possible x values.

We should also recognize that if we were evaluating over the domain of negative x values, we would replace $1 - F_1(x)$ with $F_1(x)$. This is due to the property of symmetric distributions centered at 0 where:

$$F(-x) = (1 - F(x)) \qquad x > 0$$
$$F(x) = (1 - F(-x)) \qquad x < 0.$$

This is further demonstrated in Appendix A, where we provide Equation (6) with separate terms for values of x that are less than 0 and those greater than 0.

We can then expand Equation (6) as follows:

$$Z_2/2 = 2\int_0^\infty f_2(x)\,dx - 2\int_0^\infty F_1(x)f_2(x)\,dx$$
$$Z_2/2 = 1 - 2\int_0^\infty F_1(x)f_2(x)\,dx \tag{7}$$
$$Z_2 = 2 - 4\int_0^\infty F_1(x)f_2(x)\,dx \tag{8}$$

Equations (7) and (8) are intuitively appealing as they state that credibility is lost when a competing estimate has a lower error. The constants of 2 and 4 in Equation (8) may seem disconcerting at first but we should recognize that they are twice what they would be had we integrated over both positive and negative values of x.

Furthermore, if we consider the limiting case where Method 1 has no error (i.e. it is a perfect indicator of C) then:

1. $F_1(0) = 1$
2. $\therefore \int_0^\infty F_1(x)f_2(x)\,dx = 1/2$
3. $\therefore Z_2 = 0$

We can use the trapezoidal rule to numerically integrate Equation (8). If we apply the trapezoidal rule over 1,000 evenly-spaced (unit) intervals between 0 and 1,000, we can calculate the value of the integral to be 0.342. The resulting credibility to Method 2 is 0.630.

We can also reverse the subscripts and calculate the credibility of Method 1 using Equation (8).

$$Z_1 = 2 - 4\int_0^\infty F_2(x)f_1(x)\,dx$$

The resulting credibility of Method 1 is 0.379. The sum of these credibilities don't quite equal 1 but that is simply the result of the approximation of the numerical integration. We can address this issue through normalization as presented in Table 1.

	Raw Credibility	Normalized Credibility
Method 1	0.379	0.376
Method 2	0.630	0.624
Total	1.009	1.000

Table 1: Two Method Example

3.2. Simulation

We can also use simulation if we want to avoid the effort of numerical integration. We present the R[R C13] code to estimate credibilities via simulation.

```
> set.seed(12345)
> trials <- 1000
> pd.dev.errors <- abs(rnorm(n = trials, mean = 0, sd = 300))
> rptd.dev.errors <- abs(rnorm(n = trials, mean = 0, sd = 200))
> pd.dev.cred <- length(which(pd.dev.errors < rptd.dev.errors)) / trials
> rptd.dev.cred <- length(which(rptd.dev.errors < pd.dev.errors)) / trials
> pd.dev.cred
[1] 0.375
> rptd.dev.cred
[1] 0.625
```

Alternatively, one could use the `integrate` function with the following code[7].

```
> f <- function(x) pnorm(x, 0, 300) * dnorm(x, 0, 200)
> integral <- integrate(f, 0, Inf)$value
> rptd.dev.cred <- 2 - 4 * integral
> rptd.dev.cred
[1] 0.6256659
```

We note that the results of the simulation are quite close to those calculated from numerical integration.

3.3. Assumptions

The minimally specified model includes several simplifying assumptions that we explore in this section.

[7]This code was contributed by Mark Mordechai Goldburd. Mr. Goldburd reviewed this paper for the CAS 2014 Fall *E-Forum*

- We have assumed that the residual errors are normally distributed. Is this reasonable?
- We were provided with the standard errors in the examples in Section 3.1. What approaches can we use to develop the estimated standard error?
- Can we consider the management's recorded estimate within this model?

3.3.1. Residual Standard Error

Modeling the distribution of residuals is a complex topic that is outside the scope of this paper. We should keep in mind that we are more focused on *relative errors* than *absolute errors*. Furthermore, even when we cannot calculate that uncertainty, we should be able to assign uncertainty based on judgment. That judgment (e.g. the uncertainty / volatility of the reported loss development method at 12 months is ±200) is more "testable" than the implicit assignments of credibility to methods based on the "actuarial judgment function."

Generally, we would expect that the variance for any method would decrease over time as paid claims are a greater percentage of ultimate claims. In this model, we calculate credibility separately for each maturity. As such, shifts in credibility weights between methods will occur due to *differences in the rate of decrease* in the variance as a function of maturity.

Further, we would *expect* that methods that emphasize stability will have lower variances at early maturities than those that emphasize responsiveness. However, this will depend on the deviations between the *a priori* expected ultimate claims and the current best estimates.

Below, we present one example approach to developing error estimates using the `auto$PersonalAutoIncurred` data included in the `ChainLadder` [GMZ13] package for `R`.

1. In Table 2, we present the triangle of paid claims and volume-weighted development factors.
2. In Table 3, we present the indications of ultimate claims based on the paid development method and the current valuation of paid claims.
3. In Table 4, we present the indications of ultimate claims based on paid development factors and historical valuation of paid claims from Table 2.
4. In Table 5, we present the triangle of residuals. Those residuals are calculated as the difference between the ultimates in Table 4 and those in Table 3.

Origin	Maturity 1	2	3	4	5	6	7	8	9	10
1	101125	209921	266618	305107	327850	340669	348430	351193	353353	353584
2	102541	203213	260677	303182	328932	340948	347333	349813	350523	
3	114932	227704	298120	345542	367760	377999	383611	385224		
4	114452	227761	301072	340669	359979	369248	373325			
5	115597	243611	315215	354490	372376	382738				
6	127760	259416	326975	365780	386725					
7	135616	262294	327086	367357						
8	127177	244249	317972							
9	128631	246803								
10	126288									
Incremental DF	1.990	1.285	1.137	1.064	1.031	1.017	1.006	1.004	1.001	1.000
Cumulative DF	3.278	1.647	1.282	1.128	1.060	1.028	1.011	1.005	1.001	1.000

Table 2: auto$PersonalAutoPaid{ChainLadder}

Origin	Current Claims	Cumulative Development Factor	Ultimates
1	353584	1.000	353584
2	350523	1.001	350874
3	385224	1.005	387150
4	373325	1.011	377432
5	382738	1.028	393455
6	386725	1.060	409928
7	367357	1.128	414379
8	317972	1.282	407640
9	246803	1.647	406485
10	126288	3.278	413972

Table 3: Ultimate Claim Estimates

> In Table 5, we also present the standard deviations of the residuals[8] for each age that may be used as a basis to select the standard error of the paid development method at each of the respective ages. Those standard deviations may be slightly understated in an absolute sense as our triangle includes the current valuation - which will have a deviation of 0 under this approach. However, we elected to include that diagonal to generalize the approach and allow for other estimates of the "best" estimate of ultimate claims.

This is only one example algorithm to estimate the standard errors for results of an actuarial method at a particular age. We note that the `ChainLadder` package includes many useful functions for developing these estimates and recommend that readers review the vignette accompanying that package.

3.3.2. Distribution of Residuals

The credibility model presented does not require the use of the normal model to describe the distribution of errors. Identification of the appropriate model is a complex topic that is outside the scope of this paper. However, it would seem reasonable to use a model that

[8]Mr. Goldburd also noted that the standard deviations presented in Table 5 are calculated as sample standard deviations from an estimated mean. If we assume residuals are centered on zero, it may be more appropriate to calculate standard deviations as root mean squared distance from 0, with n as the denominator rather than $n-1$, since a degree of freedom is not lost to an estimated mean. This would not apply to generalizations of the model (such as those presented in the Appendix) and the calculation of residual variance is not the focus of the paper.

	1	2	3	4	5	6	7	8	9	10
1	331488	345740	341804	344161	347521	350208	352263	352949	353706	353584
2	336129	334692	334188	341989	348668	350495	351154	351562	350874	
3	376747	375028	382190	389771	389826	388583	387831	387150		
4	375174	375122	385974	384275	381578	379587	377432			
5	378927	401227	404106	399865	394719	393455				
6	418797	427258	419182	412600	409929					
7	444549	431998	419324	414379						
8	416886	402278	407640							
9	421652	406485								
10	413972									

Table 4: Retrospective Ultimates

	1	2	3	4	5	6	7	8	9	10
1	-22096	-7844	-11780	-9423	-6063	-3376	-1321	-635	122	0
2	-14745	-16182	-16686	-8885	-2206	-379	280	688	-0	
3	-10403	-12122	-4960	2621	2676	1433	681	0		
4	-2258	-2310	8542	6843	4146	2155	-0			
5	-14528	7772	10651	6410	1264	-0				
6	8869	17330	9254	2672	0					
7	30170	17619	4945	-0						
8	9246	-5362	0							
9	15167	-0								
10	0									
sd	16105	12122	10270	6704	3676	2135	867	662	87	

Table 5: Residuals

is symmetric with a zero mode and mean. Rehman & Klugman [RK10] includes discussion related to the use of a normal distribution to describe reserve variability.

3.3.3. Credibility Assigned to Management Estimates

We can consider the recorded management estimate to simply be another indication. If we are able to compile a triangle of prior recorded estimates, we can apply a model similar to that presented above for the paid chain-ladder method. Additionally, review of Rehman & Klugman [RK10] may be useful in determining a model for describing such errors.

4. Conclusion

In this section, we summarize the findings of the research presented in this paper.

4.1. Principal Finding

In this paper, we have proposed a method for weighting methods that is based on the uncertainty of the estimate. We recognize that developing measures of that uncertainty is not a trivial matter.

Equation (6) is the primary finding of working through the minimally specified example. Equation (6) may be generalized for n methods. The initial generalization is presented below as Equation (10). That is, the credibility of method n relative to methods $1 \dots n-1$ may be calculated as follows.

$$Z_n/2 = \int_0^\infty 2^{n-1}\{[1-F_1(x)]\dots[1-F_{n-1}(x)]\}f_n(x) \tag{9}$$

$$Z_n = \int_0^\infty 2^{n}\{[1-F_1(x)]\dots[1-F_{n-1}(x)]\}f_n(x) \tag{10}$$

Equation (11) presents the final generalization and the primary finding of this paper. Specifically that the credibility of method i relative to methods $1 \dots i-1, i+1 \dots n$ is calculated as:

$$\begin{aligned} Z_i = \int_0^\infty & 2^{n}\{[1-F_1(x)]\dots[1-F_{i-1}(x)] \\ & [1-F_{i+1}(x)]\dots[1-F_n(x)]\}f_i(x) \end{aligned} \tag{11}$$

4.2. Distribution of Residuals

We acknowledge that determination of the distribution of errors is not trivial. However, it would seem that assuming a normal distribution would be reasonable. In addition, so long as we have a consistent approach to determining errors from the various methods, we should be able to apply our model. Those "consistent" approaches may include an approach that assigns uncertainty based on professional judgment.

4.3. Simulation as an Alternative

In practice, using numerical integration to calculate credibility under the proposed model is not overly difficult. It may also be appealing as setting up the model requires that we think through issues of estimation uncertainty. In the companion workbook, we present the calculation for four methods with standard errors 100, 200, 400, and 600. Of course, as n increases, using simulation to estimate credibilities becomes a more attractive option.

5. Supplemental Information

5.1. Acknowledgment

The author thanks Marc Pearl, Nina Gau, Jennifer Wu and John Alltop for their review of this paper as part of the Casualty Actuarial Society (CAS) 2014 Reserves Call Paper Program. I also thank the reviewers for the CAS 2014 Fall *E-Forum.* Finally, I thank my Philadelphia-based Oliver Wyman colleagues (Jason Shook, Alexandra Taggart and Evelyn Shen) for their reviews. Any errors that remain are the sole responsibility of the author.

5.2. Additional R Packages

In addition to those previously cited, the R packages listed below were also used to develop the model theory and associated documentation.

- reshape2 [Wic07]
- xtable [Dah14]

5.3. Author Biography

Rajesh Sahasrabuddhe is currently a consulting actuary with Oliver Wyman. He has a Bachelor of Science degree in Mathematics / Actuarial Science from the University of Connecticut. He is a Fellow of the CAS and a Member of the American Academy of Actuaries. He has previously served as Chairperson of the CAS Syllabus Committee.

5.4. Further Research

I welcome the opportunity to collaborate with others on improving and furthering the research presented in this paper. I can be reached at `rajesh1004@gmail.com`. My current professional contact information is available to CAS members on the CAS website. Non-members may obtain contact information through the Online Directory of Actuarial Memberships at `https://actuarialdirectory.org`.

References

[Dah14] David B. Dahl.
xtable: Export tables to LaTeX or HTML, 2014.
R package version 1.7-3.

[GMZ13] Markus Gesmann, Daniel Murphy, and Wayne Zhang.
Statistical Methods for the Calculation of Outstanding Claims Reserves in General Insurance, 2013.
R package version 0.1.7.

[Phi81] Stephen W. Philbrick.
An examination of credibility concepts.
Proceedings of the Casualty Actuarial Society, LXVIII:195–21, 1981.

[R C13] R Core Team.
R: A Language and Environment for Statistical Computing.
R Foundation for Statistical Computing, Vienna, Austria, 2013.

[RK10] Zia Rehman and Stuart Klugman.
Quantifying uncertainty in reserve estimates.
Variance, Volume 04(Issue 01), 2010.

[Wic07] Hadley Wickham.
Reshaping data with the reshape package.
Journal of Statistical Software, 21(12):1–20, 2007.

Appendix A Expanded Credibility Model

Equation (6) presents a model that is simplified by assuming a symmetric distribution centered at 0. The model prior to simplification (with the assumption maintained) is as follows:

$$Z_2 = \int_{-\infty}^{0} 2F_1(x) f_2(x)\, dx + \int_{0}^{\infty} 2\left[1 - F_1(x)\right] f_2(x)\, dx \tag{12}$$

We could further relax the assumption of symmetry centered at 0. Doing so would produce the following:

$$\begin{aligned} Z_2 = &\int_{-\infty}^{0} F_1(x) f_2(x)\, dx + \int_{-\infty}^{0} [1 - F_1(-x)] f_2(x)\, dx + \\ &\int_{0}^{\infty} [1 - F_1(x)]\, f_2(x)\, dx + \int_{0}^{\infty} F_1(-x) f_2(x)\, dx \end{aligned} \tag{13}$$