

# The Optimal Number of Quantiles For Predictive Performance Testing of the NCCI Experience Rating Plan

Jon Evans and Curtis Gary Dean

---

## Abstract

**Motivation.** Quantile testing is a key technique for fitting parameters and testing performance in workers compensation experience rating and the number of quantile intervals must be specified for such a test.

**Method.** A model is developed to compare the error in the quantile test empirical estimates of relative pure loss ratios to the interquantile differences between expected pure loss ratios. The ratio of these two quantities can be interpreted as a kind of noise-to-signal ratio which must be kept below a certain tolerance for the results of the quantile test to be sufficiently clear in a statistical sense. The formula for this ratio is a function of the random variation in individual risk loss ratios, a measure of the variation in experience modification (mod) values, the sample size of risks, and the number of quantiles. Theoretical model predictions are compared to empirical results from bootstrap quintile tests of the National Council on Compensation Insurance (NCCI) Experience Rating Plan (ERP). A generic type of individual risk credibility, though very different from the credibility values in the ERP, can be used to estimate the ratio of the standard deviation in individual risk loss ratios to a measure of variation in the mod values.

**Results.** The model predicts that the noise-to-signal ratio grows in proportion to the 1.5 power of the number of quantiles and in inverse proportion to the 0.5 power of the sample size of risks. Empirical quintile and decile tests of NCCI's Experience Rating Plan are consistent with model predictions.

**Conclusions.** Increasing the number of quantiles requires a much greater proportional increase in data volume to maintain a constant noise-to-signal ratio. Combined with the typical large magnitude of individual-risk loss process variance to loss parameter variance, this explains the use of few quantiles, specifically quintiles, for testing NCCI's Experience Rating Plan.

**Keywords.** Quantile Test, Quintile Test, Experience Rating, Workers Compensation, Predictive Modeling, Credibility, Lift, NCCI

---

## 1. INTRODUCTION

The National Council on Compensation Insurance (NCCI) Experience Rating Plan (ERP) is routinely tested by sorting risks based on experience modification factor (mod) values into *quintiles*, each containing an equal number of risks. A natural question is why specifically are quintiles, that is five *quintiles*, used to test the ERP. Naïve intuition would suggest that more quantiles would reveal more details about performance. However, like virtually all statistical tests, quantile testing attempts to uncover underlying systematic or signal information by filtering out random or noisy information. Using more quantiles affects both the signal and noise properties of each quantile. This paper will explain how a constraint on the relative size of noise to signal in the quantile test suggests criteria for

selecting an optimal number of quantiles.

The content of this paper is primarily related to individual risk rating, predictive testing, and credibility. Although the application is to workers compensation specifically, the methods shown are generally applicable to predictive models of losses for individual risks in all lines of property and casualty insurance. The predictive testing of individual risk rating plans has been discussed previously in [4], [8], [12], and [17]. Specifics of the NCCI Experience Rating Plan can be found in [15].

## **1.1 Objective**

This paper will provide a justification for selecting a given number of quantiles, or comparing different possible numbers of quantiles, for testing the ERP. No pretense is made of meeting the standards of mathematical rigor or rock solid logical derivation, but the model will follow from a sensible line of reasoning and whatever specific parameter values selected are intended for application with empirical data. Some model results will be shown to be reasonably consistent with patterns seen in some empirical quintile and decile tests, providing validation to the extent of practical use.

## **1.2 Outline**

The remainder of the paper proceeds as follows. Section 2 will describe a noise-to-signal (N/S) model for the resolution of a quantile test. Section 3 shows how the properties of this model are consistently demonstrated in both empirical data and hypothetical examples. Appendix A will show a connection between the N/S ratio model and credibility.

## **2. BACKGROUND AND METHODS**

The purpose of experience rating is to improve the estimate of future expected losses for an individual risk using previous actual loss experience for that risk. The basic underlying formula for the workers compensation experience rating modification factor is (2.1).

$$1 + Z_p \frac{A_p - E_p}{E} + Z_e \frac{A_e - E_e}{E} \quad (2.1)$$

$A_p$  = actual primary ratable loss from the experience period

$A_e$  = actual excess ratable loss from experience period

## *Optimal Number of Quantiles For Predictive Testing*

$E_p$  = expected primary ratable loss from the experience period

$E_e$  = expected excess ratable loss from experience period

$E = E_p + E_e$

$Z_p$  = primary credibility

$Z_e$  = excess credibility

Although there are many other refinements to this basic formula as used in the ERP, the mod essentially compares actual loss experience to manual expected losses. The  $Z_p$  and  $Z_e$  values, which vary by size of risk, determine the credibility assigned to experience. The prospective manual premium for a policy is multiplied by this mod value, which is always positive but may be greater or less than 1.0, as part of premium calculation.

### **2.1 Quintile Testing the Experience Rating Plan**

Risks are sorted by mod value and based on this order split into five quintiles, each containing an equal number of risks. The performance of the ERP is tested by comparing relative pure loss ratios – the ratio of actual losses to expected losses – across the quintiles for the policy year to which the experience mod applies both before and after the mod is applied to manual expected losses. Several adjustments are made to reported actual losses and NCCI manual rates so that losses are on an ultimate basis and exclude both underwriting and loss adjustment expenses. A scale factor is also used to set total modified expected losses equal to total actual losses by risk hazard group within each state. This equalization isolates relative differences between risks, which the mod is designed to address, from issues of aggregate rate adequacy, which are addressed by manual rate levels.

### **2.2 Lift and Equity**

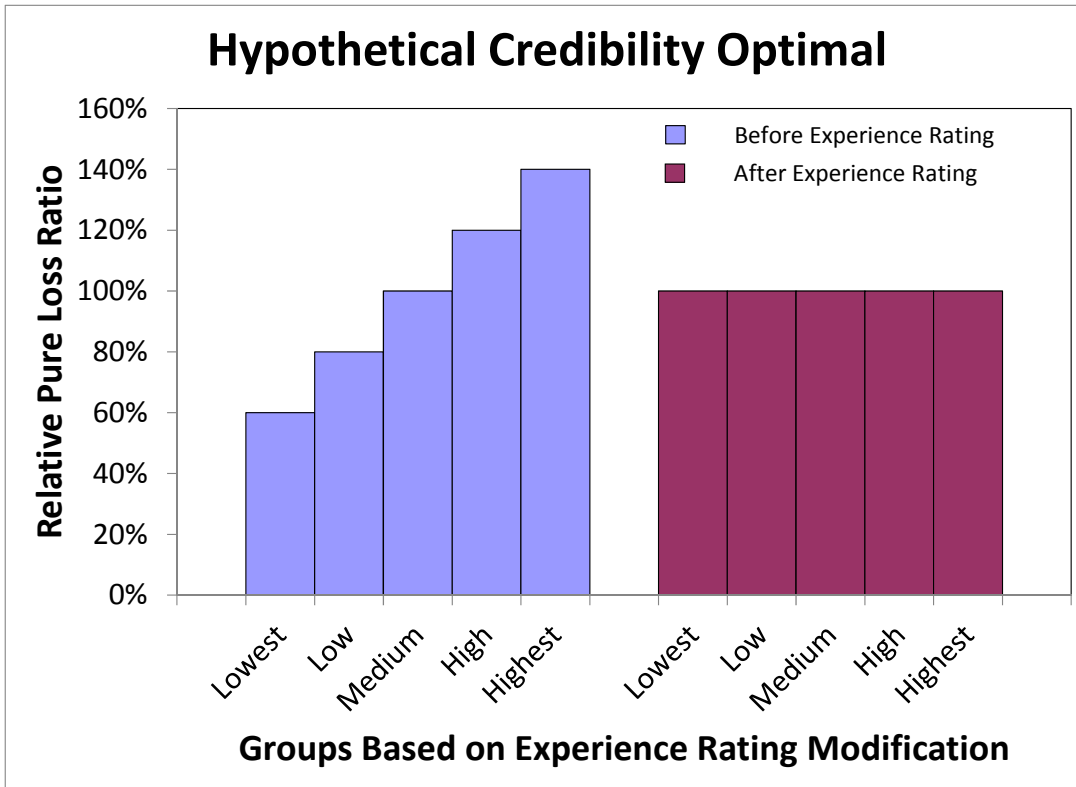
If the ERP can actually differentiate between risks then pure loss ratios calculated from manual rates should increase when moving from lower quintiles to higher quintiles. This increasing slope is called *lift*. The steeper the lift the more value the mod offers in differentiating risks.

Including the mod in the calculation of expected losses should result in nearly equal pure loss ratios, or *equity*, across quintiles if the ERP is working well and there is sufficient data in each quintile. This flatness of the relative modified pure loss ratios reveals how well the mod achieves equity between risks.

Figure 1 displays an idealized hypothetical example. The bars on the left side of the chart demonstrate the concept of lift. The mod is able to differentiate between risks and the slope shows

the degree of differentiation between them. If risks were rated only on manual rates then risks in the quintiles to the left would produce more favorable underwriting results.

Figure 1: Experience Rating Plan Displaying Lift and Equity



The second set of bars in Figure 1 demonstrates that the combination of manual rates and the mod is equitable. The underwriting results are equal across the quintiles. Hence underwriting results are the same for risks with different mod values.

Figure 2 displays a hypothetical ERP with considerable lift as seen by the bars on the left side. The five bars on the right side demonstrate that the mod has not fully corrected for predictable differences in manual loss ratios. The rating plan does not charge risks with high mods enough and overcharges risks with lower mods. This inequity is generally the result of not assigning enough credibility to individual risk experience. In contrast, Figure 3 displays a hypothetical ERP that over corrects for predictable differences – this generally results from assigning too much credibility to individual risk experience.

Figure 2: Experience Rating Plan That Undercompensates for Experience

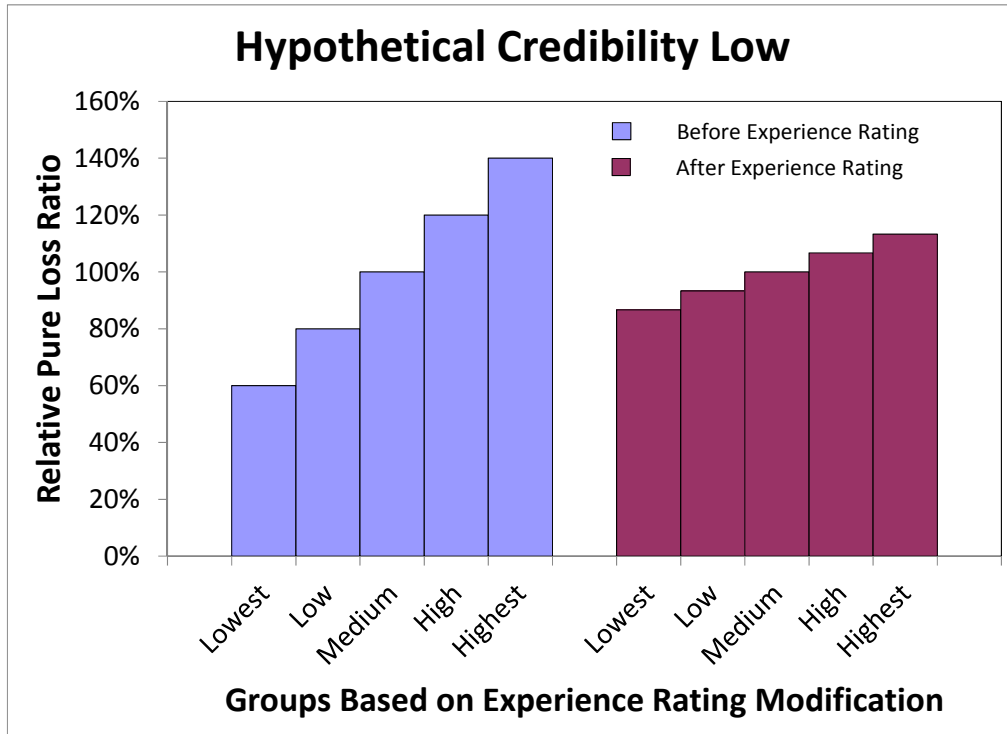
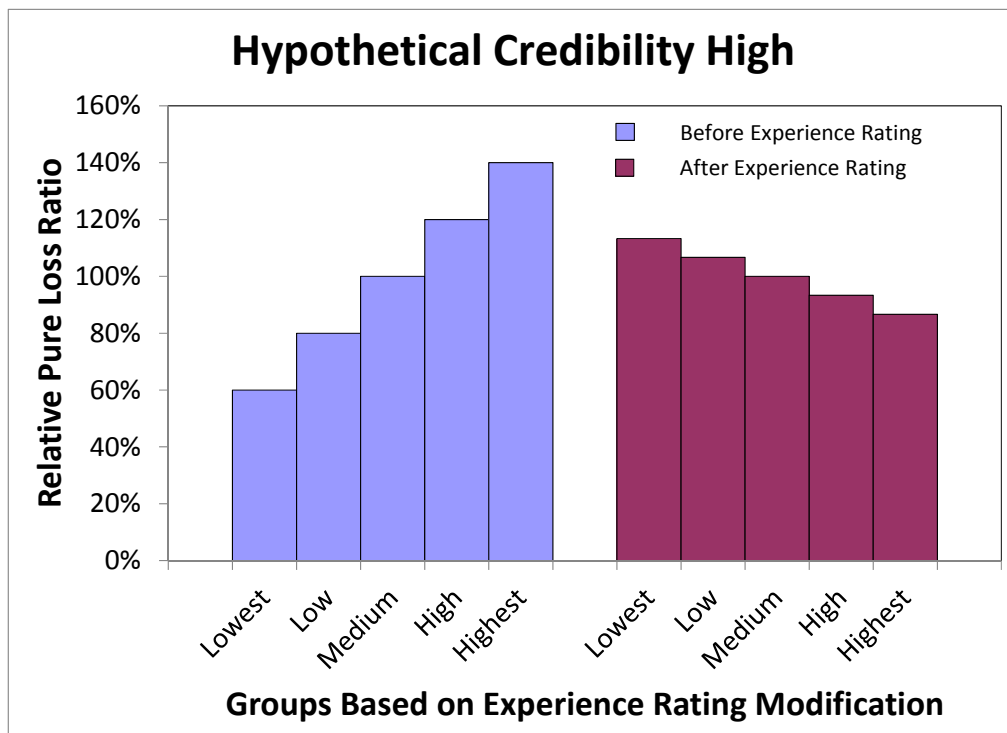


Figure 3: Experience Rating Plan That Overcompensates for Experience



### Optimal Number of Quantiles For Predictive Testing

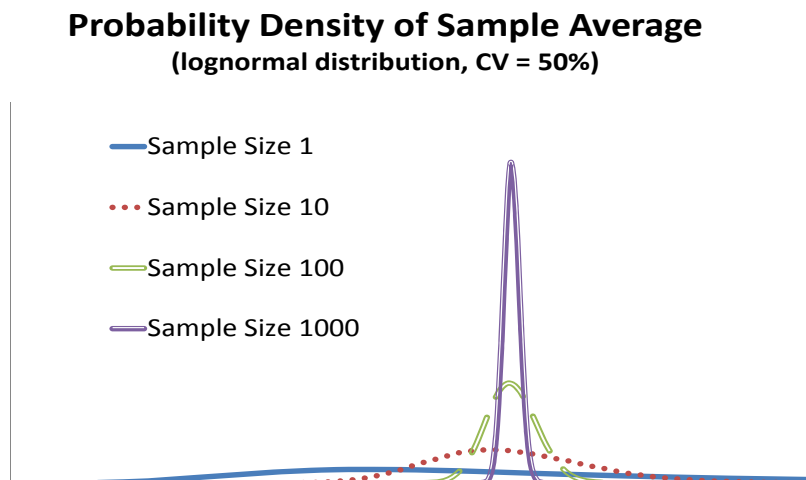
In cases where specific measures are needed to quantify how effectively the ERP is finding lift and achieving equity, the NCCI considers two statistics. The older quintile test statistic is  $B^*/A^*$ , where  $A^*$  is the variance of the un-modified quintile pure loss ratios and  $B^*$  is the variance of the modified quintile pure loss ratios: the variance in the pure loss ratios after application of the modification factors is divided by the variance in the pure loss ratios at manual rates. This traditional statistic measures equity or, perhaps more correctly, the movement towards equity with the ERP. Ideally it should be as close to zero as possible. Note that  $B^*/A^*$  would be zero for Figure 1 because  $B^*$ , the variance in the modified quintile loss ratios, is zero for this hypothetical example.

The newer quintile test statistic is  $\text{sign}(A-B)|A-B|^{0.5}$ , where  $A$  and  $B$  are equivalent to  $A^*$  and  $B^*$ , but may be calculated to include some extra variance due to bootstrapping the test, as will be discussed in Section 3.1. This newer statistic measures how effectively the ERP is at finding lift and achieving equitable workers compensation rating. Ideally it should be as large as possible. Using this new test statistic, Figure 1 trumps Figures 2 and 3 because all three have the same value for  $A$  but  $B$  takes on the smallest possible value of zero for Figure 1.

### 2.3 Sample Size and the Central Limit Theorem

The Central Limit Theorem says that the sum or average of a large number of independent random variables will be approximately Normal. Even a highly skewed distribution such as the pure loss ratio for an individual risk will tend towards a Normal distribution if an average is computed over many risks. Figure 4 illustrates this concept.

**Figure 4: The Central Limit Theorem in Action**



## Optimal Number of Quantiles For Predictive Testing

To simplify the discussion, assume that pure loss ratios for risks are independently and identically distributed with mean  $\mu$  and variance  $\sigma^2$  and that each risk has the same expected loss.  $\sigma^2/n$  is the variance and  $\sigma/\sqrt{n}$  is the standard deviation in the overall pure loss ratio for a bin containing  $n$  homogeneous risks. As can be seen in Figure 4, the decreasing variance as  $n$  increases results in a tighter curve clustered around the expected value.

The practical implication of this is that we expect for both manual and modified pure loss ratios the random observation error will be proportional to  $1/\sqrt{n}$ . Although the assumptions about homogeneity of risks are far from perfectly met in the real world, the departure will generally not cause this  $1/\sqrt{n}$  relationship to lose its practical value.

### 2.4 Systematic Differences Versus Random Variation

Individual quantiles should contain enough data so that the random variation of the manual pure loss ratios within quantiles is small compared to systematic differences in the underlying expected manual pure loss ratios between quantiles. As another hypothetical example, suppose that risks with mods uniformly distributed on the interval  $[0.80, 1.00)$  are compared to risks with mods uniformly distributed on the interval  $[1.00, 1.20)$ . Figures 5a and 5b show, for interval sample sizes 1 and 100, respectively, hypothetical distributions for manual loss ratios for risks uniformly distributed across these two intervals, assuming individual risk loss ratios are lognormally distributed with mean equal to the mod value and 0.50 standard deviation.

#### Random Outcomes versus Systematic Differences

Figure 5a

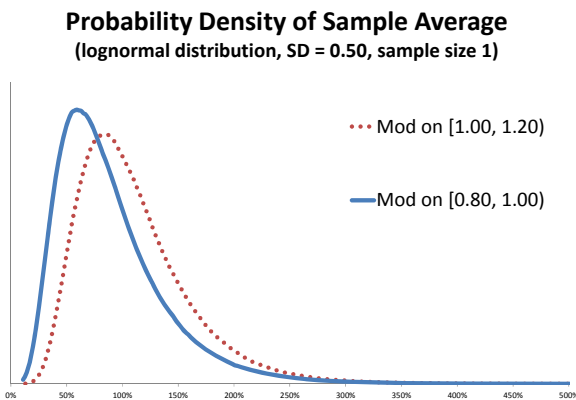
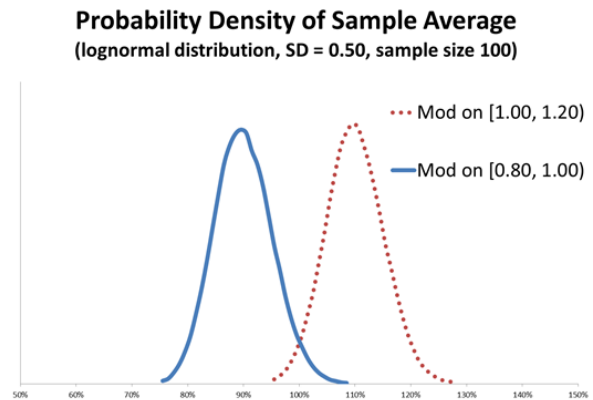


Figure 5b



### Optimal Number of Quantiles For Predictive Testing

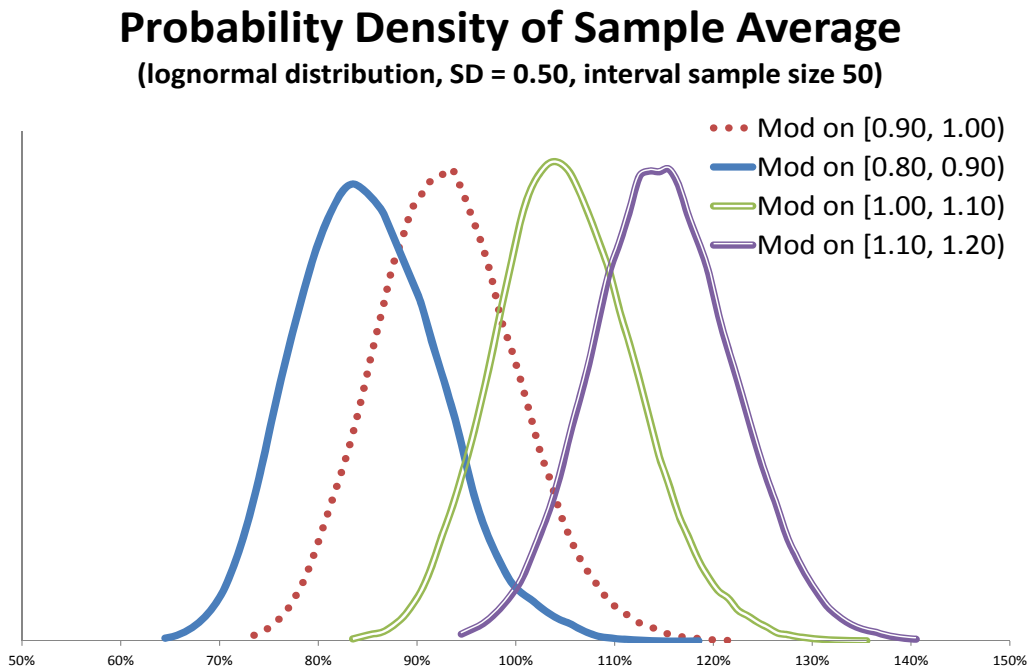
Note the substantial overlap in the distributions of actual pure loss ratios in the Figure 5a. There is a fair chance that the actual ratio for a risk from [0.80, 1.00) will exceed that for a risk from [1.00, 1.20), which would give misleading results about mod performance. One might conclude that the mod cannot effectively distinguish the loss potential of individual risks. If there are more risks in each interval, as in Figure 5b, then the distinction is almost certainly shown in the observed loss ratios.

The systematic difference between the two previous intervals can be quantified as the differences between their midpoints of 0.90 and 1.10. Assuming that there are enough risks in the intervals so that the resulting distributions are approximately Normal, standard deviations of manual loss ratios are a good way to characterize random variation. The choice of how small the random variation must be compared to the systematic difference is fairly subjective but not impractically ambiguous, as will be demonstrated in the next sections.

## 2.5 Tradeoffs in Determining the Number of Quantiles

More quantiles provide a more detailed look at how the plan is performing all along the range of mod values. However, more quantiles will increase the random variation within each quantile because a fixed number of risks will be split into fewer risks per quantile. Worse still, more quantiles will simultaneously decrease the systematic differences between adjacent quantiles.

Figure 6: Splitting Into More Intervals





## Optimal Number of Quantiles For Predictive Testing

Figure 6 shows an idealized view of the dilemma. Here, we revisit the example from Figure 5b, breaking up the two intervals, each with 100 sampled risks into four intervals, each with 50 sampled risks. First, note that we can safely ignore the contribution of variance of the mod within an interval to the variance of the loss ratio. The variance of the mod within the original 2 intervals of width 0.20 was 0.003 and for the split intervals of width 0.10 was 0.0008, trivial compared to the conditional variance of 0.25 for the loss ratio. So, there will be an increase of about 41% in the standard deviation of the overall loss ratio due to smaller sample size. Simultaneously, the distance between the midpoints of the adjacent intervals has been cut in half. Consequently, the ratio of random variation to systematic variation in Figure 6 is about 282% of what it was in Figure 5b.

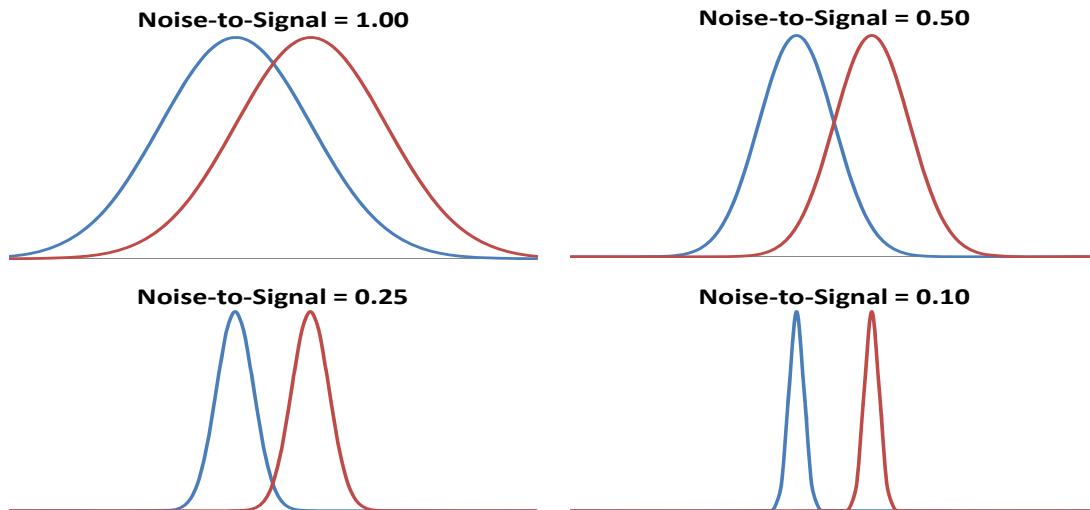
### 2.6 Noise-to-Signal Ratio

It is conceptually useful to think of random variation in observed loss ratios as *noise* and the systematic difference between the means of adjacent bins as *signal*. The ratio of random to systematic variation, that we previously alluded to, is explicitly defined and referred to as the *noise-to-signal ratio* (N/S) in (2.2).

$$N/S = \frac{\text{Standard deviation of random variations for an interval loss ratio}}{\text{Difference in expected loss ratios between intervals}} = \frac{\text{Noise}}{\text{Signal}} \quad (2.2)$$

All other things equal this ratio should be as small as possible. Figure 7 graphically illustrates the significance of noise-to-signal ratios using a Normal distribution model.

**Figure 7: Noise-to-Signal For Normal Distributions**



### *Optimal Number of Quantiles For Predictive Testing*

Table 1 shows the probabilities of reversals for a Normal distribution model. By reversal we mean that the observed loss ratio for the higher interval is actually lower than the observed loss ratio for the lower interval.

**Table 1: Noise-to-Signal Ratios and Ordering of Pure Loss Ratios**

Noise-to-Signal Ratio	Probability of Reversal
1.00	0.2398
0.50	0.0786
0.25	0.0023
0.10	$7.69 \times 10^{-13}$

A reasonable tolerance for the N/S ratio might be selected at 0.25. So, the “optimal” number of quantiles is beginning to emerge. If we can estimate the N/S ratio as a function of the number of quantiles, then the maximum number of quantiles within our N/S tolerance will be that “optimal” number.

## **2.7 The Number of Quantiles and Data Requirements**

To address the question of the number of quantiles the N/S signal ratio can be estimated by a formula in terms of the following quantities:

$b$  = number of quantiles

$\sigma^2$  = variance of the manual pure loss ratio for a single risk

$n$  = number of risks tested

$R$  = a constant corresponding to the “spread” or “variation” of mods

We will assume risks are all the same size. This is a reasonable simplifying starting assumption, but in reality risk sizes and the variances of their loss ratios span across orders of magnitude for the ERP. The resulting formula will be validated, or invalidated, for the general context using empirical data.

The standard deviation for the loss ratio in a single quantile will be  $\sigma / \sqrt{n/b} = \sqrt{b}(\sigma / \sqrt{n})$ . Again, we will ignore variation in the loss ratio due to mod difference within quantiles, as this usually contributes a very small part of the total random variation.

If there are more quantiles then the differences between the expected values for the mods of

### *Optimal Number of Quantiles For Predictive Testing*

adjacent quantiles is smaller. If the number of bins is doubled then these differences should be approximately cut in half. A reasonable rule of thumb is that the typical difference is inversely proportional to  $b$ , that is equal to  $R / b$ .

Putting the pieces above together the N/S ratio formula leads to (2.3).

$$N/S = \frac{\sqrt{b}(\sigma / \sqrt{n})}{(R/b)} = \frac{\sigma}{R} \left( \sqrt{\frac{b^3}{n}} \right) \quad (2.3)$$

Solving for sample size  $n$  leads to (2.4).

$$n = \frac{(\sigma/R)^2}{(N/S)^2} b^3 \quad (2.4)$$

Equation (2.4) shows, among other relations, that the amount of data needed increases as the cube of the number of quantiles:  $n \propto b^3$ . The ERP is currently tested using a quintile test. A decile test would require 8 times as much data for a comparable N/S ratio. Similarly, a test with twenty quantiles would require 64 times as much data.

## **3. RESULTS AND DISCUSSION**

### **3.1 Bootstrapping Empirical Data Demonstrates Effects of Number of Quantiles**

To measure the noise associated with the relative pure loss ratios in a quintile test, NCCI bootstraps the underlying data. The data set of risks used for a particular test is resampled with replacement 100 times. Each time the quintile test loss ratios are recalculated. A candlestick chart illustrates the 5th, 25th, 75th, and 95th percentiles of the bootstrapped relative pure loss ratios. The vertical spread of the quintile bars is indicative of the noise in the test and the difference in the vertical location between adjacent candles is indicative of the signal in the test. Consider the following bootstrap quintile and decile tests for policy year 2010:

- Figures 8a and 8b are countrywide, which includes 886,976 individual risks.
- Figures 9a and 9b, are countrywide for risks whose policy period expected pure loss ranges from 1k to 10k, which includes 467,887 individual risks.
- Figures 10a and 10b, are countrywide for risks whose policy period expected pure loss

*Optimal Number of Quantiles For Predictive Testing*

ranges from 100k to 1m, which includes 18,692 individual risks.

- Figures 11a and 11b, are for a large state, which includes 62,629 individual risks.
- Figures 12a and 12b, are for a small state, which includes 10,943 individual risks.

**Figure 8a: Countrywide Quintile Test**

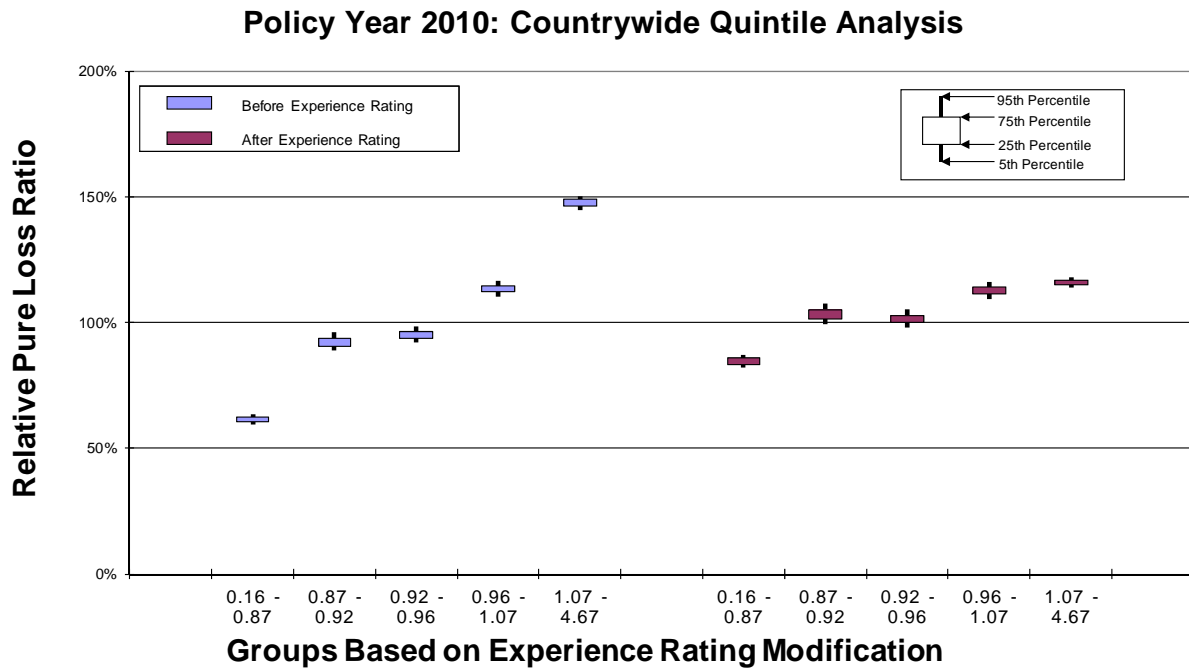
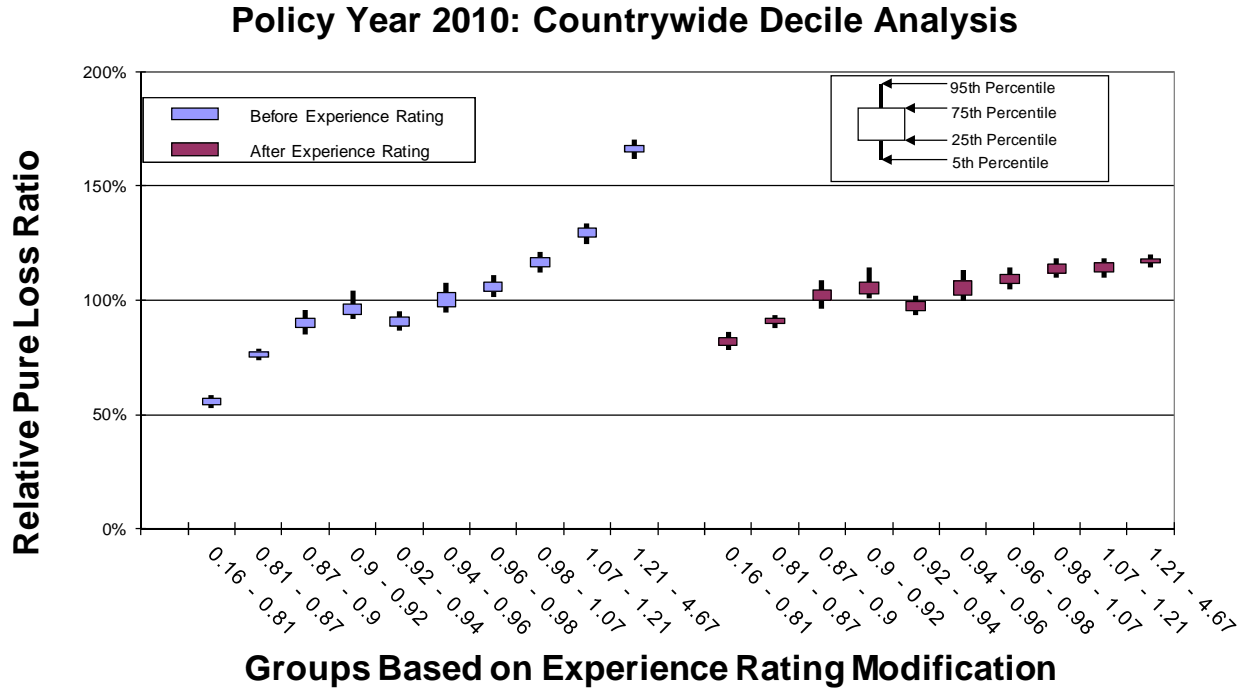


Figure 8b: Countrywide Decile Test



Figures 9a: Countrywide Small Risk Quintile Test

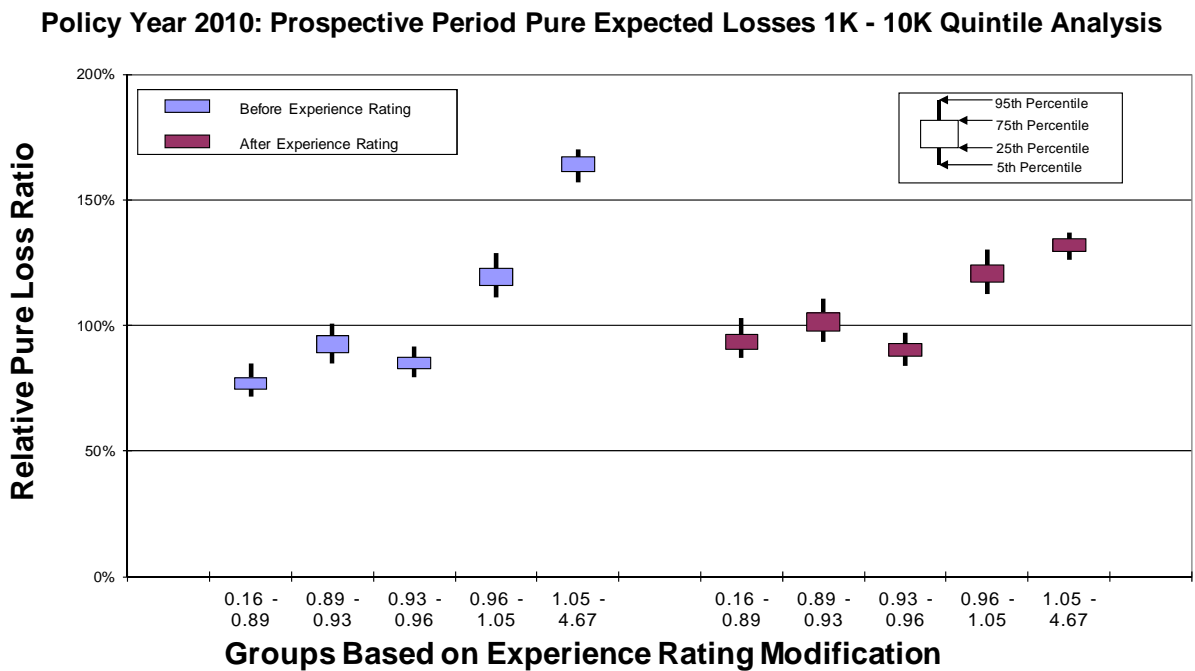
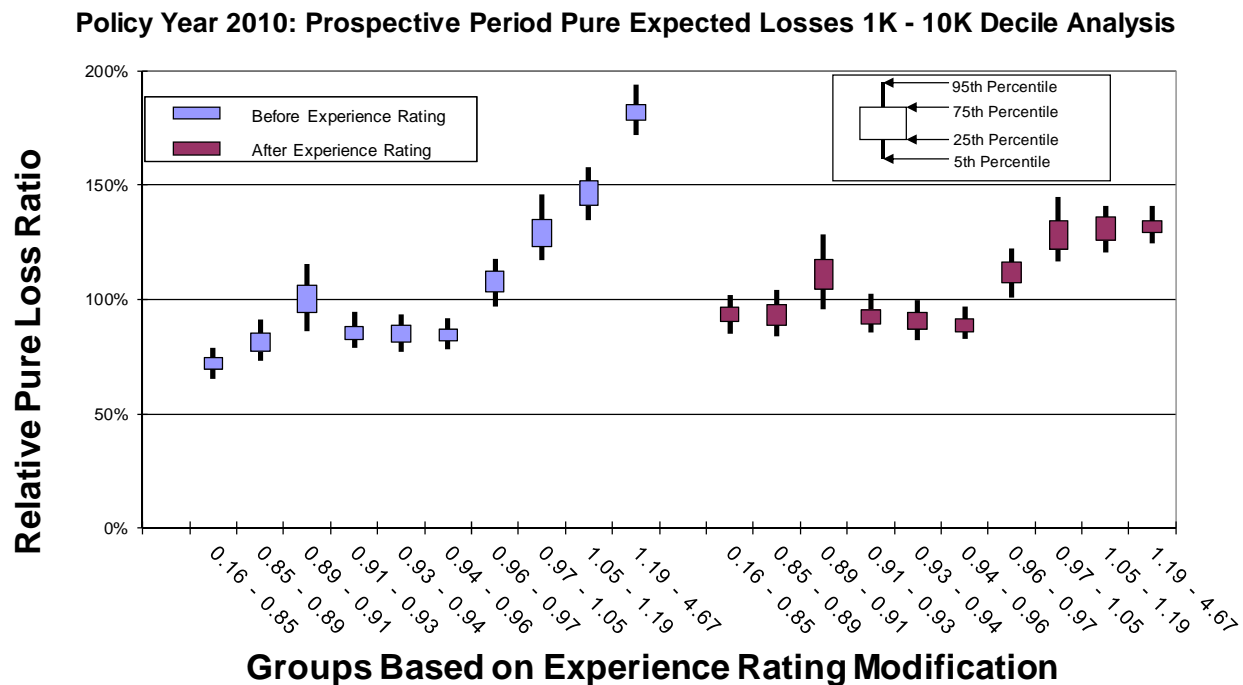


Figure 9b: Countrywide Small Risk Decile Test



Figures 10a: Countrywide Large Risk Quintile Test

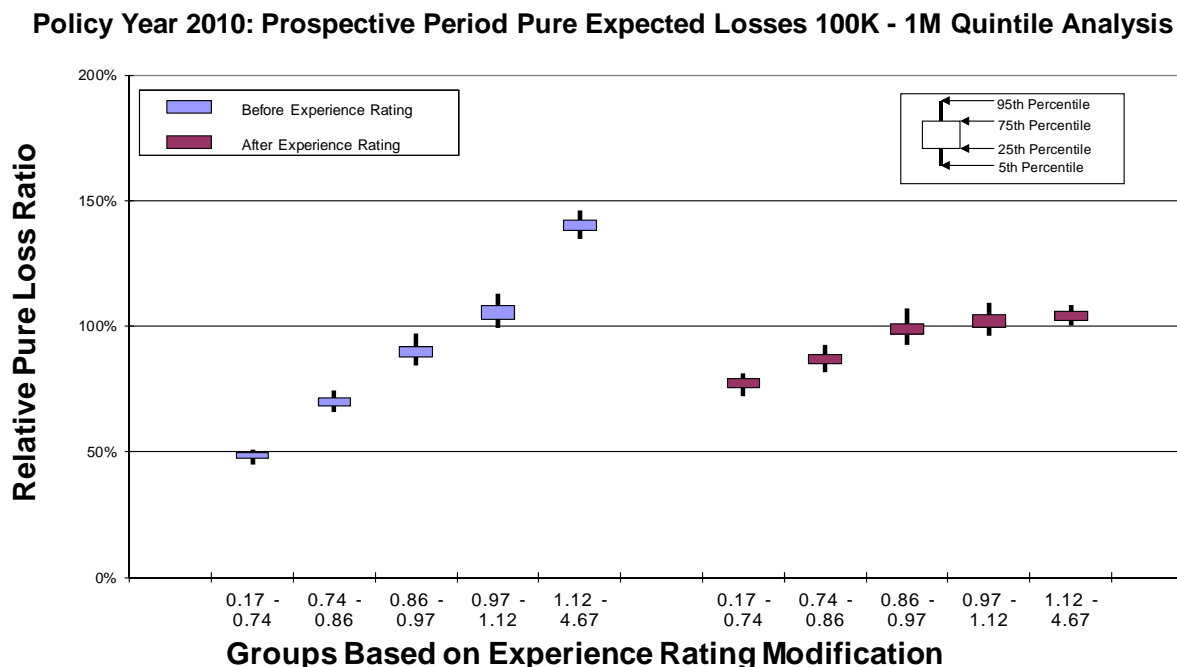
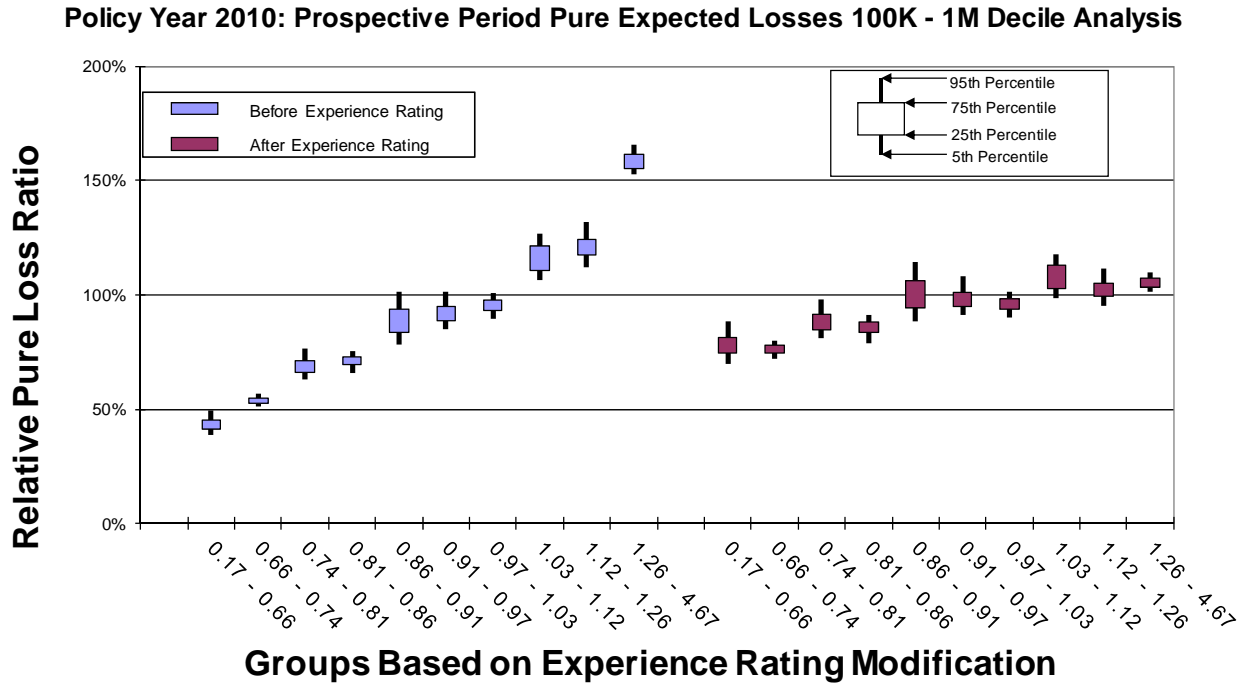


Figure 10b: Countrywide Large Risk Decile Test



Figures 11a: Large State Quintile Test

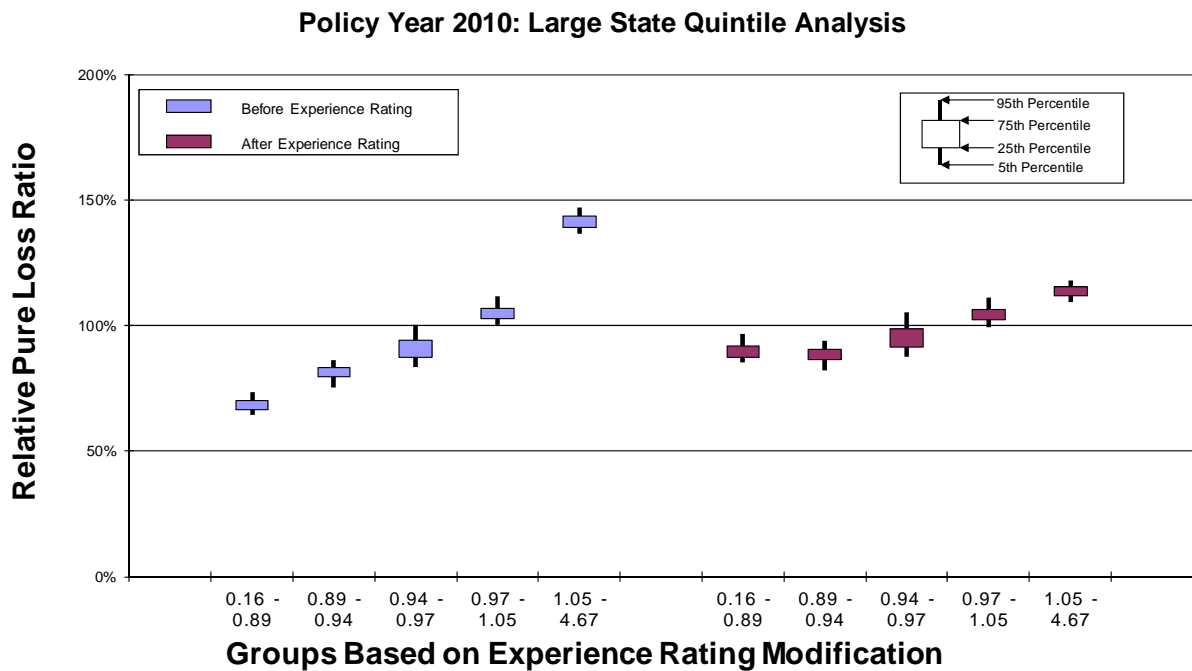
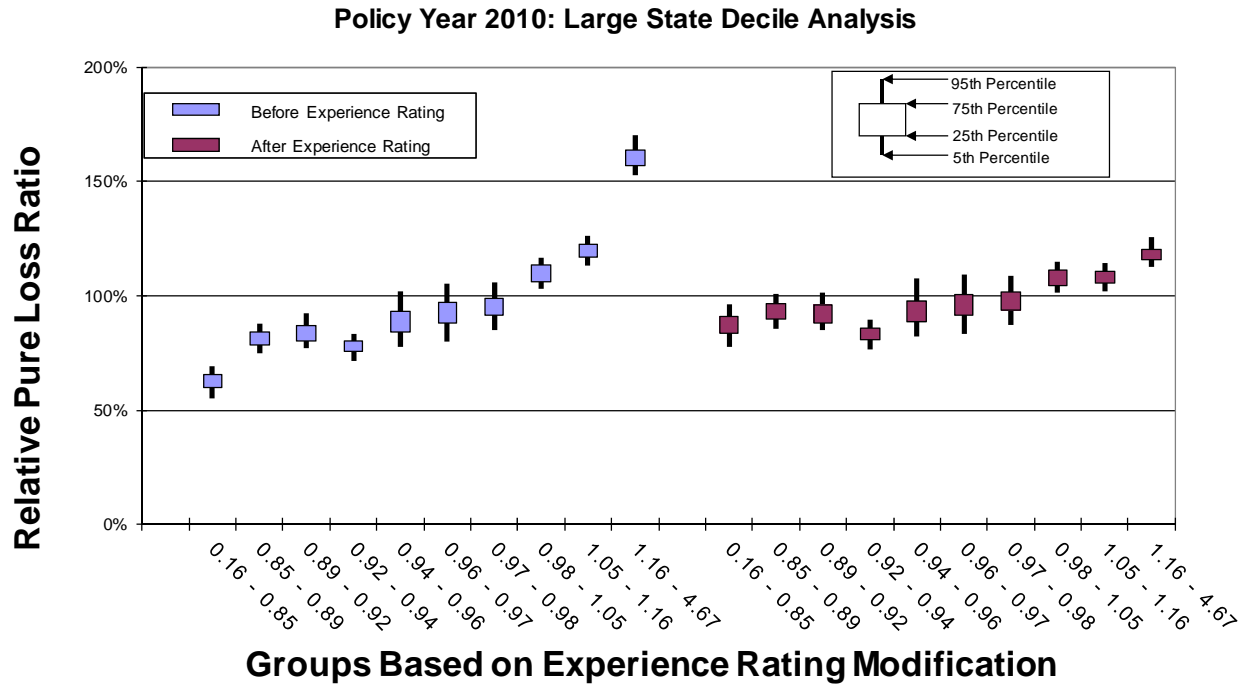


Figure 11b: Large State Decile Test



Figures 12a: Small State Quintile Test

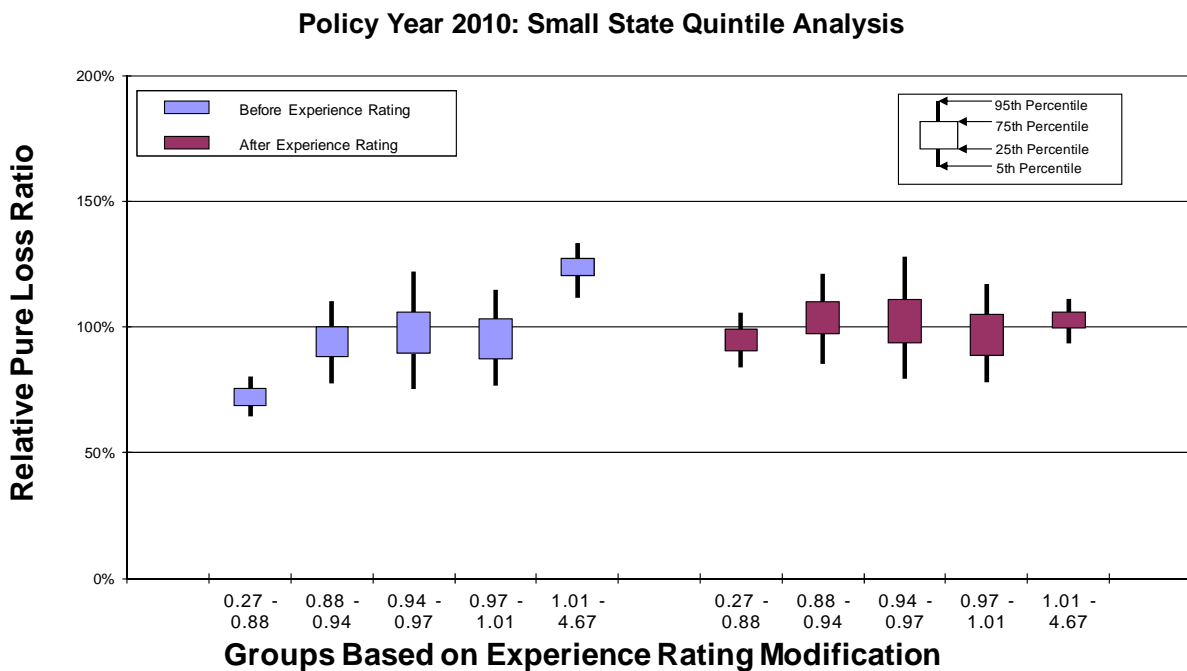
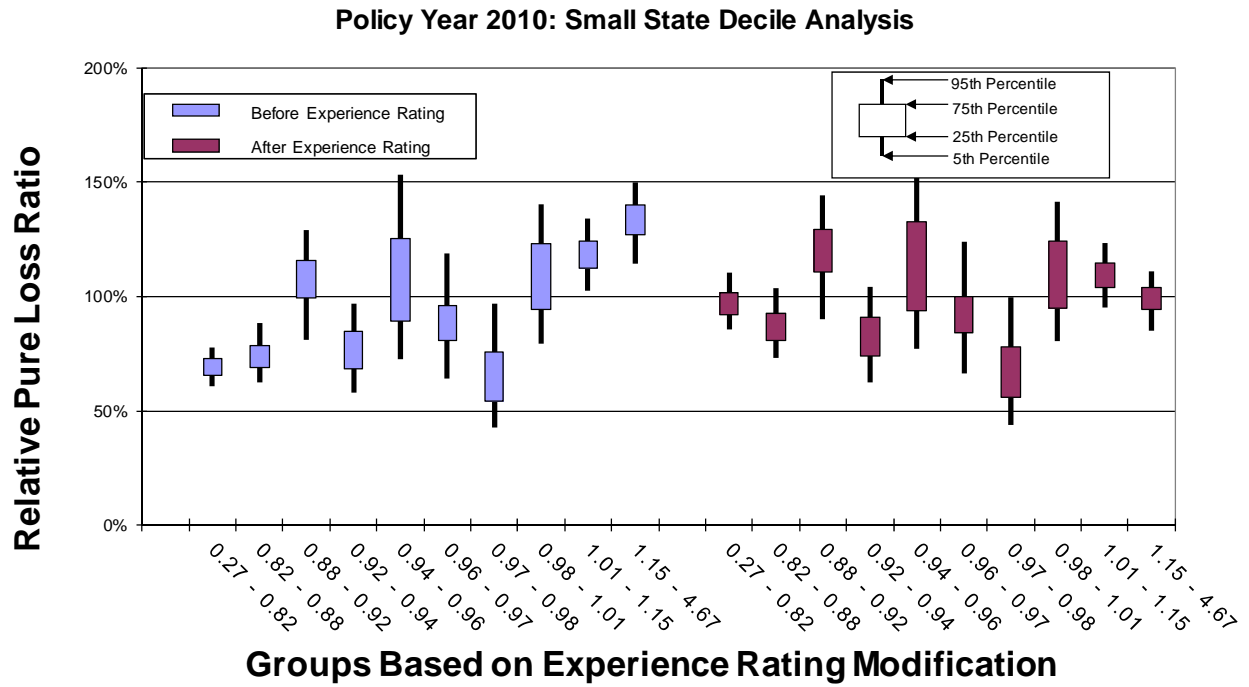




Figure 12b: Small State Decile Test



### *Optimal Number of Quantiles For Predictive Testing*

Comparison between the quintile and decile tests illustrates the three-fold increase in N/S ratio predicted by the model. Since the N/S ratio should also change in proportion to  $1/\sqrt{n}$ , where  $n$  is the number of risks, all other things equal (which they are not), we would expect the N/S to increase going from Figures 8 to Figures 11 to Figures 12. In fact, such a pattern is clearly evident. However, between tests using the same number of quantiles but different data, the random variation in the loss ratio per individual risk may vary, and will certainly decrease going from the small risks in Figures 9 to the large risks in Figures 10. Another thing that can change as the underlying data set changes is the typical signal difference between the quantiles.

The general consistency between the patterns observed in these charts and what we expect from the N/S ratio model provides a reasonable validation of the model as a practical tool.

The slight positive slopes seen in most of the empirical quintile and decile tests for policy year 2010 highlight the need for increased effective credibility that prompted NCCI's increases in the split point beginning in 2013, from \$5,000 to \$10,000 in year one, \$13,500 in year two, and \$15,000 plus two years of inflation adjustment in year three. Further routine increases based on severity indexation will follow.

## **3.2 Measuring Lift and Equity**

At the end of section 2.2, two statistics were introduced: (1)  $B^*/A^*$  (smaller is better) is a measure of how effectively the ERP achieves equity, and (2)  $\text{sign}(A-B)|A-B|^{0.5}$  (larger is better) is a combination measure that increases with more lift,  $A$  is larger, and increases with more equity,  $B$  is smaller.  $A^*$  is the variance of the un-modified pure loss ratios and  $B^*$  is the variance of the modified pure loss ratios across the quantile groups. The quantities without the  $*$  are similar quantities but may include some additional variance from bootstrapping.

Table 2 displays these statistics for the 2010 policy year data that was used to create the prior ten bootstrapping charts. The first column of numbers,  $B^*/A^*$ , shows that the ERP makes considerable progress towards achieving equity because the variance in pure loss ratios is much smaller for the modified pure loss ratios than the un-modified pure loss ratios. For example, the variance in the countrywide modified pure loss ratio quintiles is only 14.9% of the variance in the unmodified quintiles. The second column shows that the ERP is finding lift and moving towards more equitable rating.

**Table 2: Statistical Measures of Lift and Equity**

		<u>Statistics</u>	
		$B^*/A^*$	$\text{sign}(A-B) A-B ^{0.5}$
Countrywide	Quintile	0.149	0.261
Countrywide	Decile	0.140	0.265
Countrywide Small Risks	Quintile	0.265	0.272
Countrywide Small Risks	Decile	0.259	0.284
Countrywide Large Risks	Quintile	0.105	0.296
Countrywide Large Risks	Decile	0.107	0.309
Large State	Quintile	0.144	0.232
Large State	Decile	0.150	0.239
Small State	Quintile	0.044	0.157
Small State	Decile	0.474	0.155

### 3.3 Noise-to-Signal Ratio

The noise-to-signal ratio (N/S) was introduced in equation (2.2) of section 2.6.

$$N/S = \frac{\text{Standard deviation of random variations for an interval loss ratio}}{\text{Difference in expected loss ratios between intervals}} = \frac{\text{Noise}}{\text{Signal}} \quad (2.2)$$

Although an explicit formula for calculating noise-to-signal ratios was presented in equation (2.3), the ratios can also be estimated by bootstrapping empirical data as done in section 3.1. Noise-to-signal ratios corresponding to the bootstrapping results used to create the charts in section 3.1 are displayed in Table 3. As expected, the noise-to-signal ratios are significantly higher for deciles than quintiles. The last column of the table displays the ratio of noise-to-signal ratios for deciles to those of quintiles.

Equation (2.3) shows that  $N/S \propto \sqrt{b^3}$  where  $b$  is the number of quantiles. Comparing deciles and quintiles with this proportionality yields a value not too far from those in the last column of Table 3.

$$N/S \text{ Deciles} \div N/S \text{ Quintiles} = \sqrt{10^3} \div \sqrt{5^3} = \sqrt{2^3} = 2.83 .$$

The theoretical value 2.83 is somewhat greater than values in the table because the experience

modification factors spread out in the two tails of the mod distribution. This would not happen if the mods were uniformly distributed over a range.

**Table 3: Bootstrapped Estimates of Noise-to-Signal Ratios**

	Noise - to - Signal		N/S Decile / N/S Quintile
	Quintile	Decile	
Countrywide (C/W)	0.086	0.229	2.67
C/W \$1K - \$10K	0.203	0.499	2.46
C/W \$100K - \$1,000K	0.128	0.354	2.75
Large State	0.192	0.482	2.51
Small State	0.748	1.973	2.64

### 3.4 Some Hypothetical Examples

Appendix A develops the rule of thumb, shown in (3.1), to estimate  $\sigma/R$  in terms of the credibility  $Z$  for an individual risk loss ratio. This is particularly useful, as  $Z$  can act as a reasonable measure of risk size, though not on a proportional scale.

$$\frac{\sigma}{R} = \sqrt{\frac{1/Z^2 - 1}{12}} \tag{3.1}$$

However, before progressing too far we must recall that in principle a  $Z$  credibility value such as this, although it includes useful information about  $\sigma/R$ , is qualitatively very different from the  $Z_p$  and  $Z_e$  values found in the experience rating. This  $Z$  value would:

1. Estimate the true underlying mean for a single policy year using that policy year's losses, rather than using prior policy years' experience to predict the true mean for a future policy year,
2. Cover all losses unlimited, and
3. Use only one year of losses.

So, much caution must be used in attempting to apply it to the very different situation of the

### Optimal Number of Quantiles For Predictive Testing

ERP. Unfortunately, there is no clear way to relate  $Z$  to the readily available, but fundamentally very different,  $Z_p$  and  $Z_e$ . More generally, there is probably no good way to determine such a  $Z$  that would directly apply to the context of the NCCI ERP. Statement 1 suggests  $Z$  should be higher than  $Z_p$ , but 2 and 3 suggest  $Z$  should lower than  $Z_p$ . Since higher  $Z$  implies lower  $\sigma/R$  and hence lower  $N/S$  it is prudent to err on the side of a lower  $Z$ . So, on balance, it is not unreasonable to speculate that  $Z$  should be on the same order as  $Z_p$ , but somewhat lower, perhaps  $Z \approx Z_p / 2$ . Ultimately, any  $N/S$  model resulting from whatever relationship we may assume can be validated, or invalidated, using empirical data along the lines we followed in Section 3.1. Table 4 explores what different values for  $Z$ ,  $n$ , and  $b$  imply for the  $N/S$  ratio.

**Table 4: Noise-to-Signal By Z Credibility**

Noise to Signal  $\leq 0.25$  shaded

Z Credibility = 50%						Z Credibility = 25%					
Quantiles	Sample Size-->					Quantiles	Sample Size-->				
	100	1,000	10,000	100,000	1,000,000		100	1,000	10,000	100,000	1,000,000
2	0.14	0.04	0.01	0.00	0.00	2	0.32	0.10	0.03	0.01	0.00
5	0.56	0.18	0.06	0.02	0.01	5	1.25	0.40	0.13	0.04	0.01
10	1.58	0.50	0.16	0.05	0.02	10	3.54	1.12	0.35	0.11	0.04
20	4.47	1.41	0.45	0.14	0.04	20	10.00	3.16	1.00	0.32	0.10
100	50.00	15.81	5.00	1.58	0.50	100	111.80	35.36	11.18	3.54	1.12

Z Credibility = 10%						Z Credibility = 5%					
Quantiles	Sample Size-->					Quantiles	Sample Size-->				
	100	1,000	10,000	100,000	1,000,000		100	1,000	10,000	100,000	1,000,000
2	0.81	0.26	0.08	0.03	0.01	2	1.63	0.52	0.16	0.05	0.02
5	3.21	1.02	0.32	0.10	0.03	5	6.45	2.04	0.64	0.20	0.06
10	9.08	2.87	0.91	0.29	0.09	10	18.23	5.77	1.82	0.58	0.18
20	25.69	8.12	2.57	0.81	0.26	20	51.58	16.31	5.16	1.63	0.52
100	287.23	90.83	28.72	9.08	2.87	100	576.63	182.35	57.66	18.23	5.77

The formulas for  $Z_e$  and  $Z_p$  vary by state and over time, according to a fairly complicated formula that is a function of experience period expected losses, rather than policy period pure expected losses. However, the values in Table 4 would be roughly in the ballpark for recent years.

The conjecture  $Z \approx Z_p / 2$  together with Table 5 suggest that  $Z$  would be around 3% to 20% for Figures 9 and around 40% to 45% for Figures 10.

**Table 5: Approximate Recent ERP Credibility By Risk Size**

Policy Period Pure Premium	Approximate Recent ERP Credibility	
	$Z_p$	$Z_e$
1,000	5%	0%
10,000	40%	2%
100,000	80%	10%
1,000,000	90%	40%

Next recall that Figure 9 had a sample size approaching half a million risks and Figure 10 had a sample size a bit under 20,000. From Table 4 we can see that both Figures 9 and 10 would be expected to be well within the acceptable N/S for the quintile. The deciles test would be expected to be around the maximum N/S ratio. This is roughly the situation we see in the charts and therefore  $Z \approx Z_p / 2$ , although not founded in any sort of solid mathematics or logic, is still empirically validated to be in the ballpark and therefore practically useful within broad limits.

#### 4. CONCLUSIONS

The number of quantiles selected for a meaningful test of predictive performance of the NCCI Experience Rating Plan is constrained by the ratio of noise-to-signal. If this ratio is not kept below a reasonable threshold, which is subjective but can be sensibly selected somewhere in the neighborhood of 0.25, the results of the quantile test will be unclear. In a sense the “optimal” number of quantiles is the largest number that produces a N/S ratio under this threshold. The N/S ratio is proportional to the standard deviation of observed loss ratios for individual risks and the 1.5 power of the number of quantiles. The N/S ratio is inversely proportional to the variation in mod values and the square root of the number of risks in the data. Consequently, the data required to maintain a given N/S ratio is proportional to the cubic power of the number of quantiles. This huge data penalty in test resolution for increasing the number of quantiles explains the use of a relatively small number of quantiles, exactly five in the quintile test, for testing the NCCI Experience Rating Plan. These and other implications of this N/S ratio can be demonstrated consistently for both empirical tests of the ERP and hypothetical examples.

## Appendix A

### A Connection Between Noise-to-Signal Ratio and Credibility

Suppose a single credibility  $Z$  value, which can be any value in (0%, 100%) not necessarily based on any sound credibility model, is used on the total unlimited policy year loss ratio  $L$  to estimate a non-predictive individual risk modification factor. That is, a risk's actual loss ratio in a single year is used to estimate its own true mean loss ratio for that same year. The sample variance in such a mod would be (A.1).

$$\begin{aligned} & \frac{1}{n-1} \sum_{i=1}^n [Z L_i + (1-Z)\bar{L} - (Z\bar{L} + (1-Z)\bar{L})]^2 \\ &= \frac{Z^2}{n-1} \sum_{i=1}^n (L_i - \bar{L})^2 \\ &= Z^2 \hat{\sigma}_L^2 \end{aligned} \tag{A.1}$$

where  $\hat{\sigma}_L^2$  is the actual sample variance of the observed loss ratios. The remaining or “random” part of the sample variance is (A.2).

$$(1-Z^2) \hat{\sigma}_L^2 \tag{A.2}$$

Although a quantile test would be meaningless in this non-predictive situation, we can estimate a ratio  $\sigma/R$  in the N/S formula (2.3). (A.2) can be used as an estimate of  $\sigma^2$  and (A.1) can be used to derive an estimate for  $R$ . As a simplification, assume that mods are effectively uniformly distributed across an interval of width  $R$ . For a uniform distribution the difference between the means of mods in adjacent quantiles, or signal, where mods are split into  $b$  quantiles would be  $R/b$ . This is the key property of  $R$  as used in (2.3). Equating the variance expression for a uniform distribution with (A.1) leads to (A.3).

$$\frac{R^2}{12} = Z^2 \hat{\sigma}_L^2 \tag{A.3}$$

This would lead to (A.4).

*Optimal Number of Quantiles For Predictive Testing*

$$\frac{\sigma}{R} = \frac{\sqrt{(1-Z^2) \hat{\sigma}_L^2}}{\sqrt{12Z^2 \hat{\sigma}_L^2}} = \sqrt{\frac{1/Z^2 - 1}{12}} \quad (\text{A.4})$$

Assuming a uniform distribution of mods (A.4) corresponds implicitly to (A.5).

$$R = \sqrt{12} \sigma_{Mod} \quad (\text{A.5})$$

(A.5) will certainly not hold true in general, but it is worth noting how it compares to a decile test of hypothetical mods lognormally distributed around mean 1.00 with standard deviation 0.10, as shown in Table A1. (A.5) predicts that R would be 0.346 and the difference between adjacent quantiles would be 0.035 in this situation.

**Table A1: Differences between decile means for a lognormal distribution with mean 1.00 and standard deviation 0.10**

Decile	Mean	Difference
1st	0.836	
2nd	0.897	0.061
3rd	0.930	0.033
4th	0.957	0.027
5th	0.983	0.025
6th	1.008	0.025
7th	1.034	0.027
8th	1.065	0.030
9th	1.104	0.040
10th	1.186	0.082



## 5. REFERENCES

- [1] Bailey, Robert A., "Experience Rating Reassessed," *PCAS* XLVIII, 1961, p. 60-82.
- [2] Bailey, Robert A.; and Simon, Leroy, J., "An Actuarial Note on the Credibility of a Single Private Passenger Car," *PCAS* XLVI, 1959.
- [3] Bühlmann, Hans, "Experience Rating and Credibility," *ASTIN Bulletin*, Vol. IV, Part III.
- [4] Dorweiler, Paul, "A Survey of Risk Credibility in Experience Rating," *PCAS* XXI, 1934, p. 1.
- [5] Fisher, Arne, "Note on an Application of Bayes' Rule in the Classification of Hazards in Experience Rating," *PCAS* III, 1916, p. 43-48.
- [6] Gillam, William R., "Parametrizing the Workers Compensation Experience Rating Plan," *PCAS* LXXIX, 1992, p. 21-56.
- [7] Gillam, W.R.; and Snader, R.H., "Fundamentals of Individual Risk Rating," National Council on Compensation Insurance, (Study Note), 1992, Part I.
- [8] Gillam, William R., "Workers' Compensation Experience Rating: What Every Actuary Should Know," *PCAS* LXXIX, 1992, p. 215-239.
- [9] Mahler, Howard C., "An Introduction to Basic Credibility," *CAS Forum*, 1997, Winter, pp. 67-103.
- [10] Mahler, Howard C., "A Graphical Illustration of Experience Rating Credibilities," *PCAS* LXXXV, 1998, p. 654.
- [11] Mahler, Howard C., and Dean, Curtis G., "Credibility," Chapter 8 in *Foundations of Casualty Actuarial Science*, 4<sup>th</sup> Edition, Arlington, Va.: Casualty Actuarial Society, 2001.
- [12] Meyers, Glenn G., "An Analysis of Experience Rating," *PCAS* LXXII, 1985, p. 278.
- [13] National Council on Compensation Insurance, *The 1998 Adjustment to the Experience Rating Plan: Your Guide to Understanding the Changes*.
- [14] National Council on Compensation Insurance, *Individual Risk Rating Working Group Agenda; Individual Risk Rating Working Group Minutes*, August 2006 through February 2011.
- [15] National Council on Compensation Insurance, *Experience Rating Plan Manual for Workers Compensation and Employers Liability Insurance*.
- [16] Perryman, F. S., "Experience Rating Plan Credibilities," *PCAS* XXIV, 1937, p. 60.
- [17] Venter, G.G., "Experience Rating—Equity and Predictive Accuracy," *NCCI Digest*, April 1987, Volume II, Issue I, pp. 27-35.

### Abbreviations and notations

- C/W, countrywide
- ERP, experience rating plan
- Mod, experience modification factor
- NCCI, National Council on Compensation Insurance
- N/S, noise-to-signal ratio
- $A, A^*$ , variance of un-modified quintile pure loss ratios
- $B, B^*$ , variance of modified quintile pure loss ratios
- $b$ , number of quantiles
- $L$ , unlimited policy year loss ratio for a risk
- $n$ , number of risks
- $R$ , spread in modification factor values
- $\sigma^2$ , variance in pure loss ratio for one risk
- $\hat{\sigma}_L^2$ , actual sample variance of observed loss ratios
- $Z_p$ , primary credibility for an experience rated risk
- $Z_e$ , excess credibility for an experience rated risk
- $Z_i$ , credibility for an individual risk

## *Optimal Number of Quantiles For Predictive Testing*

### **Biographies of the Authors**

**Jonathan Evans, FCAS, FSA, FCA, CERA, MAAA, WCP** is an actuary at The National Council on Compensation Insurance in Boca Raton, FL. His work primarily involves research and development of NCCI's ratemaking, reserving, and catastrophe modeling procedures.

**Curtis Gary Dean, FCAS, MAAA** has been a Distinguished Professor of Actuarial Science at Ball State University in Muncie, IN since 2001. Previously, he worked 26 years as an actuary at American States, later acquired by Safeco, in Indianapolis, IN. While on leave from Ball State he oversaw the creation of the Commercial Lines predictive modeling unit at Travelers in Hartford, CT.