Introduction to the National Council on Compensation Insurance Experience Rating Plan and Its Actuarial Methodology

Jon Evans

CONTENTS

1. INTRODUCTION

Workers compensation individual risk experience rating, and particularly the National Council on Compensation Insurance's (NCCI's) Experience Rating Plan (ERP), date back to the first years of workers compensation laws in the United States and the Casualty Actuary Society, in the 1910's and 1920's (see early volumes of *Proceedings of the Casualty Actuarial Society*, available at www.casact.org). What makes workers compensation experience rating special is the central role of credibility methods and the emphasis on a predictive perspective from these earliest years. It is also unique in its universal adoption in the United States. Today (2013), all states and the District of Columbia, require some type of workers compensation experience rating. All but 9 states allow for the use of NCCI's ERP for interstate rating.

NCCI provides extensive services related to the ERP, including assigning risk IDs, tracking the experience over time and across different insurance carriers for over 600,000 different employers, and then calculating ERP modification factors. NCCI files state loss costs (generically referred to as "rates"), and in some cases full rates that include underwriting expenses and profit provisions, per $100 of payroll by individual classification codes that categorize types of employment. *Manual rates* are determined using these classification code rates. The NCCI ERP produces an experience modification factor, or *mod*, by comparing actual loss experience for an individual risk to expected losses underlying manual rates. Predictive performance testing consistently confirms that, including situations where performance is less than ideal, *modified rates* (where the mod has been used to adjust manual rates) almost always produce a very big improvement over unmodified manual rates in the equity of rates for individual risks. This results in a secondary, but more often publicly cited, social benefit of incentivizing employers toward improving workplace safety.

In this study note, we will discuss not only the structure of the calculation in the latest version of the NCCI ERP, but also the actuarial methodology behind the plan. The reader who understands this study note should not only acquire a good understanding of the NCCI ERP specifically, but also a facility with some valuable and more general actuarial methods. However, we will not go deeply into very specific details of some of the elaborate underlying calculations, such as the calculations underlying partial factors for ELRs and D-ratios, or all the special rules in the ERP. These cumbersome details are not necessary to understand the principles, function, and basic operation of the ERP.

Many simplified examples are provided and are necessary to understand the material. Be warned that many of these examples will be to varying degrees unrealistically simple and/or unrealistically extreme in the interest of illustrating the concepts. Exercises are also provided. For a more in depth treatment of quantile testing, there is a related paper "The Optimal Number of Quantiles For Predictive Performance Testing of the NCCI Experience Rating Plan," by Evans and Dean. Exercises covering material in that paper are also included in this study note.

Another caution is that the actual ERP changes over time and across the many dozens of jurisdictions in which it applies. In a real world situation, where some part of this study note does not seem to agree with the actual NCCI ERP, the ultimate authoritative written sources are NCCI's Experience Rating Plan Manual and NCCI's Experience Rating Plan User's Guide.

The information is presented in the following sections:

- Section 2 explicitly lays out the predictive perspective, as distinguished from the static perspective, for actuarial models. Aside from being a more desirable objective, the predictive perspective makes better use of what data is available for individual risk rating.
- Section 3 explains how the individual risk loss process for workers compensation is complicated enough that standard textbook credibility and statistical models are not quite adequate for the task of experience rating.
- Section 4 discusses the generalizations to basic credibility models that are required by both the predictive perspective and the complicated loss processes described in the earlier sections.
- Section 5 explains how industry data collected through the NCCI Workers Compensation Statistical Plan is used to determine the experience period actual and manual basis expected losses that appear in the ERP formula.
- Section 6 covers some of the other miscellaneous adjustments, such as the maximum cap on ERP modification factors, which are part of the ERP.
- Appendix A provides a brief history of the evolution of the structure of the ERP.

Prerequisites for fully understanding this study note are a basic general knowledge of property and casualty actuarial methods and knowledge of actuarial models and the empirical determination of such models, particularly credibility theory, as presented in recent years in the syllabi for Online Courses CA1 and CA2 and Exams 3L and 4 of the Casualty Actuarial Society. Some familiarity with workers compensation ratemaking is also helpful.

2. PREDICTIVE ESTIMATION

In this section we will describe and contrast the predictive perspective, which has always been important for experience rating, and is growing in popularity for other actuarial models, with the more traditionally common static perspective. We will use overly simplistic models of loss processes in this section, and then introduce the complications that arise in more realistic loss processes in Section 3. The predictive perspective and complications in the loss process require a more generalized credibility method, which will be described in Section 4.

2.1 Contrast With Static Estimation

Although actuarial rate calculations are intended as estimates of future losses, most follow a more *static* perspective. A static perspective involves making the best estimate of the loss process underlying past experience and assuming the same loss process will underlie future losses, possibly including some anticipated systematic changes such as inflationary trends and different statutory benefit levels.

Example 2.1 Suppose a risk's true manual pure loss ratio prior to observation is randomly distributed with mean 100% and standard deviation 50%. Also, the observed loss ratio is randomly distributed around this mean with standard deviation 100%. Then the greatest accuracy credibility of the observed loss ratio will be 20%. If the observed manual loss ratio is 300% for a specific risk, this results in a credibility-weighted estimate of 140% for the true manual loss ratio for that risk, which suggests that a modification factor of 1.40 should be applied to its manual rate in the future.

A more *predictive* perspective focuses directly on the optimal estimate of future losses using past loss experience as explanatory information.

Example 2.2 Suppose that an analysis of past experience has shown that observed manual loss ratios for individual risks have a mean of 100%, an overall variance that is fairly constant over time, and a serial correlation of 15% from one year to the next. If the observed manual loss ratio for a risk in the prior policy period is 300%, this suggests for the next policy period an estimate of 130% for the true manual loss ratio for that risk and a modification factor of 1.30 should be applied to its manual rate.

Because loss processes tend to change over time, even if slowly in some cases, the predictive perspective can produce different estimates using the same data, usually involving lower credibility of experience than credibility from the static perspective.

NCCI's aggregate and classification ratemaking procedures for manual rates follow a mostly static perspective. It is worth noting that the large sample size in some manual rating classes, at least at a countrywide level, and the even larger data volume available for statewide aggregate calculations, lends itself to such a static perspective.

A static perspective can be difficult to apply to individual risks. There is only one observation for the total losses of a single risk in each individual policy period, albeit including varying numbers and amounts of individual claims. This makes it difficult from a static perspective, to determine the split out of the systematic and random components of variance that is necessary to determine the credibility which is vital to individual risk experience rating. However, by following a very predictive perspective, the NCCI Experience Rating Plan can combine data across risks, where each observation includes the combination of prior and subsequent experience relative to manual rate levels for a single risk.

Example 2.3 If we have the manual loss ratios for 1,000 equally sized individual risks in a single policy year and the sample mean is 100% with sample variance 1.25, we cannot determine the greatest accuracy credibility, or for that matter a predictive estimate, without extraneous information and/or assumptions.

Example 2.4 If we have the manual loss ratios for 500 equally sized risks over two sequential policy years, we can use the serial correlation, together with the sample means and sample variances in each policy year, to estimate a predictive credibility formula based on simple linear regression between the policy years.

Example 2.5 Alternatively in Example 2.4 from a static perspective we could try to split out the random component of the total variance, but this involves the hazards of (in a certain sense) using samples of size 2.

The NCCI ERP compares actual losses to manual basis expected losses, with many special adjustments to be described in subsequent sections, to produce a credibility-adjusted modification factor or *mod*. Three recent policy periods are used as the experience period. A manual basis rate for the risk, determined prospectively for an upcoming policy period, is multiplied by the modification factor. There is a one year gap between the three experience period "years" and the prospective policy effective year to allow for reports of actual losses to be made.

Time →

1st Experience Policy Period	2nd Experience Policy Period	3rd Experience Policy Period	Gap Policy Period	Prospective Policy Period

Example 2.6 The manual basis expected loss of $10,000 for a policy effective on 7/1/2015 might be multiplied by an experience modification factor of 0.90, resulting in a modified expected loss of $9,000, because actual losses on policies effective 7/1/2011, 7/1/2012, and 7/1/2013 were only about 60% of manual basis expected losses. Credibility procedures might determine that a 10% credit is the predictively meaningful information that can be derived, and the rest of the 40% lower than manual expected experience is random and non-predictive.

2.2 Lift and Equity Performance Objectives

The objective of the NCCI ERP is to improve the equity of rates for individual risks, where equity is defined in terms of how closely the relative rate for a risk matches its relative expected losses. Although the mod has an overall off-balance effect (see Section 6.1), the mod does not directly address the overall or absolute adequacy of rates, which is handled directly by manual rate levels. Operationally, equity of the mod itself is defined in a more limited way; specifically, equity is viewed in terms of equalizing modified expected loss ratios across the range of different mod values. However, this says nothing about how much improvement to overall rate equity the mod adds.

Example 2.7 A trivial and meaningless mod of 1.00 for every risk would be completely equitable in the sense that expected modified loss ratios would not vary by mod, but it would do nothing to improve the equity of rates.

To capture the magnitude of potential improvement in rate equity, another objective for the mod is high lift. Lift is the variation in manual loss ratios captured by the mod.

Example 2.8 If the prospective manual loss ratio is 150% for risks in the highest decile of mod values and 50% for those in the lowest decile, then lift is high.

Example 2.9 For Example 2.8, if the average mod is 1.05 for the highest decile and 0.95 for the lowest decile, then despite high lift, as shown by the decile manual loss ratio spread of 50% to 150%, the mod is not very equitable.

Example 2.10 For Example 2.8, if the average mod is 1.50 for the highest decile and 0.50 for the lowest decile, then the mod is likely very equitable and together with high lift, does much to improve rate equity.

Example 2.11 If the prospective manual loss ratio is 101% for risks in the highest decile of mod values and 99% for those in the lowest decile, then lift is so low that even if the mod is perfectly equitable it can only do a little to improve rate equity.

2.2.1 Safety Incentives as a Secondary Benefit

A secondary benefit of experience rating is that individual risks are incentivized to promote safety. Fewer claims and less severe claims generally lead to lower mods and hence lower future premiums. However, even if there were no safety incentive or no possibility for individual risks to improve safety, experience rating would still be justified if it made a significant improvement in rating equity.

Example 2.12 The mods in Example 2.10 would very much improve rate equity even if every risk had already implemented the highest possible standards of workplace safety.

2.3 Quantile Testing Predictive Performance

As mentioned earlier, there is only one observation of the actual manual loss ratio for a single risk in a single policy period. However, many observations are available for risks with the same mod value or mod values within a certain interval.

Example 2.13 In a recent policy year in a small state, there were about 3,600 risks with mods in the interval [0.84, 0.90).

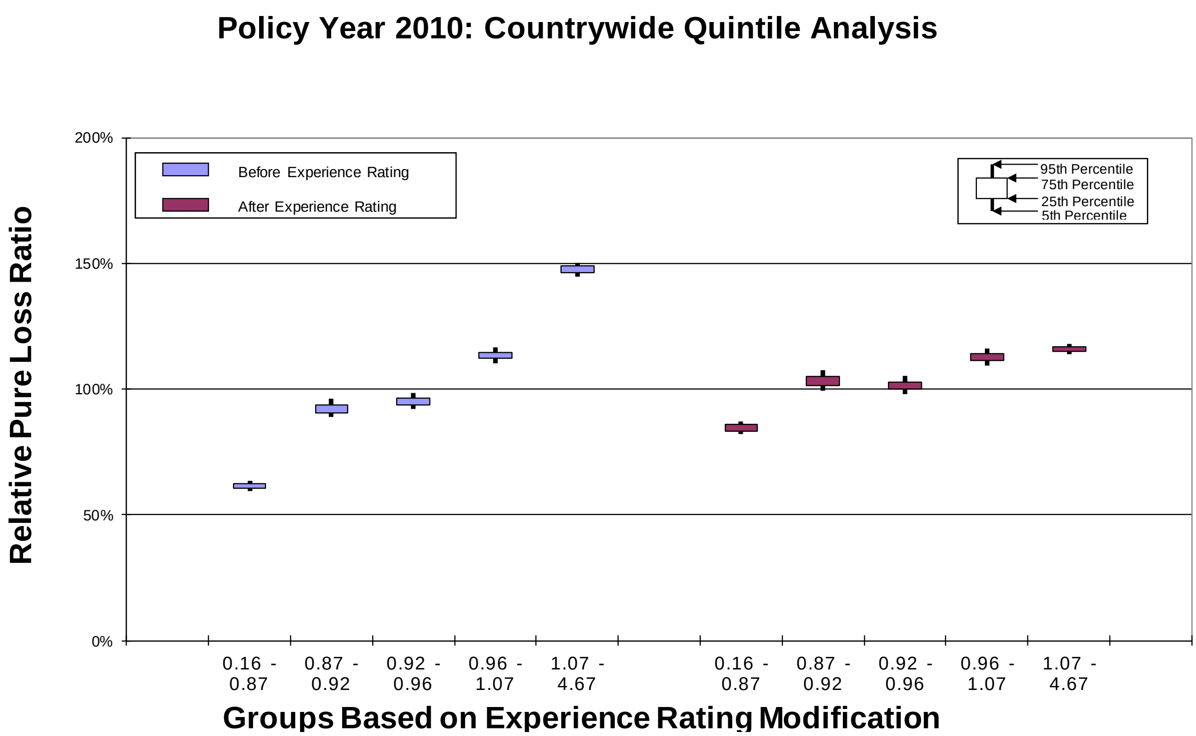
By splitting mod values into intervals and combining experience for risks into an overall manual loss ratio for each interval, most of the random effects for individual risks can be averaged away.

Example 2.14 In the 1934 paper by Paul Dorweiler (see Bibliography), tables are shown that combine data for risks based on premium size and mod ranges of width 0.10. The tables showed that modified loss ratios had higher equity than manual loss ratios. This result further improved when regression lines were fit to the loss ratios to filter out much of the remaining random effects.

To perform a similar analysis, NCCI focuses on a *quintile* test, which is specifically the five category version of the more general concept of a *quantile test*. A quantile test divides risks into intervals, each containing an equal number of risks and called *quantiles*, based on the value of the ERP modification factor. This test reveals both lift, shown by the quintile loss ratios before the mod is applied, and equity, shown by the quintile loss ratios after the mod is applied.

Example 2.15 The following candle chart displays the results of an actual quintile test of the NCCI ERP Countrywide for Policy Year 2010. The percentiles of the candles demonstrate bootstrap confidence intervals of the relative pure loss ratios. This test shows that the lift is very high, as can be seen by the steeply ascending candles on the left side of the chart (Before Experience Rating). It also shows that the application of the mod improves the equity, but the slight slope of the candles on the right side (After Experience Rating) suggests the ERP should be more sensitive. In fact, effective in 2013, NCCI implemented a significant increase in the split point (see Appendix A) which is expected to improve equity and result in the candles on the right side of a quintile test being much flatter.

Example 2.15



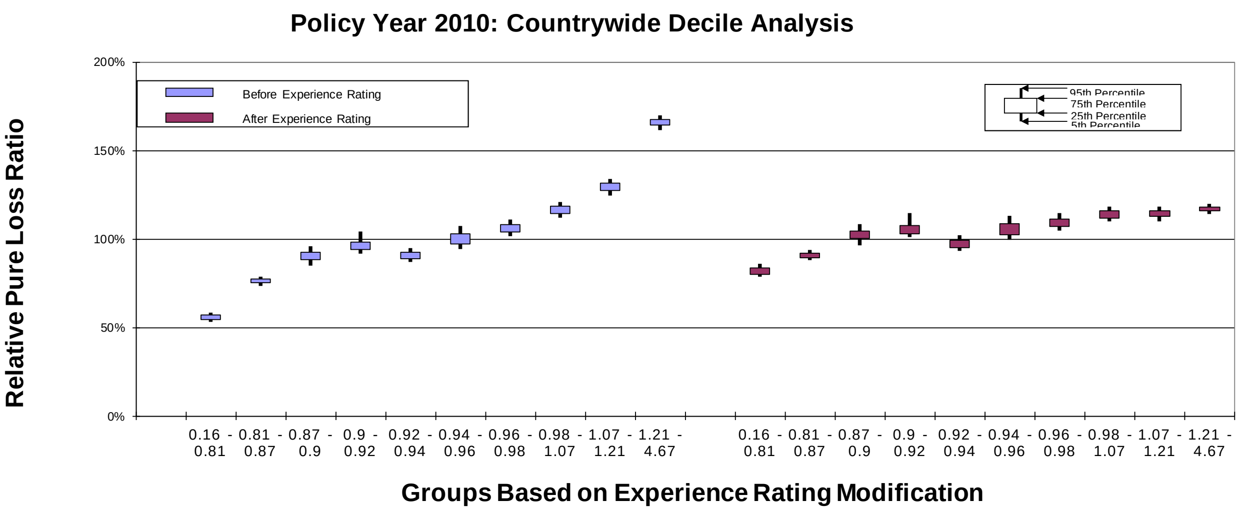
To put numerical scalar values on the concepts of lift and equity NCCI considers two statistics. The *old quintile test statistic* is B*/A*, where A* is the variance of the un-modified quintile loss ratios and B* is the variance of the modified quintile loss ratios. This traditional statistic measures equity and ideally should be as close to 0 as possible. The *new quintile test statistic* is sign(A-B)|A-B|0.5, where A and B are equivalent to A* and B*, but may be calculated to include some extra variance due to bootstrapping the test. This newer statistic measures the amount of the lift which is actually resulting in improved rate equity and ideally should be as large as possible.

Example 2.16 For the quintile test in Example 2.15, A = 0.0806, A* = 0.0803 and B = 0.0124, B* = 0.0120. The old quintile test statistic is 0.149 and the new quintile test statistic is 0.261.

NCCI uses the quintile test, with its relatively small number of five quantiles, specifically because large volumes of risk data are required to average away random effects and isolate both the lift and equity aspects of predictive performance. It turns out there is a very large penalty for increasing the number of quantiles in terms of the statistical clarity of a quantile test (see Evans and Dean in Bibliography). A decile test, with only twice as many quantiles, requires eight times as much data to provide the same level of statistical clarity as a quintile test.

Example 2.17 Here is a decile test on the same underlying data used in Example 2.15 The comparable old "decile" test statistic is 0.140 and the comparable new "decile" test statistic is 0.265 for this decile test. Although the patterns are still relatively clear, the vertical size of the candles, that is their bootstrap confidence intervals, has increased at the same time as the distances between the vertical positions of the candles has decreased. In many smaller categories than this countrywide example, such as individual states, a quintile test will produce acceptably clear results but a decile test will not.

Example 2.17



2.4 No Guarantee of Individual Risk Outcomes

Although a mod with good lift and equity improves the relative estimate of expected losses for an individual risk, there is no guarantee that the actual losses that occur will be anywhere close to this estimate. To the contrary, actual losses for an individual risk will generally be much larger or much smaller than modified expected losses.

Example 2.18 For the situation in Example 2.2, if the variance of the manual loss ratios is 1.00 then the standard deviation of the modified loss ratios will be $\sqrt{}$(1-0.152) = 0.99 Thus, actual losses in the subsequent policy year will likely differ by at least +50% from modified expected losses for 250 or more of the 500 risks.

Furthermore, if the modified expected loss were highly accurate at predicting actual losses on an individual risk basis, then credibility would tend to be close to 100%. In such a situation, the risk would almost certainly be self-insured for all losses, except possibly extreme tail events.

Example 2.19 Suppose for the situation in Examples 2.2 and 2.18 that the serial correlation, and hence credibility for this example, was 99% instead of 15%. The standard deviation of the modified loss ratios would be $\sqrt{(1\text{-}0.992)} = 0.14$ and the actual losses would still likely differ by at least +14% from modified expected losses for 150 or more of the 500 risks. To achieve a standard deviation of 5% would require the serial correlation/credibility to be 99.9%.

The value of experience rating can be seen through tests, such as quantile tests, that combine data across individual risks, but not for specific individual risks except in the case of extremely large, and effectively self-insurable, risks.

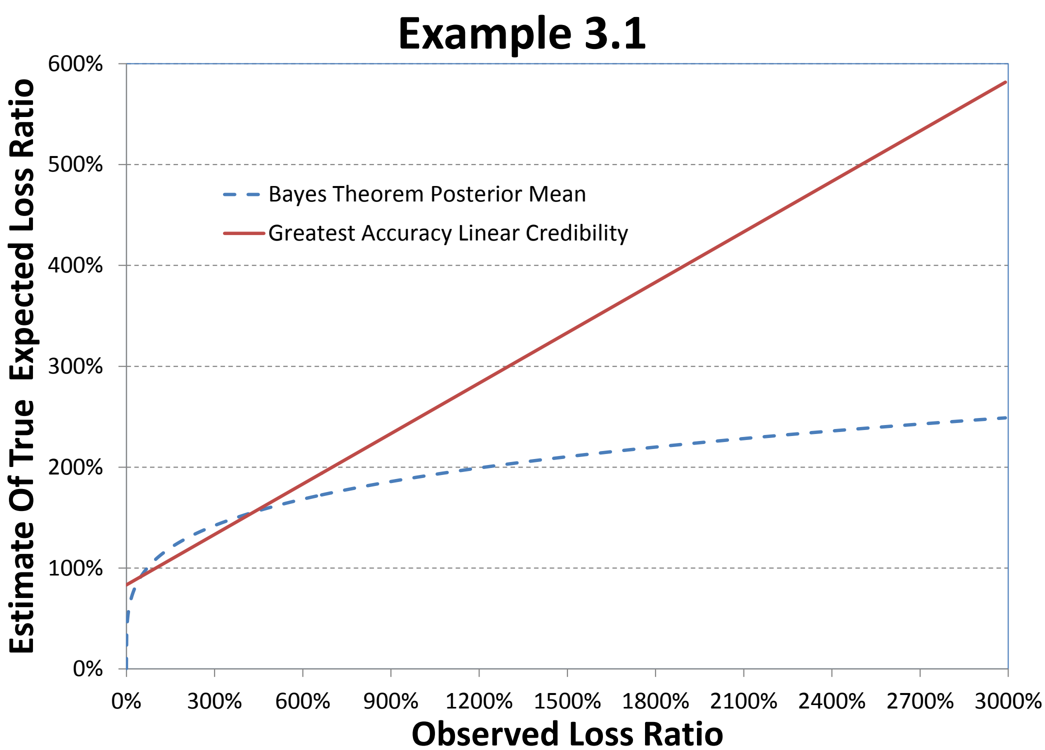
3. COMPLICATING PROPERTIES OF THE INDIVIDUAL RISK LOSS PROCESS

In this section we describe some complications in the individual risk loss process that, together with moving from a static perspective to a predictive perspective, as described in Section 2, require generalizing the basic credibility/statistical models, as will be described in Section 4, for the context of individual risk experience rating.

3.1 Contrast With Simplifying Assumptions of Basic Credibility and Statistical Models

Greatest accuracy credibility methods use simple linear (more precisely, affine) functions of observed losses to estimate true expected losses with minimal expected squared error (see Klugman et. al. in Bibliography). Similarly, simple linear regression models use least squares to fit parameters. Both of these models work very well for data processes that follow symmetric distributions with relatively thin tails, especially the Normal Distribution. However, the individual risk loss processes for workers compensation and most property/casualty insurance are very different.

Example 3.1 Suppose that prior to observing an actual loss ratio, the true mean loss ratio for a specific risk is lognormally distributed with mean 100% and standard deviation 50%. Also, suppose the observed loss ratio is lognormally distributed around the true mean with coefficient of variation 100%. The standard greatest accuracy linear credibility of the observed loss ratio as an estimate of the true loss ratio of the risk is Z = 16.7%. For very low or very high observed loss ratios, this produces very different estimates of the true mean compared to the Bayes Theorem posterior mean, which greatest accuracy linear credibility is intended to estimate.



3.2 Frequency and Severity Components of Total Loss

The total loss for a risk is the sum of a number of individual claim amounts.

Example 3.2 Suppose Risk A experienced one claim of $300,000 and Risk B experienced 10 claims each of $30,000. Both risks experienced $300,000 in total losses.

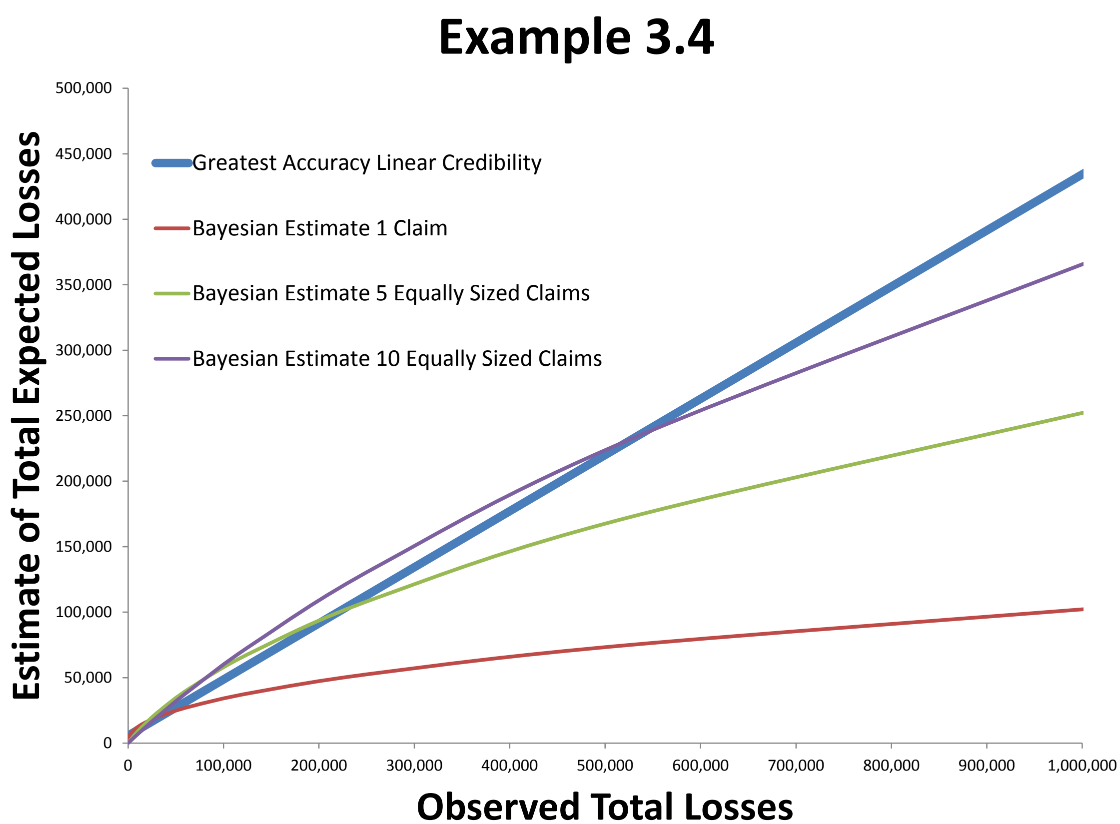
The underlying expected loss is the product of an expected number, or *frequency*, of claims, and the expected amount, or *severity*, of an individual claim. A certain fixed amount of total loss for a risk might have different implications for frequency and severity parameter estimates depending on the number and amounts of individual claims occurring.

Example 3.3 Suppose both risks in Example 3.2 had a prior manual basis expected loss of $10,000 based on a frequency of 1.0 and severity of $10,000. Risk A's experience suggests a higher severity, but Risk B's experience also suggests a higher frequency. As observed frequency tends to be more credible than an observed severity, Risk B likely has a higher true expected total loss than Risk A.

Example 3.4 Suppose that:

- Prior to observation, a risk's true claim frequency is exponentially distributed with mean 1.0.

- The true severity of its claims is independent of the true frequency and exponentially distributed with mean $10,000.
- The number of claims is conditionally Poisson distributed.
- The amount of each of the claims is conditionally independent of the number of claims and the amount of the other claims, and conditionally follows an exponential distribution.



Then the credibility of the total observed loss is Z = 42.9%. Not only does the basic credibility estimate fail to capture the curvature in the Bayesian estimate as a function of total loss dollars that are observed, but it makes no distinction regarding the frequency and severity composition of the loss dollars.

Example 3.5 Suppose the risks in Example 3.2 both had the underlying loss process in Example 3.4. The Bayesian estimates for the two different outcomes are very different from each other and from the greatest accuracy linear credibility estimate.

Risk	Observed Number of Claims	Observed Amount of Each Claim	Greatest Accuracy Linear Credibility Estimate of True Expected Losses	Bayesian Estimate of True Frequency	Bayesian Estimate of True Severity	Bayesian Estimate of True Expected Losses
A	1	$300,000	$134,286	1.0	$57,220	$57,220
B	10	$30,000	$134,286	5.5	$27,404	$150,722

Furthermore, there may be statistical dependencies between the numbers of claims and the amounts of claims that occur. There may also be dependencies between the individual amounts.

Example 3.6 If a risk experiences only one claim, the expected severity might be $9,000 but this might increase to $15,000 for each claim if two claims occur.

Example 3.7 If two claims occur, there might be a 25% correlation between the severities. If this were the situation in Example 3.6, then if one claim was $55,000 the expected value for the second claim would likely be closer to $25,000 than $15,000.

3.3 Skewness of Frequency and Severity Processes

Typically, the frequency process and the severity process are each asymmetric. Frequency processes are very positively skewed where frequency is low, but much less so when frequency is high.

Example 3.8 Suppose the frequency process is conditionally Poisson distributed with uncertainty in the frequency being Gamma distributed with coefficient of variation 10%. If the overall frequency is 1.0, then the probability of 0 claims is 37.0% versus a probability of 26.4% for 2 or more claims. If the overall frequency is 100.0, then the probability of 99 or fewer claims is 50.0% versus a probability of 47.2% for 101 or more claims.

Severity processes tend to be extremely skewed with extremely heavy tails.

Example 3.9 An overly simplistic, but not totally unrealistic, model for the overall severity process for workers compensation individual claim amounts, combining all levels of injury ranging from medical only to permanent total, would be a lognormal distribution with mean $20,000 and coefficient of variation of 700%. This model implies $2,282 is the median, $20,000 is the 83.9th percentile, 3.6% of claims exceed $100,000, and 1 in 665 claims exceed $1,000,000.

3.4 Severity Differences Between States and Over Time

The various rules for benefit levels, average wages, medical costs, and other conditions can vary considerably between states, leading to significant differences in average severity.

Example 3.10 In one recent case, the average indemnity amount on lost time claims was about three times higher in one state than in a similar neighboring state sharing a very long border and a significant concentration of population straddling the border. (See any yearly edition of NCCI's Annual Statistical Bulletin for specific examples along these lines.)

Severity also experiences very significant trends over time.

Example 3.11 The overall severity of claims more than tripled from the mid 1990's to circa 2010. (Compare average claim severity statistics from different editions of NCCI's Annual Statistical Bulletin.)

4. GENERALIZING CREDIBILITY METHODS FOR THE CONTEXT OF EXPERIENCE RATING

Having discussed the predictive perspective in Section 2 and complications in the individual risk loss process in Section 3, in this Section we show how basic static credibility is generalized to address both of these challenges. This generalized credibility is the essence of the framework that has been in place for the NCCI ERP over the last century. See Appendix A for some history.

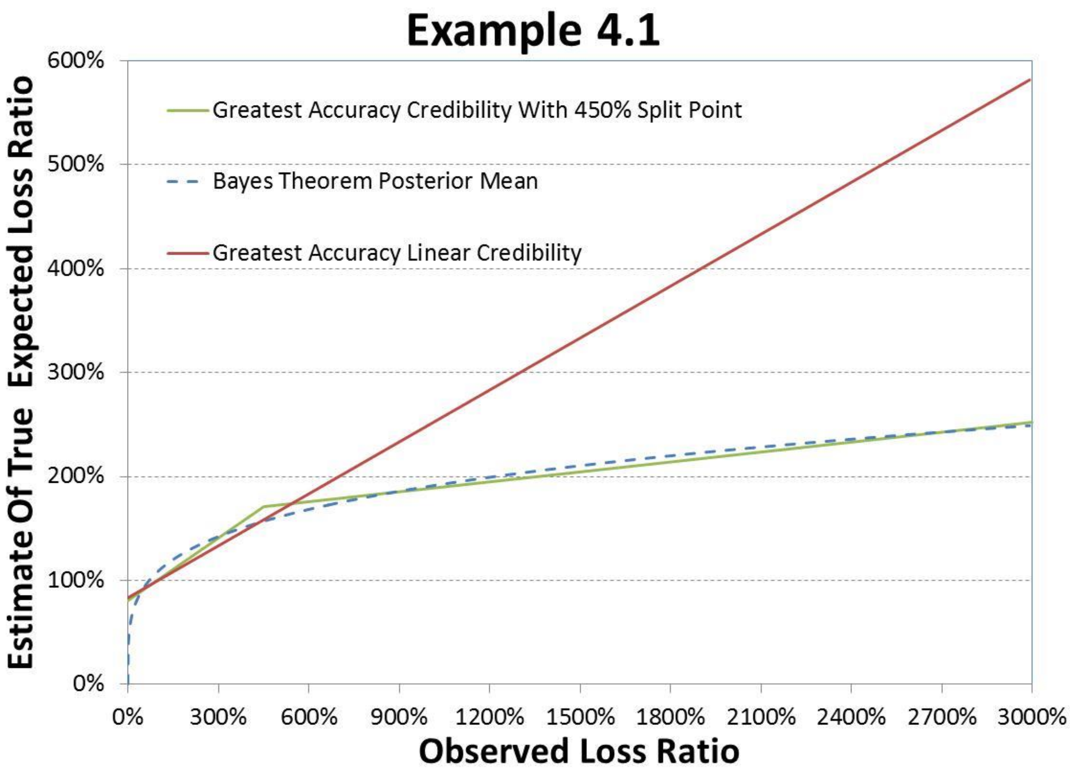
4.1 Splitting Losses Per Claim: Primary, Excess, and Over a Loss Limit

To address the highly skewed nature of losses, losses are split into separate components with linear credibility applied to each component. The combined credibility-weighted estimate for the overall true expected loss ratio can then be expressed as:

$$Z_p\, A_p + \left(1 - Z_p\right)E_p + Z_e\, A_e + \left(1 - Z_e\right)E_e$$

where the subscripts "p" and "e" are used to denote the primary and excess components of the actual observed loss ratios (Ap and Ae) and the expected loss ratios (Ep and Ee)..

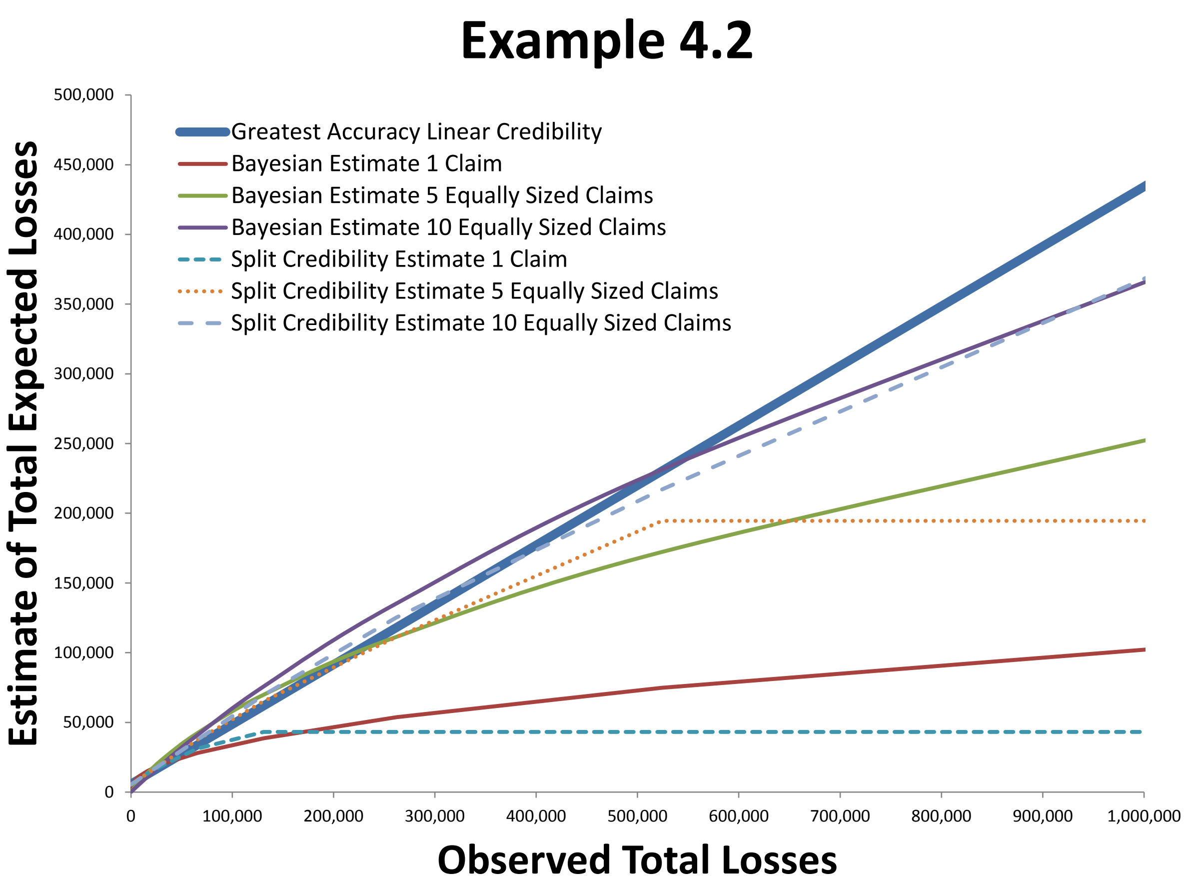
Example 4.1 In Example 3.1 we can split the observed loss ratio into a primary part under 450% and an excess part above 450%. The greatest accuracy credibilities, for each component, calculated separately with no adjustments for statistical dependencies, are Zp = 20% and Ze = 3.2%. The means prior to observing the loss ratio of the expected partial loss ratios of the layers are Ep =0.954 and Ee = 0.046.



Example 4.1 shows the power of using linear credibility in a piecewise fashion to accommodate a highly skewed loss process and much more closely approximate the Bayesian posterior mean.

To also accommodate differences arising from frequency and severity, losses can be split on a per claim basis rather than on an aggregate basis.

Example 4.2 We can apply a per claim split to Example 3.4. Suppose losses are split per claim into a primary layer of losses below $20,000 and excess losses above $20,000. Also, individual losses are capped at $100,000, which reduces losses in the excess layer to a maximum of $80,000 per claim. Greatest accuracy credibility gives Zp = 47.9% and Ze = 34.2%. Correspondingly Ep = $6,908 and Ee = $2,857. Since the loss limit reduces losses overall by -2.35%, the raw split credibility estimate should be increased by +2.41%, as the estimate is still intended to cover unlimited losses. Compared to credibility without a split, the split credibility estimates are dramatically better for 1 single claim and 5 claims, and slightly better for 10 claims.



Example 4.3 Applying the $100k split credibility from Example 4.2 to Example 3.5 produces estimates much closer to the Bayesian estimates for Risk A and slightly closer for Risk B.

Risk	Observed Number of Claims	Observed Amount of Each Claim	Greatest Accuracy Linear Credibility Estimate of True Expected Losses	$20k Split $100k Limit Credibility Estimate of True Expected Losses	Bayesian Estimate of True Frequency	Bayesian Estimate of True Severity	Bayesian Estimate of True Expected Losses
A	1	$300,000	$134,286	$43,207	1.0	$57,220	$57,220
B	10	$30,000	$134,286	$138,430	5.5	$27,404	$150,722

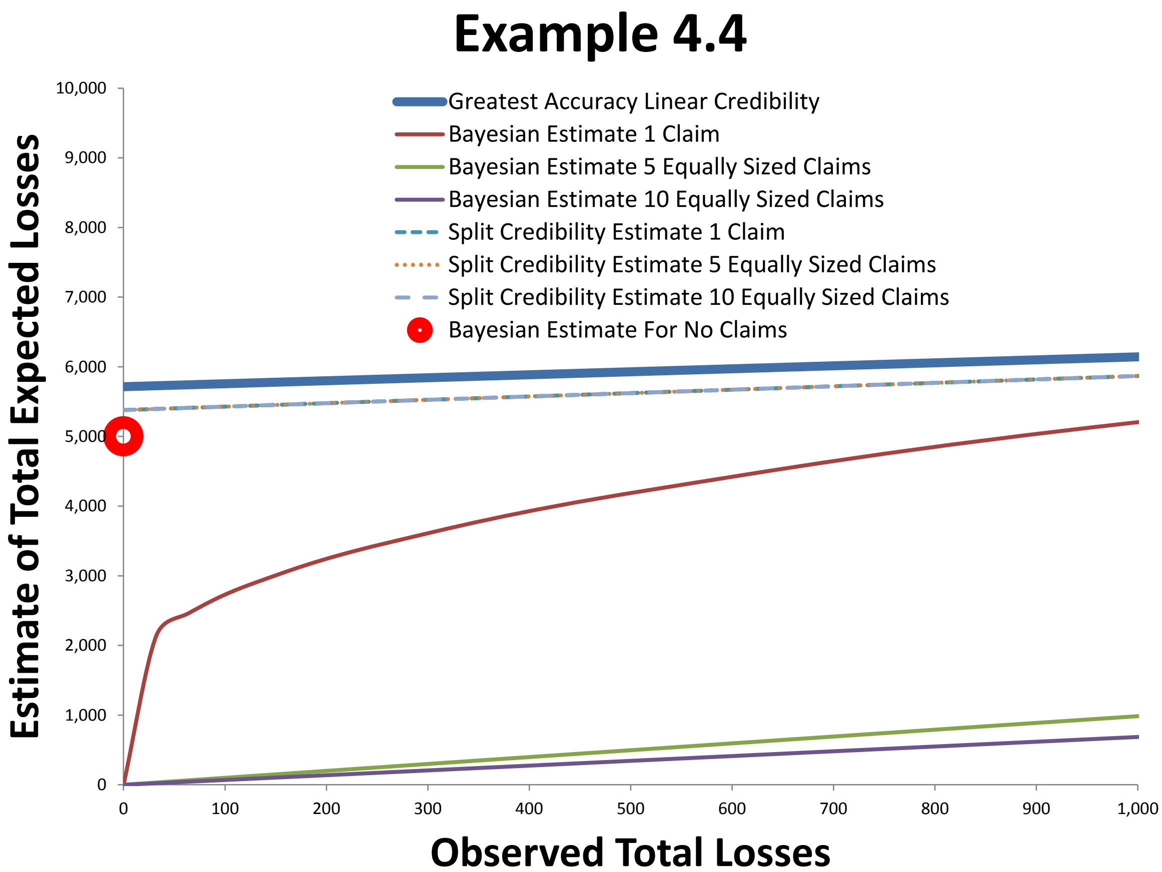
A natural question is why to not simply use a Bayesian estimate instead of a credibility approximation, split or otherwise. There are several practical reasons, including:

- Bayesian estimation requires specifying prior and likelihood distributions, but credibility only requires a much simpler specification of a few variance parameters that are not distribution specific.
- Bayesian estimation involves integral calculations and/or simulations that are usually computationally complex, but credibility only requires simple arithmetic calculations.

Another question is why to not apply credibility directly to frequency and severity rather than layers of losses split per claim. This would be a more viable alternative. However, here again the severity piece on its own is still highly skewed and would likely require some sort of non-linear adjustment. Additionally, a separate frequency estimate would be very sensitive to the claim count, which is driven by small claims and particularly medical only claims. More specifically, a particularly odd feature is that the Bayesian estimate would tend to reward a few very small losses over no losses at all.

Example 4.4 In Example 4.2, the Bayesian estimates have a discontinuity near zero observed total losses, between zero observed claims and one or more observed claims. The credibility estimates do not suffer from this problem, but gradually drop to reasonable estimates, as observed losses approach zero, independent of the number of observed claims.

Observed Number of Claims	Observed Amount of Each Claim	Greatest Accuracy Linear Credibility Estimate of True Expected Losses	20k Split 100k Limit Credibility Estimate of True Expected Losses	Bayesian Estimate of True Frequency	Bayesian Estimate of True Severity	Bayesian Estimate of True Expected Losses
0	NA	$5,714	$5,379	0.5	$10,000	$5,000
1	$100	$5,757	$5,428	1.0	$2,725	$2,725
10	$100	$6,143	$5.869	5.5	$125	$686



Also, as will be shown subsequently, the split credibility approach leads to certain parameters that can be empirically fit to optimize predictive performance. This empirical fitting implicitly leaves some slack to, at least partially, account for potential frequency and severity process correlations and other complications of the loss process.

4.2 The Current Parameterization of Credibility

The basic form of the ERP modification formula is:

$$1 + Z_p \frac{A_p - E_p}{E} + Z_e \frac{A_e - E_e}{E}$$

Ap = actual primary ratable loss from the experience period

Ae = actual excess ratable loss from the experience period

Ep = expected primary ratable loss from the experience period

Ee = expected excess ratable loss from experience period

E = Ep + Ee = expected total ratable loss from experience period

Zp = primary credibility

Ze = excess credibility

The split between primary and excess is done on a per claim basis and there is a limit on individual losses, similar to Example 4.2. See Sections 5 and 6 for details on the individual loss limit and various other adjustments. The resulting experience mod is applied to manual basis premium.

Example 4.5 Suppose the split point is \$15,000 and the individual loss limit is \$300,000. One claim of \$50,000 and another claim of \$500,000 in the experience period would lead to Ap = \$30,000 and Ae = \$320,000.

Example 4.6 If manual basis pure premium is \$20,000 and the mod is 1.15, then the modified pure premium is 23,000.

Example 4.7 Here is a hypothetical example of how the basic mod formula would react to some different actual loss outcomes for the same risk.

Zp	50%
Ze	4%
Ep	$10,000
Ee	$15,000

	Actual Loss Outcomes		
Ap	$0	$30,000	$15,000
Ae	$0	$0	$90,000
Mod	0.78	1.38	1.22

The current methodology for parameterizing Zp and Ze was implemented in 1991 (see Gillam, William R.., "Parametrizing..." in the Bibliography). We will demonstrate the motivation, or more properly inspiration, for this parameterization in terms of static credibility theory. The specific values of the parameterization are then empirically fit and validated from the predictive perspective performance goals for the ERP.

Basic credibility models like those extensively used in Sections 3 and 4 determine Z as:

$$Z = \frac{VHM}{VHM + EPV}$$

where,

VHM = variance of the hypothetical means

EPV = expected process variance

When the quantity being estimated is a ratio to the quantity of exposure (or measure of data volume), such as a manual loss ratio of losses to manual expected losses for an individual risk, standard assumptions are that VHM is constant and EPV is inversely proportional to the exposure.

$$VHM = b \qquad\qquad EPV = \frac{a}{E}$$

E = quantity of exposure (such as manual basis expected losses)

a, b = positive constants

This leads to the familiar formula for Z in terms of data volume:

$$Z = \frac{E}{E+K} \qquad K = \frac{EPV(for\ E=1)}{VHM} = \frac{a}{b}$$

More general assumptions prevent the variability around the true expected value from asymptotically going to 0 for large data volume (large risks in our context) and also recognizes that the heterogeneity of true expected values is likely greater for small data volume (small risks in our context).

$$VHM = e + \frac{f}{E} \qquad\qquad EPV = c + \frac{d}{E}$$

c, d, e, f = positive constants

This leads to a more general form for K:

$$K = E\left(\frac{cE + d}{eE + f}\right)$$

K can be rewritten as:

$$K = E\left(\frac{CE + D}{E + F}\right)$$

where C = c/e, D =d/e, and F =f/e are an alternative set of positive constants. This shows that there are only three degrees of freedom in determining K from c, d, e, and f.

Since credibility is more tied to the expected number of claims than expected losses, another generalization is to substitute a measure of expected claims for expected losses in the credibility formulae:

$$K = \frac{E}{G}\left(\frac{C(E/G)+D}{(E/G)+F}\right) = \frac{E}{G}\left(\frac{CE+GD}{E+GF}\right) \qquad Z = \frac{(E/G)}{(E/G)+K} = \frac{E}{E+K_*} \quad K_* = GK = E\left(\frac{CE+GD}{E+GF}\right)$$

G = an index of severity that may vary over time and between states

Example 4.8 Z is the same when E=$25,000 and G =5 as when E=$50,000 and G=10. In each case the implied expected number of claims is E/G = 5,000.

Example 4.9 The introduction of the severity index G could have been done earlier, with the same result, by specifying the VHM and EPV relationships as:

$$VHM = e + \frac{f}{E/G} \qquad EPV = c + \frac{d}{E/G}$$

Now if we use the credibility formulae just developed separately for primary and excess losses, but retain the same E/G measure of data volume in each case, we have the following parameterization of primary and excess credibility values:

$$Z_p = \frac{E}{E+K_{p*}} \qquad K_{p*} = E\left(\frac{C_pE+GD_p}{E+GF_p}\right)$$

$$Z_e = \frac{E}{E+K_{e*}} \qquad K_{e*} = E\left(\frac{C_eE+GD_e}{E+GF_e}\right)$$

Cp, Dp, Fp, Ce, De, Fe = positive constants

Example 4.10 The most recent NCCI update from the late 1990s of the constant credibility parameter values is shown below. There are also minimum values of 2500G and 60,000G imposed on Kp* and Ke*, respectively.

Cp	0.10	Ce	0.375
Dp	$2,570	De	$150,000
Fp	$700	Fe	$5,100

The G value can be determined easily from claim statistics and E is determined from classification payroll data for an individual risk. The six constants were determined to maximize the predictive performance of the mod formula, as will be discussed in Section 4.3

4.2.1 W and B Representation Versus Zp and Ze Representation

A simple no split modification formula using a basic credibility formula takes the form:

$$1 + Z\left(\frac{A-E}{E}\right) = 1 + \frac{E}{E+K}\left(\frac{A-E}{E}\right) = \frac{A+K}{E+K}$$

It has been long recognized that K acts as a kind of "ballast", stabilizing the mod calculation across different values of E, or effectively different risk sizes. For split credibility, Ze is generally less than Zp and it makes sense to think of Ze = W Zp, where W, on the interval [0%, 100%], is a kind of weight given to excess losses. Using this framework, and renaming Kp* as B so that Zp = E / (E + B), we can restate the split credibility mod formula as:

$$1 + Z_p\left(\frac{A_p - E_p}{E}\right) + Z_e\left(\frac{A_e - E_e}{E}\right) = \frac{A_p + WA_e + (1-W)E_e + B}{E+B}$$

Example 4.11 Consider an individual risk where E = \$25,000, Zp = 50% and Ze = 4%. This is equivalent to B = \$25,000 and W =0.08.

As a result, credibility values can be expressed entirely in terms of this alternative W and B representation, rather than the Zp and Ze representation, without any substantial difference. The W and B representation has been more common for the actual publication of tables and calculation of mod factors. When the ERP began about 100 years ago, in an era before inexpensive computing power, this form had the advantage of slightly reducing the complexity of the arithmetic calculations.

4.2.2 Primary/Excess Correlation

The ERP does not make any explicit adjustment for correlation between the primary and excess loss layers, which is generally high.

Example 4.12 The correlation between the primary and excess layers in Example 4.1 is 32.7%

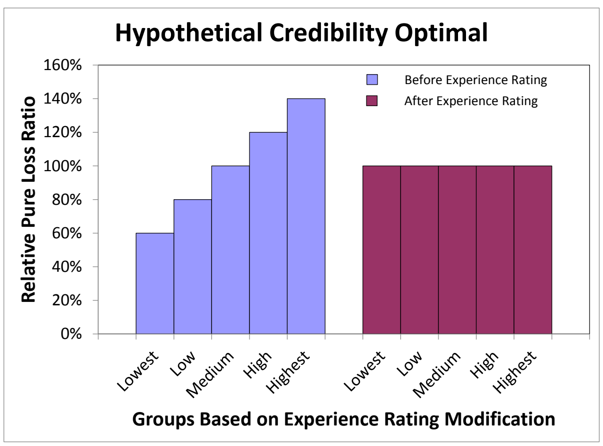
Example 4.13 The correlation between the primary and excess layers in Example 4.2 is 64.2%

Correlation between explanatory variables is generally undesirable in statistical models, particularly linear models. However, as we saw with Examples 4.1 and 4.2 for hypothetical static estimation scenarios, the sum of separate layer credibility estimates, without any correlation adjustment, can still do a good job of estimating total expected losses. Furthermore, as will be discussed in Section 4.3, the credibility parameters are empirically fit to optimize the predictive value for total expected losses, as measured by quintile tests. This predictive value has also been proven through many subsequent quintile tests over the years.

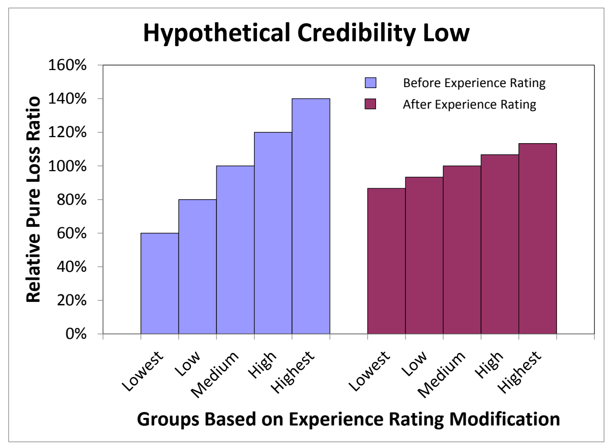
4.3 Predictive Fitting and Validation of Credibility Parameters With the Quintile Test

The parameterization set up in Section 4.2 was based on applying a static model for greatest accuracy credibility to split primary and excess layers of loss. The pure premium underlying actual modified rates is used to estimate all expected losses (save for some exceptions discussed in Section 6.5). However, the six constants from this model have not yet been estimated. So, the estimation step is where the predictive perspective objectives for estimating future total expected losses come in.

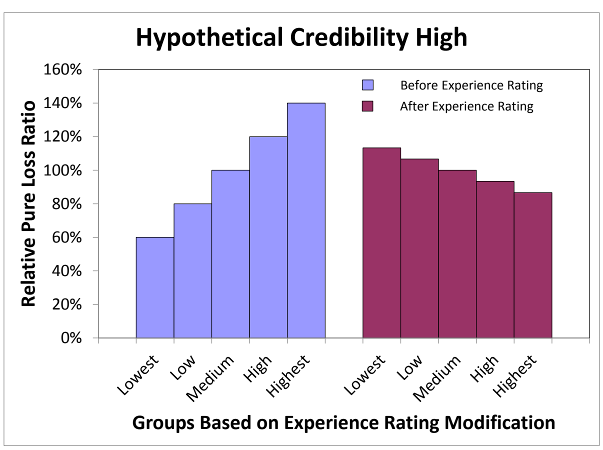
Example 4.14 Hypothetical example of desirable credibility.



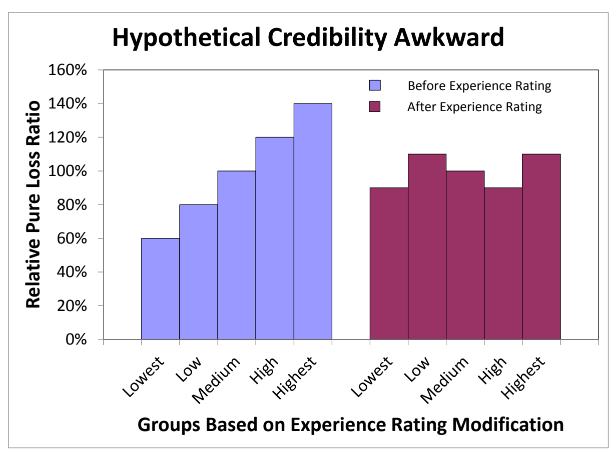
Example 4.15 Credibility should be increased likely by some combination of lower Cp, Dp, Ce, De, and/or higher Fp, Fe.



Example 4.16 Credibility should be decreased likely by some combination of higher Cp, Dp, Ce, De, and/or lower Fp, Fe.



Example 4.17 Credibility should be optimized by some combination of changes in Cp, Dp, Ce, De, Fp, and/or Fe.



Although the constants Cp, Dp, Fp, Ce, De, and Fe were derived from assumptions about parameter and process variances, no direct estimates of such variances are used to determine values for these constants. Instead, the actual values are selected to flatten the quintile test described in Section 2.2, specifically by bringing the old quintile test statistic close to 0. Effective credibility decreases as Cp, Dp, Ce, and De increase. Effective credibility increases as Fp and Fe increase. These constants were determined in the 1990s to simultaneously flatten quintile tests for different risk size ranges. The determination was made through a general hands-on trial and error iteration process, not any special mathematical search algorithm. Although the old quintile test statistic only measures mod equity, lift has always been high for ERP plans tested. By optimizing predictive performance, these six constant values are, to some extent, implicitly accommodating various complications in the loss process, such as possible correlations between frequency and severity.

After the constant values were determined, the high predictive performance of the ERP has continued to be validated by subsequent quintile tests, with credibility automatically adjusted across time and different states by the severity index G.

4.4 Indexations By Severity

Part of the parameterization of the credibility values was the inclusion of the severity index G. The individual loss limit is also indexed with each new manual rate filing. The split point will now also be subject to routine indexation. Quintile testing has demonstrated that the values for Cp, Dp, Fp, Ce, De, and Fe determined in the 1990s continue to work well due to the indexation updates built into the ERP.

Example 4.18 Suppose the split point is $5,000 and the individual loss limit is $100,000. If severity triples, then a split point of $15,000 and an individual loss limit of $300,000, along with Zp and Ze corresponding to a G-value three times higher should give close to the same performance as before severity tripled. As a first approximation, tripling severity is equivalent to exchanging all dollar amounts for a synthetic currency that trades at three units for every one dollar.

Several different types of severity indexes are utilized in the ERP. Aside from special rounding rules, these are all multiples of the State Average Cost per Claim (SACC). These indexes include:

G-value (G) -> 0.001 SACC

State Accident Limit (SAL) -> 25 SACC

State Reference Point (SRP) -> 250 SACC

The SAL is used to set the individual loss limit by state and over time. The SRP is used only as an intermediate step in calculating the SAL.

Example 4.19 G-value = 10.0 is equivalent to SAL = \$250,000 is equivalent to SRP = \$2,500,000 is equivalent to SACC = \$10,000.

4.5 The Minimum Size Eligibility Threshold

For a specific risk to be eligible for the ERP, its average premium for the three most recent years must exceed a standard amount that varies by state. The standards were established in the 1980s with few changes since, based on an estimate of the average premium for an employer with 10 employees. Current values range from a low of \$2,250 to a high of \$5,500.

Applying the ERP to a risk involves some effort and expense, both in terms of service performed at NCCI and by insurers. For very small risks with very low credibility, the potential improvement in rate equity, or at a more basic level the maximum potential dollar change in premium, is so small that applying the ERP is not practical.

Example 4.20 Consider the following hypothetical illustration. Suppose a risk has manual premium of \$1,000 for the prospective policy period. Ep = \$600, Ee = \$900, Zp = 3.85% and Ze = 0.33% and the SACC = \$10,000. For this risk, the minimum mod would be 0.98 and the mod cap (see Section 6.2) would be 1.16. Typically, this small risk will be claim free, and would receive a \$20 credit due to the minimum mod; more rarely, actual claims could produce up to a \$160 debit due to the maximum mod. Even a modest expense associated with the calculation and application of the mod would offset the small premium impact of applying experience rating for this risk.

5. EXPERIENCE PERIOD ACTUAL LOSS DATA AND EXPECTED LOSS CALCULATION

Having now set up the basic ERP mod formula with its split credibility formula in Section 4, this section will describe how experience period actual and expected losses are determined. The NCCI Workers Compensation Unit Statistical Plan, which is used to collect data for manual ratemaking, is also used in the ERP to determine actual reported losses and corresponding manual basis expected losses for the experience period.

5.1 The NCCI Unit Statistical Plan

NCCI collects audited exposure, premium, and loss information by policy in states for use in ratemaking, experience rating, actuarial analysis, and other NCCI products and services. In addition to its use in manual ratemaking, this data is also used by NCCI to calculate mods. Individual risk ID numbers are assigned to all employers, including interstate employers having payroll in more than one state. Employers may change insurance carrier, name, location, or in other respects, but in general as long as the fundamental nature of the business and its ultimate owner does not change, the risk ID is intended to be maintained. Unit statistical plan data is first reported as of 18 months past the policy effective date (called 1st report), and subsequently at 12 month intervals (called 2nd, 3rd, etc. reports) up to 126 months (10th report). Among other information, key data elements include payroll by class code and individual claim incurred loss amounts.

Example 5.1 A premium unit at 1st report from policy year 2007 might contain payroll of $500,000 for class XXXX and $1,500,000 payroll for class YYYY.

A corresponding loss unit at 3rd report might also show a permanent partial claim with $300,000 and several medical only claims totaling $50,000.

5.2 ELRs

Payroll by class from the unit reports for the three experience policy periods is multiplied by Expected Loss Rates (ELRs), which are filed together with regular manual rates by class, both specified as rates to $100s of payroll, to determine the expected experience ratable loss.

$$E = \text{ELR} \times (\text{Payroll}/100)$$

Ratable losses for the experience period differ from total prospective ultimate losses, which are included in manual rates, in several ways, such as:

- Ratable losses include losses incurred over the three experience period years, rather than the single prospective year policy.
- Losses are valued as of 1st, 2nd, and 3rd reports, respectively, for the experience years used, and are not developed to ultimate.
- Losses are adjusted by benefit levels and trends to correspond to the experience period rather than prospective policy period.
- Individual claims are limited by the State Accident Limit (SAL).
- 70% of the loss amount on medical only claims is excluded.
- Catastrophes and certain other special non-ratable losses are excluded.

Example 5.3 This table demonstrates the difference in practice between ELRs and manual pure loss costs for a hypothetical Risk A in State X. (Note: in some states, actual manual loss costs also include loss adjustment expenses)

Class	3 Year Experience Period Payroll	ELR (per $100 payroll)	Experience Period Expected Ratable Loss	Prospective Policy Period Expected Payroll	Prospective Pure Loss Cost (per $100 payroll)	Prospective Policy Period Expected Pure Loss
XXXX	$1,000,000	1.10	$11,000	$500,000	1.50	$7,500
YYYY	$1,500,000	1.80	$27,000	$600,000	3.00	$18,000
Total	$2,500,000		$38,000	$1,100,000		$25,500

Note, even though the prospective expected manual basis losses often tends to be of the same broad order of numerical value as the experience period expected ratable losses, the nature of these two expected loss numbers is very different.

The underlying ELR factors for a state are determined by Hazard Group, separately for indemnity and medical losses. Hazard Groups (HGs) are a partition of class codes into 7 groups (A through G) based on ascending claim severity. By calculating ELR factors separately by HG, differences in the impact of the state accident limit, which is much greater for the higher severity HGs, can be better accounted for. Calculating ELR factors separately for indemnity and medical losses allows a better accounting of differences with regard to loss development. The ELR for a class is determined by multiplying the ELR factors for the Hazard Group of the class by the indemnity and medical components of the pure loss cost.

Example 5.4 This table demonstrates the calculation of the hypothetical ELRs in Example 5.3 in terms of ELR factors.

Class	Hazard Group	Hazard Group Indemnity ELR Factor	Prospective Indemnity Pure Loss Cost (per $100 payroll)	Indemnity ELR (per $100 payroll)	Hazard Group Medical ELR Factor	Prospective Medical Pure Loss Cost (per $100 payroll)	Medical ELR (per $100 payroll)
XXXX	M	0.763	0.80	0.61	0.700	0.70	0.49
YYYY	N	0.650	1.20	0.78	0.567	1.80	1.02

We will not delve further into the underlying nuts and bolts of the ELR calculation except to highlight one particularly special part of this calculation, the ELAFs.

5.1.1 ELAFs

Excess Loss Adjustment Factors (ELAFs) are used in the calculation of ELRs to remove expected losses in the layer between the state accident limit and the $500k claim limit used in class ratemaking. Although SALs are currently still smaller than this $500k limit, the calculation is designed to add more expected loss, even if the SALs at some point exceed $500k.

The ELAFs are based on a totally separate excess ratio curve model rather than the excess ratio curves underlying NCCI's more familiar Excess Loss Factors (ELFs). ELAFs are designed for undeveloped losses about as mature as the actual losses in the experience period, whereas ELFs are designed for losses developed and stochastically dispersed to ultimate. The current generation of ELAF excess ratio curves consists of the arithmetic inverses of a set of simple cubic polynomials, one for each injury type.

Example 5.5 The ELAF excess ratio curve for Permanent Total claims is

$$E(r) = (0.003148\,r^3 + 0.054149\,r^2 + 1.299625\,r + 1)^{-1}$$

where E(r) is fraction of expected loss excess of r, which is the *entry ratio*, or a specific loss amount as a ratio to average severity. So, if PT claims in State X average $1,000,000 for Hazard Group Y and the SAL is $250,000 then a factor of (1-E(0.25))/(1-E(0.50)) = 0.619786 would be applied to expected undeveloped PT losses limited to $500k as derived from class ratemaking data to account for the $250k limit in the ERP.

The actual ELAFs are determined by state and by HG for all injury types combined through scaling and weighting individual injury type excess ratio curves, in a process similar to that used for ELFs.

5.3 D-Ratios

Expected ratable losses, as determined by ELRs times payroll, must be split into primary and excess layers. *D-ratios* are estimates of the fraction of total ratable losses that are primary, and are also filed alongside manual rates and ELRs by class.

$$Ep = \text{D-ratio} \times E \quad Ee = E - Ep$$

Example 5.6 If the D-ratios for class codes XXXX and YYYY are 0.40 and 0.30, respectively, then for Risk A from Example 5.3, primary and excess experience period expected ratable losses are:

Class	Experience Period Expected Ratable Loss	D-ratio	Experience Period Expected Ratable Primary Loss	Experience Period Expected Ratable Excess Loss
XXXX	$11,000	0.40	$4,400	$6,600
YYYY	$27,000	0.30	$8,100	$18,900
Total	$38,000		$12,500	$25,500

Similar to the partial ELR factors, the underlying partial D-ratios for a state are determined by Hazard Group separately for indemnity and medical losses, and then weighted together in proportion to the indemnity and medical components of the pure loss cost for each class.

Example 5.7 This table demonstrates the calculation of the hypothetical D-ratios in Example 5.6 in terms of D-ratio factors.

Class	Hazard Group	Hazard Group Indemnity D-ratio Factor	Hazard Group Medical D-ratio Factor	Indemnity Fraction of Prospective Pure Loss Cost (per $100 payroll)	Medical Fraction of Prospective Pure Loss Cost (per $100 payroll)	D-ratio
XXXX	M	0.450	0.350	50.0%	50.0%	0.40
YYYY	N	0.360	0.260	40.0%	60.0%	0.30

6. ADJUSTMENTS FOR OTHER CONSIDERATIONS

In Section 4 we arrived at the basic ERP mod formula and its credibility. Section 5 described the data collection and basic calculations behind determining experience period actual and expected losses for the mod formula. This section covers various other special adjustments.

6.1 Off-balance

The ratio of total modified pure premium in a state to total manual pure premium is called the *off-balance*. This is equivalent to the pure premium weighted average mod including non-rated risks, which have an implicit mod of 1.00. The off-balance tends to be less than 1.00, primarily because large risks with high credibility tend to have lower actual to manual basis expected experience than small low credibility risks which tend to have higher actual to manual basis experience. NCCI adjusts its manual rates overall to target statewide adequacy of pure premium on a modified basis, which rather than manual basis is indicative of the effective rate in practice.

Example 6.1 Suppose overall statewide NCCI level off-balance is 0.95. Then modified pure premium is 5% lower than manual basis NCCI level pure premium. NCCI targets rate adequacy on a modified pure premium basis. Therefore, aggregate manual pure premium is a little more than 5% higher than if the off-balance were 1.00, but the actual aggregate modified pure premium is the same as if the off-balance were 1.00.

6.2 The Cap on the Maximum Mod

Mods are limited to a maximum value of 1.10 + (0.0004 x E/G), where E is the experience period expected loss and G is the G-value or approximately the State Average Cost per Claim (SACC) / $1,000.

Example 6.2 Suppose a small risk with E = $5,000 and Ep = $2,000 in a state with SACC = $10,000 experiences two claims of $5,000 each in different occurrences over the experience period. The mod formula produces a mod of 1.50, but the actual mod is reduced to the cap of 1.30

This cap offers a further buffer, in addition to split credibility and the loss limit, against the impact of extreme actual loss outcomes for small risks.

6.3 Exclusion of 70% of Medical Only Losses

In some cases, the total increase in future modified premium and other employer administrative expenses due to a small claim can exceed the cost of the claim itself. If these claims involve only medical expenses, policyholders might be inadvertently incentivized to pay the losses directly and not report them to insurers.

Example 6.3 In Example 6.2, had the risk experienced no losses in the experience period the mod would have been 0.87. Alternatively, if this risk experienced one $5,000 medical only claim, the mod without the 70% exclusion of medical only losses, would have been 1.19. If annual manual policy premium for this risk is $7,000, the difference between the mod due to the one claim and a loss free mod would amount to $2,240 per year in additional premium, or a total of $6,720 over the three years in which this claim remains in the experience period. The employer might be incentivized to pay the claim directly and fail to report it to the insurer, in order to save $1,720 and possibly some other administrative expenses.

These small claims, if not reported, still contain predictively valuable information that is not available for the mod calculation. Therefore, losses on medical only claims, which make up the vast majority of the small claims are reduced by 70% to minimize this non-reporting incentive.

Example 6.4 Applying the 70% exclusion of medical only losses in Example 6.4, only $1,500 of actual loss would be used in the mod formula, leading to a mod of 0.96. The 3 year increase in premium due to reporting the claim would only total $1,890 and the insurer would not likely see any net incentive to directly pay the $5,000 claim rather than reporting the claim to the insurer.

6.4 Caps on Multiple Claim Occurrences

Total primary actual losses from a single occurrence are limited to twice the split point.

Example 6.5 Suppose a risk has a multiple claim occurrence resulting in 3 lost time claims of $15,000 each. Assuming a split point of $15,000, only $30,000 of the total incurred loss would be counted as actual primary loss, with the remaining $15,000 counted as excess loss.

Total ratable actual losses from a single occurrence are limited to twice the SAL.

Example 6.6 Suppose in Example 6.5 the SAL was $250,000 and the three lost time claims from this single occurrence were $200,000 each. The primary actual loss would still be $30,000 but the actual excess loss would be limited to $470,000, even though none of the individual claims exceeded the SAL.

These per occurrence limits offer yet another buffer against the impact of extreme actual loss outcomes, particularly for small risks.

6.5 Exclusions of Catastrophes and Certain Non-ratable Losses

Some types of non-ratable losses are not used to calculate the mod. Similarly, some types of premium charges are not multiplied by the mod factor. Examples of non-ratable losses or charges which the mod usually does not apply to include: some types of catastrophes and terrorism, expense constants, premium discounts, and some types of disease exposures.

Example 6.7 A risk with payroll $500,000 might be charged $0.10 /($100 payroll), or $500 total for this risk, as a provision for terrorism to which the mod is not applied. If the risk had a mod of 1.10 and manual premium of $10,000, not including the terrorism provision, then after modification the total including the terrorism provision would be $11,500.

6.6 Net Experience Rating (or Net Reporting)

NCCI's ERP is intended to be used with losses gross of any deductible that may be elected by the employer. Some states, however, require the use of losses net of deductibles for experience rating. (In rare cases, states require the reporting of net losses in unit data. This is referred to as net reporting.) In these states, while the actual experience period losses are net of policy deductibles, the ELRs and D-ratios do not account for deductibles. This mismatch results in double counting the impact of the deductible. In addition to producing systematically lower mod values, a deductible credit is also applied to manual basis premium.

Example 6.8 Suppose a risk with $10,000 in manual premium has a deductible that reduces expected losses by 10%. Due to net reporting, the expected mod for this risk might average 0.95 rather than 1.00. Although actuarial principles suggest the expected premium should be $9,000, the double counting effect of net experience rating results in an expected premium of $8,550.

Although NCCI aggregate ratemaking procedures prevent net experience rating from resulting in a statewide inadequacy, it does create an inequitable benefit for risks with deductible policies versus those without. In the rare case of net reporting, the claim values gross of the deductible are not even reported to NCCI. In a few rare cases, an option exists for no deductible credit and the premium benefit flows through the mod after a while, thereby eliminating any potential double counting.

Example 6.9 If there were no deductible credit in Example 6.8, the mod calculated on a net reporting basis would, on average, partially account for the deductible due to credibility limits. For example, the deductible might, through the mod, result in an expected credit of $500.

What is most consistent with actuarial principles is to treat deductibles as non-ratable proportional credit. This would involve calculating the mod with both actual and expected experience period losses gross of deductibles and applying deductible credits separately to premium.

Example 6.10 If the mod were calculated gross of deductibles in Example 6.8, the expected premium would be $9,000.

6.7 Interstate Risks

As of this writing (2013), 41 states and DC participate in NCCI's interstate ERP. Nine states have their own form of workers compensation experience rating. For interstate risks, experience mods must be calculated separately for the ERP and each of the nine non-ERP states.

Example 6.11 A interstate risk might have a mod of 0.85 calculated entirely based on its exposure and losses in ERP states and a mod of 1.10 calculated entirely based on its losses and exposure in California. These mods are not combined, but are separately applied to manual rates in the ERP states and California, respectively. If the manual pure premium in the ERP states is $500,000, then modified pure premium is $425,000; likewise, if manual pure premium in California is $200,000 then modified pure premium is $220,000.

When a risk has payroll exposure in more than one of the states participating in the interstate ERP, experience period actual and expected losses must be separately calculated according to the SALs, ELRs and D-ratios in each state. These losses are summed across states, and a single set of credibility values must be determined for the mod formula. The current procedure is to weight together individual state W and B credibility values using experience period expected losses.

Example 6.12 Suppose Risk A has $8,000 of experience period expected ratable losses in State X where the SACC is $7,000 and $12,000 of expected loss in State Y where the SACC is $15,000.

State	Experience Period Expected Ratable Loss	State Weights	State Average Claim Cost	Weight	Ballast
X	$8,000	40%	$7,000	0.060	$17,500
Y	$12,000	60%	$15,000	0.054	$37,500
Total	$20,000	100%		0.056	$29,500

At the time of this writing, NCCI is considering replacing the current weights and ballasts with a method of combining implied claim counts across states to determine Zp and Ze for interstate risks.

Example 6.13 Here is Example 6.12 repeated with an implied claim count method of determining credibility.

State	Experience Period Expected Ratable Loss	State Average Claim Cost	Implied Claim Counts	Zp	Ze
A	$8,000	$7,000	1.143		
B	$12,000	$15,000	0.800		
Total	$20,000		1.943	43.7%	3.1%

ACKNOWLEDGMENT

I am thankful for the support of the Actuarial and Economic Services Division of NCCI in providing the allocation of my time for writing. Specific thanks go to Gerald Ordoyne, Karen Ayres, and Tony DiDonato, of NCCI, and Gary Dean, of Ball State University, for help in clarifying specific parts of this study note.

APPENDIX A: A BRIEF OVERVIEW OF HISTORICAL GENERATIONS OF THE NCCI EXPERIENCE RATING PLAN

A.1 Early 20th Century

Individual risk experience rating adjustments were being implemented in workers compensation in the years immediately following the passage of the first state workers compensation laws in the United States in 1911. Even before the NCCI itself was formed in 1922, its two predecessor organizations (the National Workmen's Compensation Service Bureau and the National Council on Compensation Insurance) had recognized the importance of credibility in experience rating and also the need to split experience period losses into pieces with separate credibility values. By 1918, these organizations began to offer experience rating plans that split losses into two categories, "Death and Permanent Total" losses and "All Other" losses. By 1923 the categories had been changed to "normal" and "excess". The credibility formulas followed the familiar forms of:

$$Z = \frac{P}{P+K} \quad \text{or} \quad Z = \frac{P+C}{P+K+C}$$

Early plans also typically included minimum thresholds for experience rating and a self-rating point at which a large risk was given 100% credibility.

A.2 Mid-20th Century

By 1940, the ERP used a multiple split for primary losses by dividing individual claims into layers of \$500 with the loss in the nth layer allocated to primary in proportion to (2/3)n . This formula rises toward an asymptotic limit of 1,000 as the actual loss for a claim becomes large.

Example A.2.1 Under the 1940 split formula, an actual experience period claim of \$7,000 would be allocated \$500(2/3) + \$500(4/9)+…+\$500(2/3)14 = \$997 to primary and \$6,003 to excess.

In 1961, the split of primary and excess was changed to a continuous split that counted all losses for a claim of \$2,000 or less as primary and used the following formula for claims exceeding \$2,000.

$$A_p = \frac{10000A}{8000 + A} \qquad A > 2000$$

This formula rises toward an asymptotic limit of \$10,000 as the actual loss for a claim becomes large.

Example A.2.2 Under the 1961 split formula, an actual experience period claim of \$7,000 would be allocated \$4,667 to primary and \$2,333 to excess.

For both the 1940 and 1961 ERPs, the credibility values were determined by three constants: K, Q, and S. S was the self-rating point, above which both Zp and Ze were 100%. Below Q, excess credibility was 0%.

	$0 < E < Q$	$Q < E < S$	$S < E$
Zp	$\frac{E}{E+K}$	$\frac{E}{E+(1-W)K}$	1
Ze	0	$\frac{W\,E}{E+(1-W)K}$	1
W	0	$\frac{E-Q}{E-S}$	1

B	*K*	*(1-W)K*	*0*

A.3 RERP

In 1991, NCCI introduced the Revised Experience Rating Plan (RERP) with the following main changes:

1. A single split point was adopted where losses per claim under $5,000 count as primary.
2. The Ballast formula was changed to increase with risk size.

$$B = Max\left[7500, E\left(\frac{0.1E + 2570G}{E + 700G}\right)\right]$$

3. The Weight formula was changed, resulting in an increase in weight for small risks and a decrease in weight for large risks.

$$W = Max\left[0.07,\ \frac{E + B}{E + Max\left[150000,\ E\left(\frac{0.75E + 203825G}{E + 5100G}\right)\right]}\right]$$

4. The limit on individual losses, or State Accident Limit (SAL), was changed from 10% of the Self Rating Point (SRP) to a smaller value based on 25 times the State Average Claim Cost (SACC).

The new W and B formulas were based on a parameterization motivated by some assumptions about systematic and random variances of the primary and excess losses. These formulae kept primary and excess credibility always below 100% and eliminated the SRP.

A.4 GERT

In 1995, NCCI introduced the Graduated Experience Rating Table (GERT) with the following main changes:

The minimum Ballast value was changed to be 2,500 G where the G-value is the SAL / $25,000.

$$B = Max\left[2500G, E\left(\frac{0.1E + 2570G}{E + 700G}\right)\right]$$

The Weight formula was change to include an indexed component, also using the G-value.

$$W = \frac{E + B}{E + Max\left[60000G,\ E\left(\frac{0.75E + 203825G}{E + 5100G}\right)\right]}$$

The cap on the maximum mod was changed from a table of interval ranges to a continuous formula, partially indexed by the G-value.

1 + 0.00005(E + 2 E/G)

The cap for interstate risks was set at the effective cap for the state with largest experience period expected losses.

A.5 ERA

In 1998, NCCI introduced the Experience Rating Adjustment (ERA) with the following main changes:

1. 70% reduction on ratable medical only claims to reduce underreporting incentives
2. A change in the weight formula to improve performance by increasing weight values.

$$W = \frac{E + B}{E + Max\left[60000G,\ E\left(\frac{0.375E + 150000G}{E + 5100G}\right)\right]}$$

ERA also included a provision for indexing the split point, based on review of country wide averages costs per case. However, this review process had no fixed schedule and claim severities were fairly flat around the time when ERA was introduced. As a result, the split point remained at $5000.

A.6 2013 ER Changes

In 2013, NCCI introduced several changes to the ERP based on a multi-year review. The main changes included:

1. The split point was raised beginning in 2013, from $5,000 to an initial $10,000 in year one, $13,500 in year two, and $15,000 plus two years of severity indexation adjustment in year three.
2. Subsequently, the split point will be updated with annual rate filings based on the countrywide cost per case index.
3. The mod cap was changed to a fully indexed form which is higher for the smallest risks and lower for other risks.

$$1.10 + 0.0004\ (E/G)$$

EXERCISES

These exercises may require computer capability for calculation, simulation, and/or graphics. Some exercises require creative answers and/or may not have a definite correct answer. Although all calculations can be done with algebra, calculus, and numerical analysis, other approaches include simulation and specifically Gibbs sampling (using WinBUGS, JAGS, etc.).

Section 2

1. In Example 2.1 what is the expected squared error of the credibility estimate ?
2. In Example 2.2 what is the expected squared error of the predictive estimate ?
3. In Example 2.3 create some hypothetical extraneous information and/or assumptions and calculate the greatest accuracy linear credibility ? What sort of testing would you do to validate this model ?
4. In Example 2.4 what is the formula for the slope and intercept of the linear regression model in terms of the items mentioned ? What sort of testing would you do to validate this model ?
5. Suppose an individual risk's true expected manual loss ratio prior to observation is uniformly distributed on [50%, 150%] and the actual manual loss ratio was exponentially distributed around this mean. Construct a static perspective estimate of the true manual loss ratio, using the actual observation that has very good lift but almost no equity.
6. Repeat the previous exercise, but this time reduce the lift by "half" but maximize the

equity.

7. Repeat the previous exercise trying to get very good lift and very good equity at the same time.
8. Describe as many economic and social benefits that you can think of from experience rating individual risks.
9. Can you think of any ways in which experience rating might complicate some economic or social objectives?
10. Considering Example 2.10, can you describe several different plausible situations where there would be no safety incentives resulting from experience rating individual risks ?
11. Below is the result for a hypothetical quintile test.

Quintile	Experience Period Loss Ratio	Prospective Period Unmodified Loss Ratio	Prospective Period Modified Loss ratio
1	0.30	0.60	0.90
2	0.55	0.80	1.05
3	1.10	1.05	0.95
4	1.50	1.20	1.10
5	2.00	1.40	1.30

   a. Calculate the old and new quintile test statistics.
   b. Estimate the effective scalar credibility Z in this mod.
   c. Estimate the optimal effective scalar credibility Z in this situation.
   d. Recalculate the old and new quintile test statistics using your credibility estimate from c.

12. Use Lognormal distributions, including negative values, to approximate all of the distributions in Example 2.1 Estimate the values of the unmodified and modified quintile loss ratios if you determined the credibility estimate for a large set of different risks in one year and then applied the corresponding modification factors to another year of data. Calculate the old and new quintile test statistics.
13. For Example 2.2 assume the annual variance of loss ratios is 1.25 and repeat quintile loss ratio and statistics calculations as in the last exercise for the 15% serial correlation derived modification factors.
14. For Example 2.1 using a Lognormal distribution derive an estimate for the prospective modified loss ratio of a single risk with a mod value of $m > 0$.
15. For Example 2.2 using a Lognormal distribution derive an estimate for the prospective modified loss ratio of a single risk with a mod value of $m > 0$.
16. In Example 2.1 suppose that the expected loss of each risk changed from one year to the next by a multiplicative factor, independently distributed between different risks with mean 1.0 and standard deviation 0.10. What should the credibility factor Z be?
17. Discuss the contradiction between an insurable loss process for an individual risk and a loss amount that can be accurately predicted for an individual risk.
18. Explain the sense in which a cross subsidy between risks can occur if individual risk experience rating is not applied.
19. Contrast manual ratemaking with individual risk experience ratemaking.
20. Is individual risk ratemaking more valuable for a simple manual ratemaking system or a more complicated manual ratemaking system?

21. Would it be meaningful or practical to apply a quintile test to a retrospective individual rate plan? Discuss lift and equity for this context.
22. You are given the following data:

Risk #	Year 1 Manual Expected Loss	Year 1 Actual Loss	Year 2 Manual Expected Loss	Year 2 Actual Loss	Year 3 Manual Expected Loss
1	\$5	\$0	\$5	\$0	\$4
2	5	1	5	2	3
3	10	2	5	22	6
4	20	5	23	2	18
5	25	0	19	3	15
6	50	11	30	34	28
7	75	353	70	26	85
8	100	109	142	179	107
9	150	105	106	4	122
10	200	179	265	1,170	296
11	300	470	447	248	471
12	500	356	488	147	381
13	750	2,029	506	720	388
14	1,000	763	582	1,015	704
15	2,000	1,667	1,243	442	1,121
16	4,000	4,100	2,616	2,796	3,232
17	6,000	5,794	7,103	6,014	9,339
18	8,000	17,481	5,408	12,471	7,475
19	9,000	9,887	8,500	8,848	7,047
20	10,000	4,787	6,972	5,575	7,335

Using manual expected loss as a measure of volume calculate the empirical non-parametric estimate for the credibility constant K and for each risk the corresponding estimate for Year 3 expected losses. Recalculate K and the Year 3 expected losses from a predictive perspective.

23. Using the actual outcome losses below perform a 2 category quantile test on each of the results from #22 and make a chart of the results.

	Year 3
Risk #	Actual Loss
1	$2
2	0
3	2
4	18
5	122
6	5
7	38
8	39
9	172
10	565
11	369
12	219
13	125
14	918
15	1,191
16	2,698
17	9,212
18	13,296
19	4,923
20	5,395

Section 3

1. For Example 3.1 prove (show through calculation) that the credibility is Z = 16.7%
2. In Example 3.1 if the standard deviation of the true mean prior to observation is $x > 0$, instead of 50%, and the conditional coefficient of variation is $y > 0$, instead of 100%, what is the general formula for the credibility Z.
3. Redraw the chart in Example 3.1 using x = 75% and y = 200% as defined just defined in 2.
4. Redraw the chart in Example 3.1 using an Inverse gamma distribution with coefficient of variation 100% instead of a lognormal distribution for the conditional distribution of the loss ratio.
5. Why does the Bayesian estimate in Example 3.1 show the curvature it does?
6. What assumptions about credibility lead to the straight line in Example 3.1?
7. For Example 3.4 prove (show through calculation) that the credibility is Z = 42.9%
8. Discuss qualitatively why the number and amount of claims for the same total dollars should lead to such different estimates.
9. For Example 3.5 recalculate all the estimates for 2 claims of $500,000 each and then for 8 claims of $125,000 each.
10. Discuss how the situation in Example 3.6 might affect the estimation of expected losses.

11. Discuss how the situation in Example 3.7 might affect the estimation of expected losses.
12. Calculate probabilities of various numbers of claims, as in Example 3.8, when the Gamma distribution has coefficient of variation 15% and the overall frequency is 0.5 or 50, respectively.
13. Calculate probabilities of various claim amounts, as in Example 3.9, using an Inverse Gamma distribution, instead of a Lognormal, with a mean of 15,000 and standard deviation of 75,000.
14. Discuss in general terms, the practical impact of skewness on loss data and actuarial estimation.
15. Why do ordinary credibility and basic statistical models handle skewness poorly?
16. Discuss the real world reasons that drive large changes in severity over time.
17. Discuss the real world reasons that lead to big differences in severity between states.
18. Why do differences in severity between states and over time complicate actuarial estimation?

Section 4

1. For Example 4.1 prove (show through calculation) that Zp = 20%, Ze = 3.2%, Ep = 0.954, and Ee = 0.046
2. For Example 4.1 recalculate Zp, Ze, Ep, and Ee for x = 75% and y = 200% as described in Exercise 3 for Section 3.
3. For Example 4.1 recalculate Zp, Ze, Ep, and Ee using 300% as the dividing line between primary and excess losses.
4. For Example 4.2 prove (show through calculation) that Zp = 47.9%, Ze = 34.2%, Ep = $6,908 and Ee = $2,857.
5. For Example 4.3 suppose the loss limit at $100,000 was another split point. Calculate Zx for the layer excess of $100,000.
6. For Example 4.3 suppose split point was $30,000 and the loss limit was $90,000. Recalculate Zp, Ze, Ep, and Ee.
7. For Example 4.3 recalculate all the estimates for 2 claims of $500,000 each and then for 8 claims of $125,000 each.
8. For Example 4.4 recalculate all the estimates for 1 claims of 10 and then for 5 claims of 150 each.
9. Discuss the relative merits of split credibility and loss limits versus a direct Bayesian calculation.
10. What is the difference between a split point and a loss limit?
11. Discuss why correlation between covariates, such as the primary and excess experience losses, is a problem for linear regression models.
12. Discuss the difference between linear regression models and credibility models.
13. How does the Weight and Ballast representation of credibility slightly reduce the complexity of calculating the mod?
14. What is the effective result for credibility of placing a minimum value on the Ballast?
15. Write Ze and Zp as functions of x = E/G for the latest credibility constants and minimum constraints in the ERP.

16. Using the latest credibility constants and minimum constraints in the ERP calculate the credibility values for the table below.

E	Ze	Zp	W	B
10 G				
100 G				
1,000 G				
10,000 G				
100,000 G				
1,000,000 G				
10,000,000 G				
100,000,000 G				
1,000,000,000 G				

17. For Example 4.1 estimate optimal values for Zp and Ze in terms of flattening a quintile test.
18. For Example 4.2 estimate optimal values for Zp and Ze in terms of flattening a quintile test.
19. Why is credibility more dependent on the expected number of claims than on the total expected loss amount?
20. Why is it important to test the ERP, or for that matter statistical models in general, on a data set other than the data set used to fit the parameters of the model?
21. What changes over time in the loss process for individual risks might severity indexation fail to capture?
22. Aside from practical limitations to what extent, in principle, can experience rating improve rate equity for very small, low-credibility risks?

Section 5

1. Why is it necessary to assign risk IDs to individual employers and track them over time?
2. Discuss the challenges and ambiguities of assigning risk IDs to individual employers and tracking them over time.
3. Discuss why expected ratable losses differ from prospective expected losses in all the ways listed in Section 5.2
4. If the 70% exclusion of medical only losses were eliminated what impact would you expect it to have on ELRs, ELAFs, and D-ratios?
5. If the SAL limit on individual losses were eliminated what impact would you expect it to have on ELRs, ELAFs, and D-ratios?
6. If ELRs increase, all other things staying the same, what impact would you expect on mod values and on the equity and lift of the mod?
7. If D-ratios increase, all other things staying the same, what impact would you expect on mod values and on the equity and lift of the mod?
8. If the fraction of losses attributable to Permanent Total claims increases what impact would you expect it to have on ELRs, ELAFs, and D-ratios?
9. If D-ratios were used that were much too high what effect would it have on the quintile test? Draw a hypothetical chart of such a quintile test.
10. If D-ratios were used that were much too low what effect would it have on the quintile test? Draw a hypothetical chart of such a quintile test.

11. If ELRs were used that were much too high what effect would it have on the quintile test? Draw a hypothetical chart of such a quintile test.
12. If a ELRs were used that were much too low what effect would it have on the quintile test? Draw a hypothetical chart of such a quintile test.
13. Discuss the relative importance of Ze, Zp, ELRs, and D-ratios for mod performance.

Section 6

1. Discuss the contrast between off-balance versus equity and lift.
2. What would be the consequence if off-balance became an extreme value, such as below 0.50 or above 1.50?
3. What determines the minimum possible mod value?
4. Using the latest credibility constants and minimum constraints in the ERP and the cap on the maximum mod calculate the values for the minimum and maximum possible mods for the table below.

E	Minimum Mod	Maximum Mod
10 G		
100 G		
1,000 G		
10,000 G		
100,000 G		
1,000,000 G		
10,000,000 G		
100,000,000 G		
1,000,000,000 G		

5. If there are three claims in the experience period, each of which exceeds the split point but is less than half of the SAL, how does the multiple claim occurrence limit affect the mod value?
6. Could a debit mod under rating gross of deductibles become a credit mod under rating net of deductibles?
7. Could a credit mod under rating gross of deductibles become a debit mod under rating net of deductibles?
8. Is predictive information lost by excluding 70% of medical only losses?
9. Why does it make more sense to use the total size across states of an interstate risk to determine credibility values rather than just calculating an intrastate mod in each state for an interstate risk?
10. For interstate risks under what circumstances does weighting W and B versus determining Zp and Ze by implied expected claim counts result in the greatest and least differences, respectively, for the mod calculation ?
11. Discuss arguments for and against applying the mod to typical non-ratable losses or charges which the mod usually does not apply to, such as: some types of catastrophes and terrorism, expense constants, premium discounts, and some types of disease exposures.

Appendix

1. Discuss some plausible differences in underlying assumptions between these two familiar formulas for credibility.

$$Z = \frac{P}{P+K} \quad \text{or} \quad Z = \frac{P+C}{P+K+C}$$

2. Discuss arguments for and against a self rating point in terms of risk size.
3. What would a generalized version of the 1940 split formula (as in Example A.2.1) be?
4. How could the 1940 split formula (as in Example A.2.1) be rewritten to include a severity index?
5. What would a generalized version of the 1961 split formula (as in Example A.2.2) be?
6. How could the 1961 split formula (as in Example A.2.2) be rewritten to include a severity index?
7. How could the 1940 and 1961 formulas for Zp, Ze, W, and B be rewritten to include a severity index?
8. Suppose and the 1940/1961 formulas for Zp and Ze use K = \$50,000, Q=\$25,000, and S = \$1,000,000. For what values of G and E would the ERA formulas for Zp and Ze give the same values, lower values, and higher values, respectively?
9. Compare the mod cap under GERT to the mod cap under ERA. For what values of E and G are the caps equal, the GERT cap higher, and the ERA cap higher, respectively?
10. Draw a line chart of Zp under RERP, GERT, and ERA, respectively, for E ranging from \$100 to \$1,000,000 and G = 5, 10, and 15, respectively.
11. Draw a line chart of Ze under RERP, GERT, and ERA, respectively, for E ranging from \$100 to \$1,000,000 and G = 5, 10, and 15, respectively.

Related Paper: "The Optimal Number of Quantiles For Predictive Performance Testing of the NCCI Experience Rating Plan"

1. Ignoring noise-to signal resolution limits on statistical clarity constraints per se, discuss whether more quantiles might or might not reveal more information about ERP performance.
2. Can you think of similar simple alternative metrics to the old and new quintile test statistics?
3. Discuss the difference between lift and what exactly is being measured by the new quintile test statistic.
4. The difference between A* versus A and B* versus B is a historical anomaly, as bootstrapping was done when A and B were defined for the new quintile test statistic. Nevertheless, discuss what significance there might be to this difference between including bootstrapping or not in these variances.
5. Suppose an individual risk has a 10% chance of having total losses > 0 and these aggregate losses > 0 conditionally follow a log normal distribution with mean \$10,000 and coefficient of variation 500%. How many such risks must be sampled so that there is a 99.9% probability that the sample average will lie in the interval from \$8,000 to \$10,000.
6. Redraw Figure 4, from the paper, using the individual risk loss process in the last problem.
7. In Example 3.1, in this study note, estimate σ/R for: a. the Bayes theorem estimate, and b. the greatest accuracy credibility estimate.
8. In Example 3.1, in this study note, estimate the minimum sample size for a quintile test to produce a noise-to-signal ratio <= 0.25 for: a. the Bayes theorem estimate, and b. The

greatest accuracy credibility estimate.

9. In Example 3.1, in this study note, estimate the minimum sample size for a decile test to produce a noise-to-signal ratio <= 0.25 for: a. the Bayes theorem estimate, and b. the greatest accuracy credibility estimate.
10. In Example 3.4, in this study note, estimate σ/R for: a. the Bayes theorem estimate, and b. the greatest accuracy credibility estimate.
11. In Example 3.4, in this study note, estimate the minimum sample size for a quintile test to produce a noise-to-signal ratio <= 0.25 for: a. the Bayes theorem estimate, and b. The greatest accuracy credibility estimate.
12. In Example 3.4, in this study note, estimate the minimum sample size for a decile test to produce a noise-to-signal ratio <= 0.25 for: a. the Bayes theorem estimate, and b. the greatest accuracy credibility estimate.
13. In Example 4.1, in this study note, estimate σ/R for the split credibility estimate.
14. In Example 4.1, in this study note, estimate the minimum sample size for a quintile test to produce a noise-to-signal ratio <= 0.25 for the split credibility estimate.
15. In Example 4.1, in this study note, estimate the minimum sample size for a decile test to produce a noise-to-signal ratio <= 0.25 for the split credibility estimate.
16. In Example 4.2, in this study note, estimate σ/R for the split credibility estimate.
17. In Example 4.2, in this study note, estimate the minimum sample size for a quintile test to produce a noise-to-signal ratio <= 0.25 for the split credibility estimate.
18. In Example 4.2, in this study note, estimate the minimum sample size for a decile test to produce a noise-to-signal ratio <= 0.25 for the split credibility estimate.
19. Would you pick another rule of thumb in Section 3.2, from the paper, rather than $Z \approx Zp/2$? Discuss the reasons for your pick or the reasons for sticking with $Z \approx Zp/2$.
20. Make tables like the tables shown in Table 2, from the paper, for all of the estimates in Examples 3.1, 3.4, 4.1, and 4.2. There should be 6 tables in total.
21. Using the credibility rule of thumb, and any other reasonable assumptions you might make, estimate what the N/S would be expected to be for Figures 8-12. What would you expect the maximum number of quantiles that would still keep the N/S <= 0.25 to be in each of these situations?
22. Recalculate Table A1 for: a. a quintile test b. a 20 category quantile test.
23. How does the whole concept of the R value described in the paper break down when the distribution of estimates follows an unbounded distribution, like a lognormal, and the number of quantiles used becomes very large ? Construct an example.
24. Discuss the problems with the noise-to-signal measure when the loss process is very skewed, as it usually is, and the sample size is small?
25. NCCI usually does not limit actual loss amounts from the prospective policy period in quintile tests but relies on large data volume. Discuss the possible advantages and disadvantages of using loss limits in the prospective policy period. Particularly, what types of biases might these limits might cause?

5. BIBLIOGRAPHY

Accompanying this study note is a more in depth paper on quantile testing the NCCI EXP:

Evans, Jonathan; and Dean, Curtis Gary, "The Optimal Number of Quantiles For Predictive Performance Testing of the NCCI Experience Rating Plan," to appear in CAS eForum.

The ultimate written sources on the specifics of the NCCI ERP are:

National Council on Compensation Insurance, *Experience Rating Plan Manual for Workers Compensation and Employers Liability Insurance.*
National Council on Compensation Insurance, *Experience Rating Plan User's Guide for Workers Compensation and Employers Liability Insurance.*

For basic statistics on individual claim severity see:

National Council on Compensation Insurance, *Annual Statistical Bulletin.*

Aside from NCCI itself the most important resource on the history and theory behind workers compensation experience rating, and related actuarial methodology is:

Proceedings of the Casualty Actuarial Society, 1914-2005, available at www.casact.org.
Superseded by:
Variance, 2006-present, also available at www.casact.org.

Also of note for related actuarial methodology is:

ASTIN Bulletin – The Journal of the IAA, 1958- present, available at www.actuaries.org.

The classic landmark paper on workers compensation experience rating is:

Dorweiler, Paul, "A Survey of Risk Credibility in Experience Rating," *PCAS* **1934**, XXI, p. 1.

Some other specific key materials include:

Bailey, Robert A., "Experience Rating Reassessed," *PCAS* **1961**, XLVIII, p. 60-82.
Bailey, Robert A.; and Simon, Leroy, J., "An Actuarial Note on the Credibility of a Single Private Passenger Car," *PCAS* XLVI, 1959.
Bühlmann, Hans, "Experience Rating and Credibility," Astin Bulletin, Vol. IV, Part III.
Edwards, John; and Woodward, Joseph, "The Experience Rating of Workers' Compensation Risks," *PCAS* **1915**, II, p. 356-369.
Fisher, Arne, "Note on an Application of Bayes' Rule in the Classification of Hazards in Experience Rating," *PCAS* **1916**, III, p. 43-48.
Gillam, William R.., "Parametrizing the Workers Compensation Experience Rating Plan," *PCAS* **1992**, LXXIX. p. 21-56.
Gillam, W.R.; and Snader, R.H., "Fundamentals of Individual Risk Rating," National Council on Compensation Insurance, (Study Note), 1992, Part I.
Gillam, William R.., "Workers' Compensation Experience Rating: What Every Actuary Should Know," *PCAS* **1992**, LXXIX. p. 215-239.

Greene, Winfield W., "Should the Compensation Premium Reflect the Experience of the Individual Risk?" *PCAS* **1915**, II, p. 347-355.
Johnson Jr., Roger A, "The Multi-Split Experience Rating Plan in New York," *PCAS* **1925**, XI, p. 15-36.
Klugman, S.A.; Panjer, H.H.; and Willmot, G.E., *Loss Models: From Data to Decisions*, Fourth Edition, 2012.
Kormes, Mark. "The Experience Rating Plan as Applied to Workmen's Compensation Risks," *PCAS* **1934**, XXI, p. 81-182.
Kormes, Mark. "The Experience Rating Plan as Applied to Workmen's Compensation Risks - Part II," *PCAS* **1935**, XXII, p. 81-108.
Mahler, Howard C., "An Introduction to Basic Credibility," CAS Forum, 1997, Winter, pp. 67-103.
Mahler, Howard C., "A Graphical Illustration of Experience Rating Credibilities," *PCAS* **1998**, LXXXV, p. 654.
Mahler, Howard C.; and Dean, Curtis G., "Credibility," Chapter 8, *Foundations of Casualty Actuarial Science,* 4th Edition, 2001.
Meyers, Glenn G., "An Analysis of Experience Rating," *PCAS* **1985**, LXXII, p. 278.
Michelbacher, Gustav F., "The Practice of Experience Rating," *PCAS* **1918**, IV, p. 293-324.
Perryman, F. S., "Experience Rating Plan Credibilities," *PCAS* **1937**, XXIV, p. 60.
Senior, Leon,"Experience Rating In Rem and In Personam," *PCAS* **1925**, XI, p. 211-217.
Smick, J. J., "Merit Rating - The Proposed Multi-Split Experience Rating Plan and the Present Experience Rating Plan," *PCAS* **1939**, XXVI, p. 84-129.
Stevens, Waldo, "A Review of the Experience of Massachusetts Workmen's Compensation Experience Rated Risks," *PCAS* **1959**, XLVI, p. 87-122.
Uhthoff, D. R., "The Compensation Experience Rating Plan-A Current Review," *PCAS* **1959**, XLVI, p. 285.
Venter, G.G., "Experience Rating—Equity and Predictive Accuracy," NCCI Digest, April 1987, Volume II, Issue I, pp. 27-35.
Whitney, Albert, "The Theory of Experience Rating," *PCA* **1918**, IV, p. 274-292.