DISCUSSION OF PAPER PUBLISHED IN VOLUME LXXXIII

THE COMPLEMENT OF CREDIBILITY

JOSEPH A. BOOR

DISCUSSION BY SHOLOM FELDBLUM

> It is the concept of credibility that has been the casualty actuaries' most important and most enduring contribution to casualty actuarial science.
>
> —Matthew Rodermund

1. INTRODUCTION

Credibility theory is the crown jewel of casualty actuarial science. The statistician measures the significance of empirical findings, and the businessman uses judgment to select among diverse recommendations. Credibility theory meshes these two traditions, enabling us to combine varied indications based upon the relative predictive power of each of them.

In the quotation above, Rodermund speaks of credibility theory itself. Joseph Boor reminds us that the determination of credibility is only half of the pricing actuary's task. The other half, of no less importance, is to choose the figure that receives the complement of credibility.

Boor focuses on the practical track. There are dozens of fine actuarial papers on the theory of credibility, not all of which are easily implemented by the practicing actuary. Boor's paper, in contrast, was written first for the actuarial student, as a study note for the CAS ratemaking examination. It is equally valuable for the experienced actuary, keeping our gaze focused on the practicalities of insurance pricing.

2. BAYESIAN VS. CLASSICAL CREDIBILITY

It is instructive to contrast Bayesian and classical credibility procedures in the light of Boor's paper. Classical credibility assigns a credibility value to the experience data based upon its predictive power: that is, based upon the probability that the indication derived from the historical experience will be relatively close to the true expectations (see Longley-Cook [5]). The term "indication" here refers to the claim frequency, the pure premium, the loss ratio, or any similar ratemaking figure. Following Boor's illustrations, this discussion also uses examples of pure premiums.

In the classical tradition, the credibility assigned to the experience data is independent of the figure that is accorded the complement of credibility. Indeed the qualities of the figure that is assigned the complement of credibility are not relevant for determining the credibility that it receives, and they are therefore not relevant for determining the final credibility-weighted indication.

For instance, traditional automobile liability ratemaking procedures may assign a credibility value to the experience loss ratio or the experience pure premium based upon the number of claims in the experience period, using a full credibility standard of 1,024 claims.[1] For experience with fewer than 1,024 claims, the credibility assigned to the experience data equals $\sqrt{N/1024}$, where N is the number of claims in the experience.[2] If the experience contains 164 claims, then the credibility formula requires

[1]The rationale for this approach stems from a normal approximation to a Poisson claim distribution, whereby 1,024 claims in the experience period with no subsequent changes in the parameters of the Poisson distribution provides a 95% confidence interval that the true claim frequency is within $\pm 5\%$ of the indicated claim frequency (Stern [8]; Longley-Cook [5]).

[2]The rationale for this partial credibility formula is that the variance of the experience indications varies with the square root of the volume of the experience. It is difficult to construct a partial credibility rule from this rationale, since classical credibility does not consider the predictive accuracy of the information which receives the complement of credibility.

that we assign 60% weight to some other information, such as the current pure premium, or the pure premium from some other classes, some other territories, or some other time periods. Boor asks: "How should we choose this other information? In particular, what characteristics should this other information have?"

We will come back to this question in a moment, to see whether it is indeed well formulated. Let us first consider the workings of classical credibility. Most fundamentally, the attributes of the information that receives the complement of credibility do not affect the amount of credibility to be assigned either to this data or to the experience data. Whether we use the current pure premium, or the pure premium from some other classes, some other territories, or some other time periods makes no difference. The experience data still receives 40% credibility, and the other information receives the remaining 60% credibility.

In contrast, Bayesian credibility procedures do not speak about the predictive power of the experience data. In Bayesian credibility, the experience data in one ratemaking scenario may be sparse and volatile, but they will be assigned high credibility simply because we have nothing else that is more accurate. In another ratemaking scenario, the experience data may be voluminous and steady, but they may be assigned a lower credibility because we have other, equally good information.

If Boor's paper is important for classical credibility, then it is doubly important for Bayesian credibility. In classical credibility, the determination of credibility comes first. Before we have chosen the data to be assigned the complement of credibility, we know the amount of credibility that will be assigned to the experience data. A proper choice of the data to be assigned the complement of credibility improves the ratemaking indication only because it improves the part of the indication stemming from the complement of credibility.

In Bayesian credibility, we do not know the amount of credibility to be assigned the experience data until we know the type of information to which the complement of credibility will be assigned. The attributes of this information, such as its predictive power and its freedom from bias, affect the weight that we assign both to this data and to the experience data.

Many actuaries conceive of classical credibility and Bayesian credibility as two points along a continuum. Actuarial ratemaking seeks to produce the most accurate indications possible. Classical credibility was an early attempt to achieve this, and it remains the most practical technique in most ratemaking environments. Bayesian credibility is a more refined method of achieving the same end. To summarize: Bayesian credibility is a statistically justified procedure for optimizing the accuracy of the rate indications. Classical credibility is an early, less sophisticated attempt to do the same.

3. LEAST FLUCTUATION CREDIBILITY AND GREATEST ACCURACY CREDIBILITY

Gary Venter [9] has put forth an alternative perspective. The aim of classical credibility is not solely the achievement of accurate rate indications. Rather, the aim of classical credibility is to limit the fluctuation of rate levels from year to year, unless there is good statistical justification for a change. As Venter says [9, pp. 383, 384]:

> The basic philosophical difference between these methods is as follows. The limited fluctuation approach aims to limit the effect that random fluctuations in the data can have on the estimate; the greatest accuracy approach attempts to make the estimation errors as small as possible. The most well developed approach to greatest accuracy credibility is *least squares credibility*, which seeks to minimize the expected value of the square of the estimation error. The term "classical credibility" has sometimes been used in North America to denote limited fluctuation credibility …

Let us clarify this distinction with an illustration. Suppose that we are making rates for a new coverage, and we have no historical experience and no prior expectations upon which to base the rate. It is a stated amount coverage for $1,000, with a maximum of one claim per policy period, so we know that the pure premium must be between $0 and $1,000 (depending on the claim frequency).[3]

Since there is no statistical information or prior expectations for this coverage, let us suppose that the regulator chooses a random number between $0 and $1,000 as the pure premium for the first five years. The random number is $670. We add risk load, expenses, and profit margins to set the rate.

After the five years go by, we have some historical experience, which indicates the pure premium should be $245. However, the historical experience is sparse, and the true expected pure premium may be much different. We want to use this experience and credibility theory to set the pure premium for the next five years.

Assume that we have no prior expectation and no external information. The only information we have is the $245 historical pure premium. Under Bayesian credibility theory, the historical experience gets full credibility, and our best estimate for the expected pure premium is $245.[4]

Classical credibility is designed to avoid undue fluctuations in the rates. Both the customer and the regulator are accustomed to a pure premium of $670. Because the data are sparse, the rate level will change significantly if we rely on Bayesian credibility. Classical credibility will change the rate only to the extent that we have "credible" historical experience for doing so.

---

[3]For example, consider term life insurance coverage on an insured whose mortality expectation is entirely unknown.

[4]We may want to set a higher risk load, since there is a greater chance that we will lose money with lower rates. However, the risk load is distinct from the expected pure premium.

Suppose that we determine that the classical credibility is 40%, based on the statistical parameters which we choose, such as the size of the confidence interval. The new pure premium is $40\% \times \$245 + 60\% \times \$670 = \$500$.

With Bayesian credibility, the indicated pure premium moves as rapidly as possible towards the true expected pure premium, though the sequence of pure premium changes may have wide swings. With classical credibility, the indication moves less rapidly towards the true expected pure premium, but the sequence of pure premium changes has fewer and narrower swings. Bayesian credibility emphasizes accuracy; classical credibility emphasizes stability.

This example is admittedly extreme. In general, we have *a priori* expectations for the pure premium, and the previously charged rate is rarely so different from the experience data. However, this distinction between classical and Bayesian credibility is true for any ratemaking scenario, though the differences between the two methods will rarely be as great.

4. THE COMPLEMENT OF CREDIBILITY: CLASSICAL APPROACH

The implications for Boor's thesis are important, assuming that we interpret the distinction between classical and Bayesian credibility in the manner proposed by Venter. For classical credibility, the primary concern is limiting the fluctuations in rates. If so, the choice of the information which should be accorded the complement of credibility is clear. It is the current rate, adjusted (if necessary) for factors unrelated to potentially random fluctuations in experience.

Since this is the essence of classical credibility, it warrants further explanation. Suppose that

- the underlying pure premium is \$100 per exposure,
- there is no monetary inflation affecting claims costs (i.e., no loss cost trend),

- there is no expected change in claim frequency, and
- there are no changes in the compensation system that might affect loss costs.

Random loss occurrences, however, affect the experience pure premiums. Sometimes the experience data indicate a $90 pure premium, and sometimes they indicate a $110 pure premium. We don't know whether the true expected pure premium per exposure is $90, $100, $110, or some other amount. Unless there is good actuarial justification for doing so, the insurance company will not change the underlying pure premium. To the extent that the experience is "credible," the company will indeed change the rate to bring it more in line with the historical experience.

Without reliable experience indications, the company and the regulator are reluctant to change rates because (i) the public has come to expect a $100 pure premium and (ii) there is no good actuarial justification for changing the rates. There may be some external factor affecting the expected pure premium. In that case, the pricing actuary aims to reflect that factor in the price change. For instance, if there is 10% monetary inflation affecting loss costs (i.e., the loss cost trend is +10%), then the company, the regulator, and the public expect a 10% increase in premium rates if we have no other information. This is the rationale for credibility weighting the experience pure premium (or experience loss ratio) not with the underlying pure premium (or the expected loss ratio) but with the *trended* underlying pure premium (or the trended expected loss ratio).

The same is true for any other external change affecting the expected loss costs, such as changes in the expected claim frequency, or changes in the insurance compensation system. In practice, these factors affect both the experience data and the underlying pure premium (or the expected loss ratio). For instance, if there is a non-zero loss cost trend, the trend factor is applied both to the experience data and to the underlying pure premium.

5. THE COMPLEMENT OF CREDIBILITY: IN PRACTICE

The central question of Boor's paper is "What are the desirable characteristics of the information to be assigned the complement of credibility?" For classical credibility, we have answered this question. If the goal is limited fluctuation, then the information to be assigned the complement of credibility should be the current rate (or the current pure premium, or the expected loss ratio), adjusted for all factors other than the uncertainty inherent in the insurance process.

This is indeed what is done in most primary lines of business. In automobile liability, for instance, the experience loss ratio is credibility weighted with the expected loss ratio, adjusted (if necessary) for loss cost trends. Similarly, the indicated territorial or classification rate relativity is credibility weighted with the current territorial or classification rate relativity.

Boor's illustration of Harwayne's method of determining workers compensation pure premiums is particularly instructive, since it demonstrates numerous aspects of good ratemaking technique. In Harwayne's method, there are three components given credibility weights in calculating the pure premium (see Harwayne [4]):

1. the indicated pure premium,
2. the national pure premium, and
3. the underlying pure premium.

The second component of the formula, the national pure premium, reflects greatest accuracy credibility. The third component of the formula, the underlying pure premium, reflects limited fluctuation credibility. We discuss in Section 9 the rationale for this rate making procedure, as well as the adjustments made to the national pure premium for the greatest accuracy component. For now, let it suffice to say that the underlying pure premium is adjusted for all external influences, as described above.

General liability has a more complex procedure for combining the experience pure premium (or loss ratio) with the underlying pure premium (or expected loss ratio). This procedure, termed the "C-factor" by Graves and Castillo [3], was introduced by the Insurance Services Office in the 1980s, and it is illustrated in Boor's paper.[5]

The procedure looks like limited fluctuation credibility, but its rationale is different. The desirable characteristics of the information receiving the complement of credibility follow directly from the rationale of this credibility procedure.

Boor's central concern is to determine the desirable characteristics of the information receiving the complement of credibility. Boor lists these characteristics at the beginning of his paper. He then discusses several commonly used credibility procedures, and he discusses how well each one measures up to these characteristics.

The primary purpose of this discussion is to show how the desirable characteristics of the information receiving the complement of credibility follow from the rationale of the credibility procedure. We do this for each separate use of credibility: limited fluctuation, proxy for past experience, greatest accuracy, and marketplace pricing. The results sometimes agree with the conclusion in Boor's paper, and sometimes they expand on them. Keep in mind this theme: if we wish to determine the desirable characteristics of the information receiving the complement of credibility, we must know why we are using credibility in the first place.

Let us illustrate the C-factor procedure, so that its workings are clear. We then differentiate it from limited fluctuation cred-

[5]Graves and Castillo, who use a loss ratio ratemaking procedure, credibility weight the experience loss ratio with the trended and adjusted expected loss ratio. Boor uses the same procedure, though he credibility weights the pure premium with a trended and adjusted underlying pure premium.

ibility, and we re-examine Boor's central question: "What are the desirable characteristics of the information that receives the complement of credibility?"

For the sake of clarity, let us simplify the illustration by ignoring the time lags that are needed for data collection and rate filings. Suppose that we are making rates for a policy to be effective on January 1, 1999, using experience from accident year 1998. (For simplicity, we are making rates for a single policy, not for a policy year. Were we making rates for a policy year, we would have an additional half year of trend in the mathematics below.) The current pure premium is $100 per exposure unit, which was filed and became effective on January 1, 1998. Using accident year 1998 experience, the developed pure premium, trended to the average effective date under the anticipated rates, is $135 per exposure unit. The loss cost trend is +10% per annum. The credibility to be assigned to the experience pure premium is 60%, based upon classical credibility procedures. What is the credibility weighted pure premium for the rate filing?

Limited fluctuation credibility says the following: the public and the regulator have seen a pure premium of $100 per exposure unit in 1998. Loss costs are increasing by +10% per annum, so they expect a pure premium of $110 per exposure unit in 1999. We are willing to change the pure premium to conform with our experience only to the extent that this experience is "credible" (regardless of the quality of other information). The classical credibility is 60%, so the credibility weighted pure premium is

$$60\% \times \$135 + (1 - 60\%) \times \$110 = \$125.$$

6. EXPONENTIAL SMOOTHING AND ACTUARIAL SHORT-CUTS

The formula above is correct, if our goal is limited fluctuation in rate levels. But limited fluctuation is not the only rationale for classical credibility. Let us change the interpretation of the credibility procedure; this in turn changes the appropriate formula.

The new pricing actuary asks: "How many years of data should one use for ratemaking?" The general answer is straightforward, though the specific parameters vary from case to case:

1. One should use as much experience as available, as long as it relates to the type of coverage presently being offered.
2. One should assign higher weight to the more recent experience, since it is likely to be a better predictor of future experience.[6]
3. The additional benefit of maintaining, trending, and adjusting older years of data declines rapidly, and this benefit is soon outweighed by the cost of this work. Actuarial short-cuts can improve the efficiency of the ratemaking process.

This actuarial short-cut is another use of credibility. Let us resume with the previous illustration. We are making rates for a policy to be issued on January 1, 1999, using data from accident year 1998, and assigning 60% credibility to the experience. For simplicity, let us assume that we have always assigned 60% credibility to the experience when making rates for this coverage.

Let PP_t be the pure premium charged in year t, and let EX_t be the pure premium indicated by the experience in year t. The pure premium charged in 1999 is

$$PP_{99} = 60\% \times EX_{98} + 40\% \times PP_{98}.$$

Assuming that the same 60% credibility value was used in the past, we substitute for PP_{98} to give

$$PP_{98} = 60\% \times EX_{97} + 40\% \times PP_{97}.$$

[6]This statement is more applicable for rapidly developing experience. Mahler [5] uses an illustration from baseball "win-loss" statistics, where there is no development. When significant and especially volatile development is expected, as in casualty excess-of-loss reinsurance, some actuaries are inclined to rely more heavily on older, more mature years of experience; compare Cook [2].

We combine the two equations to get

$$PP_{99} = 60\% \times EX_{98} + 40\% \times [60\% \times EX_{97} + 40\% \times PP_{97}].$$

We continue this substitution process to express the indicated pure premium for 1999 as a function of the experience pure premiums in all previous years for which we have data. If the credibility each year is Z, then the indicated pure premium for year k equals

$$PP_k = Z \times \sum_{t=1}^{\infty}\{(1-Z)^{t-1} \times EX_{k-t}\} \tag{5.1}$$

where the summation runs over all preceding years for which data are available ($t = 1,2,3$, etc.).[7]

This equation says that the indicated pure premium for year k is a weighted average of the experience pure premiums in each preceding year, where the weights form a decreasing exponential series. Intuitively, this makes sense. All experience provides some information useful for determining the new rates, but the older the experience is, the less useful it is.[8]

There are three problems with using the general equation for PP_k given above:

1. To use the general equation, we must retain all past experience, and we must re-analyze it each year. This can be a cumbersome task, and the costs might outweigh the benefits.

---

[7]See Mahler [6], pp. 255–256. If there are w years of data available, then the sum of the coefficients in Equation (5.1) equals $1-(1-Z)^w$. Thus, in theory, all the coefficients should be multiplied by $1/[1-(1-Z)^w]$. In practice, this is about the same as using a slightly higher Z value. Since values of Z within a fairly broad range work about equally well, an attempt to optimize the value of Z by the correction noted here would not be cost-efficient. See Mahler [6], pp. 256–257, on the relative efficiency of different Z values.

[8]For a more complete exposition of this rationale for the ratemaking credibility procedure, see Mahler [6]. Different rationales for the credibility procedure lead to different credibility formulas. Mahler extends the analysis by showing how the covariance structure for the risk parameters affects the optimal credibility to be assigned to the experience.

2. The general equation used here assumes that the value of Z remains the same from year to year. In fact, the value of Z may change from year to year, particularly if the volume of business is changing from year to year. If it does, the mathematics become much more complex.

3. The ratemaking process cannot be reduced to a rote formula. Every rate review requires the careful judgment of the pricing actuary to discern anomalies in the data, shifts in the external environment or in the company's operations that might affect the anticipated loss costs, and changes in compensation systems or consumer behavior that might affect the company's ultimate claim payments. Each year the pricing actuary may subjectively adjust the experience indication up or down based upon an analysis of the data and of the insurance environment. The general equation would require us to somehow retain all these adjustment factors.

The last problem listed above is critical. We do want to use all the past experience, but we also want to use the actuarial judgment of the ratemakers who analyzed this past experience.

The Credibility "Short-Cut"

The solution to all three problems is the same, as is clear from the derivation of the formula above. The underlying pure premium serves as a proxy for the experience of all prior years. The historical experience itself need not be retained by the company or re-analyzed each year by the pricing actuary. Credibility weights may have varied from year to year, and at each filing the pricing actuary may have adjusted the indications. The effects of all these factors are retained in the underlying pure premium.

Let us examine the rationale of this credibility formula in order to address Boor's fundamental question. We are credibility weighting the experience pure premium with the current pure premium, so we are tempted to think of limiting the fluctuation in rates. But the current pure premium is used here as a proxy for the experience pure premiums from past years. We use all this past experience in order to produce the most accurate indication. Our goal is greatest accuracy, not limited fluctuation.

Proxy Problems

The information properly assigned the complement of credibility is the underlying pure premium—as long as the underlying pure premium is indeed an accurate proxy for the indicated pure premiums from past years. If this is not true—that is, if it is not an accurate proxy—then a different complement is required.

There are two ways in which the current (underlying) pure premium may not be an accurate proxy for the indicated pure premiums from past years:

1. The pricing actuary, when reviewing the experience from past years, judgmentally adjusted the data and erred in doing so. In this illustration, we have implicitly assumed that there were no errors; we trust the previous pricing actuary's judgment. The current pure premium is the best proxy for the indicated pure premium from past experience years, after adjustment for data outliers, system changes, operational changes, and so forth.
2. The pricing actuary, after reviewing the experience from past years, filed one pure premium, but the state insurance department approved only part of the rate request. If we trust the judgment of the pricing actuary, we would use the filed pure premium, not the approved one. We are assuming that the state insurance department's actions were motivated by non-actuarial concerns, such as a political desire not to raise rates more than a certain amount.

Let us consider a numerical example for the second proxy problem. Suppose that the 1998 pure premium is \$100 per exposure, and there is a +10% annual loss cost trend. Based upon the 1998 experience, the indicated pure premium for the 1999 policy is \$135. The credibility to be assigned to the experience is 60%. As demonstrated above, if our goal is limited fluctuation in rate levels, then the credibility weighted indicated pure premium is

$$60\% \times \$135 + (1 - 60\%) \times \$110 = \$125.$$

If, instead, we are using the current pure premium as a proxy for the indicated pure premium based on past experience years, then we must know the filed and approved pure premiums for the 1998 policy. Suppose that the pricing actuary had filed for a +50% rate increase, but the insurance department had granted only a +25% rate increase.

These figures tell us that the pure premium for the 1997 policy was $\$100 \div 1.25$, or \$80. The indicated pure premium for the 1998 policy based on the 1997 experience was $\$80 \times 1.50$, or \$120. If we want to use the current pure premium as a proxy for the indicated pure premium based on past experience, we must assign the complement of the credibility to the \$120 adjusted for trend and calculate the credibility weighted indicated pure premium as

$$60\% \times \$135 + (1 - 60\%) \times \$120 \times 1.10 = \$133.80.$$

7. BAYESIAN CREDIBILITY

For applications of classical credibility, whether as limited fluctuation credibility or as actuarial short-hand for older experience, the information that receives the complement of credibility is determined by the purpose of the credibility procedure. So where do Boor's six characteristics come into play?

Venter describes Bayesian credibility as greatest accuracy credibility. If the rationale of the credibility procedure is to improve the accuracy of our indications, then characteristics such as predictive power, independence, and freedom from bias seem natural.

Nevertheless, a careful analysis leads to less firm conclusions. So let us tread gingerly over this terrain, beginning with the rationale for Bayesian credibility.[9]

There are some immediate problems with the question of desirable characteristics for the complement of credibility. In the Bayesian view, there is no qualitative distinction between the information that receives the credibility and the information that receives the complement of credibility. Pricing actuaries tend to think of the experience data as being assigned the credibility and some other data as being assigned the complement of credibility. Those of us who are steeped in classical credibility tend to think of credibility as a function of the reliability or the predictive power of the experience data.[10] This may indeed reflect the thinking of most pricing actuaries, but it is not a Bayesian view.

---

[9]When Boor speaks of Bayesian credibility, he uses an illustration of territorial relativities. This may confuse some readers, since there are two independent dimensions:

- Classical credibility versus Bayesian credibility, and
- Credibility for statewide indications versus credibility for territorial relativities.

When making rates for territorial relativities, most actuaries use classical credibility techniques, not Bayesian credibility. Since the aim is often to limit fluctuation in territorial relativities, the figure that receives the complement of credibility is the current territorial relativity. For territorial ratemaking, this objective is sometimes explicitly stated. See, for instance, Conger [1] on the objectives of personal automobile territorial ratemaking in Massachusetts.

[10]The CAS Statements of Principles and the American Academy of Actuaries Standards of Practice show how deeply ingrained this perspective is in the actuarial community. Theoretical actuaries may have discarded classical credibility in favor of its Bayesian counterpart, and the CAS examination syllabus extols the elegance of the Bayesian-Bühlmann procedures. Yet the CAS Statements of Principles and the ASB Standards of Practice show no trace of the Bayesian influence. The "Credibility" paragraphs in the Statements of Principles begin "Credibility is a measure of the predictive value that the actuary attaches to a particular body of data" (Statement of Principles Regarding Property and Casualty Insurance Ratemaking, lines 88-89; compare also Statement of Principles Regarding Property and Casualty Loss and Loss Adjustment Expense Reserves, line 193). The Actuarial Standard of Practice #25, "Credibility Procedures Applicable to Accident and Health, Group Term Life, and Property/Casualty Coverages," even defines the "full credibility" standard as "the level at which the subject experience is assigned full predictive value based on a selected confidence interval." This definition, albeit incorrectly worded (see the correct wording given earlier in this discussion), is based entirely on classical credibility theory; there is no concept of a "full credibility standard" in Bayesian credibility theory.

In the Bayesian view, we have two or more sets of data, each of which tells us something about the number that we seek to estimate. We would like to use all of these data to develop our estimate. That is, we seek a weighted average of the various estimators. The Bayesian credibility procedure give us relative weights to assign to each set of data.

We cannot speak of the desirable qualities of the information that receives the complement of credibility as if this information were somehow different from our other ratemaking data. No set receives the credibility with some other set receiving the complement of credibility. It is only by convention that we speak of the experience data as receiving the credibility and of some other data receiving the complement of credibility. This convention come from classical credibility, not from Bayesian credibility.

8. BAYESIAN RATEMAKING

Perhaps we can rephrase Boor's question as: "What are the desirable characteristics of the data that receives some portion of the credibility?" This is, indeed, a proper question for the pricing actuary to ask, and Boor's six characteristics are a valid set of characteristics. But this question has nothing to do with credibility. It is a question about data quality: "What are the desirable characteristics of ratemaking data?"

Actually, five of Boor's characteristics can apply to data quality. One of the characteristics deals more specifically with credibility. Boor's five characteristics of good ratemaking data are:

1. accuracy as a predictor of next year's mean loss costs,
2. absence of bias as a predictor of next year's mean subject expected losses,
3. availability of data,
4. ease of computation, and
5. clarity of relationship to the subject loss costs.

If one wishes to use two or more sets of ratemaking data, and to combine them by means of a Bayesian credibility procedure, then the ratemaking procedure is enhanced to the degree that the two or more sets of data are relatively *independent* of each other.

But we start with the data sets. For optimal ratemaking, we should use all the data available. Suppose we have three sets of data, A, B, and C. Set A is the historical experience. It is the best data, and it is the most acceptable data for the state insurance department, so we surely want to use set A. Sets A and B are relatively independent of each other. Sets A and C are relatively dependent. Under Boor's thesis we should assign the credibility to set A, and assign the complement of credibility to set B, not to set C. This will optimize the ratemaking procedure.

At first glance, Bayesian credibility doesn't say this at all. Rather, Bayesian credibility says that we should use all three sets of data and assign the proper weights to each of them. There is no constraint limiting us to only two data sets. In fact, it is common practice to use three or more data sets in many ratemaking applications. Property ratemaking uses five years of data (mandated by statute in many jurisdictions), with different weights applied to each year. Most common is a 10%–15%–20%–25%–30% weighting, with the higher weights applied to the more recent years. In theory, the optimal weights may be determined from a Bayesian analysis, though the accuracy of the final indication may not depend that strongly on the weights chosen, as long as they are within a reasonable range (see Mahler [6]). The experience loss ratio is then credibility weighted with a permissible loss ratio. Indicated territorial relativities are credibility weighted with the current relativities. In sum, the new rates for a particular territory are a weighted average of the indications from five separate years of experience, the current statewide rates, the current territorial relativity, and the indicated territorial relativity. These weights are not chosen by a Bayesian analysis. Rather, classical credibility procedures along with *ad hoc* weighting schemes are used. But for classical credibility, as noted above, Boor's paper is irrelevant. If Bayesian analysis were used for all the weights, then Boor's thesis does become relevant. However, we do not choose the two data sets that are most independent. We choose weights to optimize the accuracy of the indication, given all the data that we have available.

Workers compensation ratemaking provides another good example. The formula pure premiums are derived from six sets of data:

A. Financial Data (all classifications)
    (A.1) Calendar year experience
    (A.2) Accident year experience
    (A.3) Policy year experience

B. Unit Statistical Plan experience (by classification)
    (B.1) Indicated partial pure premiums
    (B.2) Underlying partial pure premiums
    (B.3) National partial pure premiums.

In fact, the procedure is even more complex, since between the unit statistical plan classification experience and the financial data statewide experience there is class group experience (manufacturing, contracting, and all other). In workers compensation, just as with property, the credibility weights stem from the early days of actuarial ratemaking, before Bayesian analysis caught the fancy of pure actuaries. In theory, though, Bayesian analysis could be used here as well.

9. HARWAYNE'S PROCEDURE

Harwayne's procedure for a three-way credibility weighting of workers compensation partial pure premiums is one of the most illuminating of Boor's examples. Indeed, Harwayne's procedure is a wonderful example of actuarial practice. It was a critical advance in workers compensation ratemaking, and it has since been applied to other lines of business as well.

How does it relate to Boor's thesis? It is a wonderful example, but what exactly does it show?

There are various elements in Harwayne's procedure. Some relate to ratemaking in general, some relate to credibility considerations, and some relate to the characteristics of the data that receives the complement of credibility. We must separate these strands, so that we can focus on the last of these issues.

To appreciate its elegance, Harwayne's procedure must be viewed in the history of workers compensation ratemaking.[11] The procedure uses three sets of information:

- the statewide classification experience, giving an *indicated* pure premium,
- the current classification pure premium (the *underlying* pure premium), and
- the classification experience from other states, giving the *national* pure premium.

Class Plan Refinement

The first two sets of information are routinely used in actuarial ratemaking. For instance, when making personal automobile insurance rates for the state of New York, the pricing actuary uses the New York experience and the current New York rates (perhaps adjusted for trend and similar influences). The pricing actuary would not consider Massachusetts personal auto experience or Illinois personal auto experience or national personal auto experience. So why is workers compensation different? Why does it combine the statewide pure premium with the national pure premium?

In personal auto, the classification scheme is a well structured, multi-dimensional system. For any classification, there is

[11] Harwayne's procedure is summarized in Boor's paper. It is presented in detail in Harwayne [4], along with the justification for its use.

generally sufficient experience in New York to set credible rates. Moreover, the Massachusetts and Illinois automobile compensation systems are so different from the New York system, that the Massachusetts and Illinois data won't help much. New York has a no-fault compensation system with a strong verbal tort threshold. Massachusetts has a no-fault compensation system with a much abused monetary tort threshold, and Illinois has a tort compensation system.

The states also have different classification systems, and they have different statutory constraints on underwriting, such as those relating to gender-based differentiation. Finally, they have vastly different rate filing systems. New York has a prior approval system, Illinois has open competition, and Massachusetts has a state rating bureau using a mandated financial pricing model.

In sum, the states are incomparable: Massachusetts experience and Illinois experience are nearly impossible to convert to "New York type" experience.

Workers compensation is almost the exact opposite of the personal auto situation. Workers compensation has a simple, one-dimensional classification system. Each state has about six or seven hundred classes, many of which have relatively little experience in any one state. Moreover, the states generally use similar class definitions.[12] Finally, the workers compensation systems seemed to be of the same type (at least to the founding members of the CAS), though there were differences in the parameters by state. Medical benefits are unlimited, and indemnity benefits are generally paid as some percentage of the pre-injury wage.

State workers compensation benefits were introduced rather suddenly in the early years of the twentieth century. Before the introduction of workers compensation laws, workplace accidents were handled through the tort liability system, with injured em-

---

[12]This is particularly true in the NCCI states, and it is even true in states which have their own rating bureaus.

ployees suing employers for negligence. The applicable insurance was employers' liability coverage, not workers compensation. When first setting workers compensation rates for a state, pricing actuaries had no prior experience from that state. For the smaller and medium sized classes—which comprised the majority of the workers compensation classes—a dozen years might elapse before there would be credible experience from the state under consideration. So how might one begin a workers compensation pricing structure?

Reduction Factors

Massachusetts was one of the first states to initiate a workers compensation system. As other states began their own systems, pricing actuaries took the Massachusetts experience and converted it to the benefit levels of the other states. For instance, suppose that the Massachusetts statute provided benefits equal to two-thirds of the pre-injury wage, and that the statute of another state provided benefits equal to 60% of the pre-injury wage. To set initial rates for the other state, begin with the Massachusetts rates and multiply them by 90% (= 60% ÷ 66.7%).

This procedure is straightforward and logical, enabling the efficient development of a complete workers compensation pricing structure. The founding members of the CAS meticulously calculated all the required "reduction" factors to convert rates from one state system to a second state system, considering not just different compensation rates but also various maximum and minimum benefit limitations, durations of compensation for various types of disability, and variations in state statutes regarding dependency awards. The result, as embodied in some of the first *Proceedings* papers, was a truly elegant actuarial procedure.

Unfortunately, it did not work. The founding fathers of the CAS spent months of painstaking work determining reduction factors to convert workers compensation loss costs from one state to another, only to have their results empirically invalidated by the emerging experience. The rigorous analysis, for instance, may have said that State X's loss costs should be 25% greater than State Y's, but the emerging experience showed that they were 15% lower. These results were surprising, but they were not wrong. There are many differences between compensation systems that are difficult to quantify. The administrative procedures in one compensation system, for instance, may encourage attorney involvement in workers compensation claims, while those in another state may discourage attorney involvement. The effects on loss costs can be dramatic, but these effects are rarely amenable to actuarial quantification.

Actuaries live by numbers. If one could not quantify the appropriate reduction factors, how could one use the experience of other states in setting rates? The first generation of actuaries rushed to develop reduction factors and to use the experience of other states in setting rates. The next generation of actuaries, discouraged by the empirical discrepancies, were ready to abandon these techniques and to use the experience of each state in isolation.

The flaw with the original procedure was the attempt to quantify *a priori* the reduction factor from one state to another. To the early actuaries, this had seemed essential: how could one use Massachusetts experience for a certain classification to help set the New York classification rate unless one knew how the Massachusetts classification loss costs would appear under the New York compensation system?

Harwayne saw a solution to this problem. Indeed, there are no reduction factors at all in Harwayne's procedure, because there are too many powerful but invisible factors that affect loss costs. Rather, Harwayne's procedure assumes that these invisible factors affect all classifications equally. Massachusetts loss costs may be unusually high because of greater attorney involvement in workers compensation claims, greater claims consciousness among the populace, or any such unquantifiable factor. But if we can empirically quantify the *overall* effect, then we can use the Massachusetts experience to help set other states' *classification* rates.

To highlight the advance made by Harwayne, let us consider a simplified example. Suppose that we have classification rates for State A, which is a large state with credible experience in most classes. We need to set classification rates for State B, which is a small state, with sparse experience in many classes.

If we look at the benefit structures in these two states, we might say that State B loss costs will be 25% higher than those in State A. This conclusion is not really helpful, since there are so many factors than affect the relative loss costs in the two states. Rather, we look at the overall empirical loss costs per exposure in the two states. We might find that the State B loss costs are, on average, 15% lower than those in State A. Using this figure as the implicit reduction factor, we multiply each classification rate from State A by 85% to get the indicated State B classification rates.

Unfortunately, this doesn't work either. We need an "overall loss costs per exposure" for each state. But there is no such thing as an overall loss cost per exposure. Some classes are high-risk, and they have high loss costs per dollar of payroll. Other classes are low-risk, and they have low loss costs per dollar of payroll. Perhaps State A has more high-risk classes and State B has more low-risk classes.

Think of the problem in the following fashion. We would like to derive the average loss cost per dollar of payroll. But the exposure base is not dollars of payroll. The exposure differs for each class: it is dollars of blacksmith payroll in the blacksmith class, dollars of carpentry payroll in the carpenters class, and so forth. One can not add blacksmith payroll to carpentry payroll.

Harwayne's solution was to translate every state to the same classification mix. Suppose that State A has two blacksmiths for each carpenter, and State B has two carpenters for each blacksmith. Harwayne's procedure calculates the overall loss costs per exposure in each state by taking 2/3 of that state's carpentry pure premium and 1/3 of that state's blacksmith pure premium. This puts the experience of both states on the same classification mix basis.

Harwayne's procedure solves the workers compensation problem, but this problem is unrelated to credibility considerations. Harwayne wants to use State A experience to set rates in State B by classification. He is not concerned with credibility.

As mentioned above, Harwayne's procedure deals with three issues:

1. The procedure adjusts the national pure premium to the benefit level of the state under consideration. This is the crux of the procedure. It relates to the general ratemaking issue of ensuring that the ratemaking data is not biased. It does not relate to issues of credibility.
2. Harwayne's procedure uses a complex three-way credibility weighting formula.
   A. The indicated partial pure premium has a full credibility standard based upon the expected losses in that classification, with the full credibility standard differing for serious indemnity, non-serious indemnity, and medical benefits.
   B. Partial credibility is set by a "three-halves" rule.[13] The three-halves rule says the following: If \$$X$ of

[13]The term "three-halves rule" stems from the obverse of this formula. If one needs \$$X$ of expected losses for full credibility, then for $Z\%$ credibility, one needs \$$X \times Z^{3/2}$.

One is tempted to delve into statistics textbooks to find a rationalization for the three-halves rule. In fact, the three-halves rule is used because it looks actuarial and it works. This justification of the three-halves rule has served admirably for over half a century now, and it should not be dismissed lightly. Any formula with a two-thirds power and used by actuaries all over the country must be mathematically unassailable; no one would simply make it up. And it works, in the sense that regulators and underwriters consistently defer to the pricing actuary's expertise in using this formula. They can't possibly argue with the formula, since they can't possibly understand it.

In a fascinating addendum to this school of thought, Howard Mahler has shown that the formula actually works. In fact, he shows that almost any formula works, as long as the credibility weights are within a reasonable range. Furthermore, they all work about equally well. Given the advantages to three-halves formula noted above, the formula is unimpeachable.

expected losses suffices for full credibility, then the credibility for \$$Y$ of expected losses is $(Y \div X)^{2/3}$.

C. The national partial pure premium has a full credibility standard based upon the national claim count in that classification. Once again, the full credibility standard differs for serious indemnity, non-serious indemnity, and medical benefits.

D. Partial credibility for the national partial pure premiums is set by a three-halves rule, similar to the rule for indicated partial pure premiums, except that claim counts are used instead of expected losses.

E. The credibility for the national partial pure premium may not exceed one-half of the complement of the credibility for the indicated partial pure premium. For instance, suppose that the indicated partial pure premium receives 40% credibility, and the three-halves rule would give a credibility of 50% for the national partial pure premium. The limit on the credibility for the national partial pure premium is $(1 - 40\%)/2 = 30\%$, so this is the credibility assigned to the national partial pure premium.

F. The remaining credibility is assigned to the underlying partial pure premium. In the example in the preceding paragraph, this remaining credibility is $(1 - 40\% - 30\%) = 30\%$ If this were a very small class, and the credibilities for the indicated and national partial pure premiums were 10% and 20%, respectively, then the underlying pure premium would receive $(1 - 10\% - 20\%) = 70\%$ credibility.

In Venter's terms [9], this procedure combines limited fluctuation credibility with greatest accuracy credibility. Since it does not purport to justify any of the parameters statistically, it would not be reasonable for us to rationalize the parameters after the fact.

Harwayne's Procedure and Boor's Thesis

The final issue in Harwayne's formula pertains to Boor's thesis. Why do we go to all the trouble of adjusting the national experience to the benefit level of the state under review? Why isn't it sufficient to credibility weight with the underlying pure premium, as is done in other lines of business?

Boor's paper provides the answer. The underlying pure premium is not independent of the indicated pure premium. This is particularly true for small classifications in workers compensation, more so than for most other blocks of business.

To see why this is so, let us consider a simple example. Suppose that we are setting rates for a new insured in State A in classification W. The classification is small; in fact, suppose that there are only five other insureds in classification W in State A. The historical experience is not fully credible. In other contexts, when we say that historical experience is not fully credible, we mean that random loss fluctuations may cause a significant disparity between the observed pure premium and the expected pure premium. In this case, the lack of full credibility has a more expansive meaning. Specifically, these five insureds may be better or worse than average, so we do not want to rely totally upon their experience to set rates for other insureds.

In actuarial terms, it is not simply that the historical experience is too volatile. Rather, we are afraid that the historical experience may be biased, though we do not know the magnitude of the potential bias or even the direction of the bias. To reduce the effects of the potential bias, we want to credibility weight the historical experience with additional information.

What other information should we use? The standard ratemaking answer is to use the underlying pure premium. In fact, many novice actuaries will indeed credibility weight with the underlying pure premium (or with the expected loss ratio). But this does not do the trick at all. The five risks in this classification have been insured for many years, and the underlying pure premium is based upon their experience in past years. We are concerned that they are not representative of the average risk. The underlying pure premium is just as problematic as the experience pure premium.

The national pure premium, however, is independent of the experience pure premium. It is based on the experience of other risks. The five risks in this classification in this state may be better or worse than average, but the hundred and fifty risks in this classification in the rest of the country are more likely to reflect the true average.

Contrast this workers compensation example with a corresponding personal automobile example. Suppose that we are making personal auto rates for a small classification W in state A. The classification is not fully credible, because there are only 500 drivers in this classification.

Here we are concerned with random loss fluctuations, not with bias. We are not worried that these 500 drivers may be better or worse than the average classification W driver that the company will insure. Rather, we are concerned with volatility. Perhaps the true expected claim frequency is 10%, so we should expect 50 claims. Actual experience may have been 40 claims or 60 claims, so the indicated rates may be 20% too low or too high.

The loss volatility affects each accident year separately. The most recent experience may be too high or too low, so we credibility weight with the underlying pure premium (or with the expected loss ratio), which reflects the experience of prior years, along with the business judgment of the past pricing actuaries. The underlying pure premium is not that interdependent with the historical pure premium, so there is less need to turn to external information.[14]

[14]The remarks made earlier about compensation system differences apply here as well. In workers compensation, if classification W has twice the average loss costs per dollar of payroll in State A compared to the statewide average, than it probably also has twice the average loss costs per dollar of payroll in State B compared to the statewide average.

We can now state Boor's thesis in rigorous terms:

> If the ratemaking data may be biased (though neither the magnitude of the bias nor even the direction of the bias are known), it is useful to credibility weight the experience indication with information that is relatively independent of the ratemaking data set.

Two characteristics of this revised thesis are of particular import:

1. All data sets used in ratemaking should have the five desirable characteristics drawn from Boor's paper. These characteristics are equally relevant for the information that receives the complement of credibility as they are for the basic ratemaking experience. The only difference is a practical one: often the data that receives the complement of credibility must be carefully adjusted in accordance with these characteristics, as is true in Harwayne's method.
2. Independence is particularly important when one believes that the historical experience may be biased, and especially when one does not know the magnitude or the direction of the bias. If the historical data is simply sparse, and random loss fluctuations may distort the indications, then independence is not of great concern. A larger volume of data is all that is required. It is the bias problem that demands a solution of independence.

In personal auto, the classification differentials are heavily dependent upon the compensation system and the underwriting structure. For instance, young unmarried male drivers may have an expected pure premium five times the statewide average in a tort liability state but only three times the statewide average in a no-fault state with a strong verbal tort threshold. Similarly, driver experience, or "years since first licensed," may be a more powerful classification variable in a state that does not permit underwriting by age of the driver than in a state which does allow such underwriting. Finally, the major classification dimension in personal auto is territory, which serves as a proxy for a host of hard to quantify loss cost drivers, such as attorney involvement in insurance claims and medical treatment of automobile injuries. Territorial relativities are peculiar to each state. One cannot credibility weight an indicated territorial relativity with information from other states.

10. RATEMAKING VERSUS PRICING

> … let me tell you how I use credibility. When I need a higher rate, I choose a credibility factor that gives me a higher rate. When I need a lower rate, I choose a credibility factor that gives me a lower rate.
>
> (A prominent American pricing actuary, 1988)

The previous discussions of credibility apply primarily to small and medium sized insurers whose experience is intermittently rocked by random loss occurrences. In personal automobile, most of the coverage in the United States is written by large carriers with thousands of claims in many states, such as State Farm, Allstate, USAA, GEICO, Farmers, and Liberty Mutual. The traditional formulas generally assign full credibility to their historical experience. Do they have any need for considering the complement of credibility?[15]

If the characteristics of the complement of credibility are important for the small insurer, they are crucial for the large insurer—though they are entirely different. The actuarial apprentice begins with traditional ratemaking, advances through financial pricing models and multi-year ratemaking procedures, and finally graduates to the tasks of the master actuary: marketplace competition, underwriting cycle movements, elasticity of supply and of demand, and the relationship of risk quality to price.

We want to examine the relationship of Boor's thesis to actual insurance pricing, not simply to traditional actuarial rate reviews. To understand the determinants of insurance pricing, we must first understand the economics of risk.

---

[15] In a similar vein, Richard Woll [10] points out that there is insignificant "process risk" in the claim costs of these large insurance companies, though "parameter risk" remains for them, just as it affects other insurers. Classical credibility theory—at least in the traditional treatment by Longley-Cook [5]—pertains to process risk, not to parameter risk. Indeed, these companies generally accord credibility of 100% to their historical experience in their formal rate reviews. However, the rate-setting practices of these insurers are far more market-oriented than are the corresponding rate-setting practices of the more traditional independent-agency companies.

Insurance Risk

Novice actuaries are often told that insurance operations are particularly risky, since the costs of coverage are not known until after the policy has expired. The nature of insurance risk has important implications for policy pricing and for Boor's thesis, so the dictum in the previous sentence warrants careful analysis.

Compare the auto manufacturer to the auto insurer. The auto manufacturer—so the argument goes—knows the costs of its inventory, its work force, its equipment, and its supplies before it sets a price for the final product. This price can be set as a fixed mark-up over the costs, ensuring a steady return for the manufacturer.

The automobile insurer, in contrast, needs actuaries to peer into the future—to convert raw historical records into prophecies of future costs. These prophecies are uncertain, so auto insurers need an extra margin of profit to compensate them for this risk.

This argument would be laughable if it were not so frequently repeated, in one guise or another, in actuarial circles. Yes, there are some risks that are indeed peculiar to insurers. Asbestos and pollution risks have hurt many large commercial lines carriers, and natural catastrophes have hurt many personal lines companies.[16] But these are the extraordinary events that have ruined the rare insurer: sometimes the overly aggressive insurer, sometimes simply the unlucky insurer. Insurers writing mostly the "bread-and-butter" lines with carefully considered reinsurance programs have largely avoided these risks.

Let us consider the true business risks to the manufacturer and to the insurer. Consider first the auto manufacturer. Most auto makers must design new model cars at least 36 months

[16]In truth, this argument sheds more light on the myopic view of many casualty actuaries and other businesspeople than on the attributes of the insurance industry. Asbestos has bankrupted its manufacturers, and pollution liabilities have devastated many chemical concerns. Asbestos and pollution have siphoned billions of dollars from the insurance industry, but most carriers will weather the storm.

before they are brought to the market.[17] The investment is enormous: retooling plants and equipment, sometimes building whole new factories, setting up production lines, producing hundreds of parts that will be needed with the new chassis, developing extensive advertising and promotional activities, educating an entire sales force of independent dealers with the characteristics of the new model.[18] Sometimes the new model will sell well, and the auto manufacturer will earn hundreds of millions of dollars. Sometimes the new model will flop, and the auto manufacturer will have lost hundred of millions of dollars.[19]

This is risk. It has nothing to do with Poisson distributions or inverse power curves.

Insurance does not have these risks. To produce insurance policies, the insurer must purchase a word processor and an office copier, hire an underwriter, and contract with an agent. It does not need a plant or a factory or a laboratory. The insurer does not spend tens of millions of dollars designing a product, buying parts, producing the final goods, advertising them in expensive campaigns. The insurer hangs out a shingle and sells the policy. Well … maybe it's not *that* simple. But the underlying principle is correct: the insurer does not face the large up-front capital commitment that represents manufacturing risk.[20]

---

[17]This time lag was about 60 months through the mid-1980's, until the intensified global competition from Japanese firms forced U.S. auto manufacturers to streamline their production schedules.

[18]As an example of the size of the investment, the decision to produce the Saturn automobile required General Motors to set up a new branch —the size of a major U.S. firm—many years before a single car would be sold.

[19]Other industries have equally great investments. Pharmaceutical companies, for instance, routinely spend tens of millions of dollars in research and development a dozen years or more before they expect to bring a new prescription drug to market.

[20]The formal economic expression of this is that manufacturers, utilities, pharmaceutical companies, and similar enterprises have high operating leverage, so their returns are sensitive to changes in market demand. Insurers have low operating leverage, since almost all their costs are variable. Even most of the expenses that casualty actuaries call "fixed expenses" are considered variable expenses by economists: they do not vary in direct proportion to premium, but they do vary with overall business volume. As a result, insurance profits are far less sensitive to changes in market demand.

In many industries, brand name differentiation adds to the business risks. It is not just the expense of manufacturing a new car that represents risk. To successfully bring a car to market, the auto manufacturer must convince dealers and consumers that the new model is superior to dozens of existing models. Insurance policies, in contrast, look more or less like one another across the industry. Product differentiation is hard to achieve in insurance.

Is insurance then riskless, or at least of low risk? Not at all, but the risk is of a different sort.

The ease of entry into the insurance market—or at least the apparent ease of entry into the insurance market—highlights the actual risk of insurance operations. Many insurance products are like commodities, with standard terms and multiple suppliers. Customer loyalty is high; that is, repeat sales are not as sensitive to price as new production is. As a result, many insurers are sometimes misled. They do not see high price elasticity in the majority of their business (that is, in the renewal customers), so they presume that customer service is more important than price.[21]

In fact, the opposite is true. Price is the dominant variable for new business production in most lines of business, and (because of high retention rates) new business production is of primary importance for overall volume and ultimately for the viability of the insurance enterprise.

Pricing: Cost-Based and Market-Based

The preceding paragraphs lay the groundwork; let us now return to Boor's thesis. The pricing actuary is in a quandary. The question is not what price best reflects the costs of the product. Actuarial ratemaking skills are so well-honed, and the law of large numbers so effectively eliminates much of the loss volatility, that actuarial techniques are accurate predictors of future

---

[21] It seems that every American insurer (by its own admission) provides exceptional service—or, at least, this is true for every failing American insurer.

costs. But the dilemma of the pricing actuary is different. If the price is too high, the insurer will lose market share: imperceptibly in the short run, but significantly in the long run. If the price is too low, the insurer will lose money on the policies that it sells.

The novice actuary retorts: "If our techniques work so well, the price will never be too high or too low." This actuary has confused ratemaking and pricing. Whether the rate indication is too high or too low depends on the technical skills of the actuary. Whether the price is too high or too low depends on market conditions (such as supply and demand) and the prices charged by competitors, which fluctuate with the underwriting cycle, not just with random loss occurrences.

Actuaries seem to espouse cost-based pricing to the exclusion of market-based pricing. This seems strange, since Western economists are virtually unanimous that market-based pricing—that is, pricing based on supply and demand considerations—is the linchpin of free-market capitalist systems. Cost-based pricing, in contrast, is not a rational pricing system for free markets. It has been used in regulated markets, such as in utility markets before the 1990's, but it would be useless in the competitive markets for property/casualty insurance that now prevail in most states.

In truth, the apparent predilection for cost-based pricing is an artifact of actuarial theory, not of actuarial practice. Actuarial theory emphasizes rigorous mathematical procedures. Cost-based pricing can be made as rigorous as desired, regardless of how relevant it is for the real world, so the actuarial literature is replete with formulas for cost-based pricing. Market-based pricing may be the crux of actual practice, but there are no theorems and few formulas, so the actuarial literature is devoid of papers on market-based pricing.

Boor's thesis is fundamental to the issues raised above. What is the ideal data that should receive the complement of credibility? The data from one's own company is inherently suspect, for two reasons. First, if the data relates to the coverage at issue, it is rarely independent of the historical experience. Second, such data gives us more information for cost-based pricing. A large insurer has all the information it needs for cost-based pricing. It needs instead information relevant to market-based pricing.

The rates charged by peer companies are the ideal data set for the complement of credibility, as long as they can be converted to the underwriting basis of one's own company. This conversion is critical for real-world pricing. Suppose that you are setting personal automobile insurance rates in a certain state. After working out the rate indications based on your own company's experience, you examine the rates of a major competitor. You find that your competitor's rates are about 40% higher than your own rate indications.

The first question should be: "Is the coverage the same? That is, are the underwriting criteria the same for the two companies?" Your company may be selling policies to standard or preferred risks, whereas your competitor may be selling to substandard risks. If your substandard rates are about 40% higher than your standard rates, then the disparity between your rates and your competitor's rates may be ascribed to underwriting differences, not to pricing differences.[22]

Competitors' rates tell us two things:

1. They tell us about the expected costs of the coverage, based on independent historical data, probably some differences in the ratemaking method, and differences in actuarial judgment.

---

[22]This is analogous to Harwayne's procedure. Harwayne adjusts for differing benefit levels and cost levels by jurisdiction. Here you are adjusting for differing underwriting practices by company.

2. They tell us a great deal about market place forces, competitive pressures, underwriting cycle movements, and supply and demand considerations.

Some actuaries are loath to incorporate market-based considerations into their rate recommendations. They say: "The actuary determines the proper rates—rates that are equitable for both insurers and consumers. Market-based pricing is irrational, based on seemingly bizarre underwriting cycle movements. Actuaries, as the champions of rigorous theory, should not be abetting irrational behavior."

This is a wonderful argument, but it is irrelevant. Real world insurers prefer market-based recommendations over mathematical elegance. Actuarial rigor is firmly established in traditional ratemaking departments. Actuaries who wish to be heard must seek the light of the marketplace.

Consider again the quotation at the beginning of this section. Yes, the language is a bit facetious: even actuaries should be allowed a sense of humor. But the underlying intent is serious. The actuary who made the remark—the chief actuary of one of the country's largest and most successful insurers—was particularly skilled at anticipating the rate movements of competitors, to know when it was safe to raise rates, and when other pricing or underwriting actions would have to substitute. He made this remark in response to a theoretical presentation on the credibility that should be assigned to the experience loss ratio. The elegant expositions so often heard at actuarial seminars and conferences are often irrelevant to real world pricing.

Let us rephrase the quotation in accordance with Boor's thesis. The pricing actuary ponders:

> My actuarial student, upon examining the company's experience, has obtained a rate indication of +6%. The marketing department says that our major competitors are about 8% to 10% above our rates. My guess is that

> our competitors will take rate increases of around 5% this year.
>
> This means that we could take a rate increase of as much as 14% or 15% this year without exceeding the market rate. Perhaps the rate indication of +6% is understated: maybe the trend estimate is too low, or maybe we had some particularly lucky experience this past year. Even if the +6% is accurate, we have all this leeway between +6% and +15%. Should we take something closer to +15% and reap the profits? Or should we take something closer to +6% and try to gain market share?

This is the essence of the complement of credibility thesis. Sophisticated pricing means weighting together independent indications to determine the rate request that is actually filed. If two indications stem from the same set of data, then these indications are probably interdependent, and they may contain little more information than a single data set would provide. If the two indications stem from different sources, and particularly if the rationale for the indications are different (e.g., one is cost-based and one is market-based), then the indications are probably independent, and the two indications provide more information than either one alone contains.

Actuaries well-versed in traditional rate-making techniques will object, saying: "How can one determine the proper credibility to assign to the historical data versus to the rates of peer companies? This is too subjective; there is no rigor in this." So these actuaries give up on real-world pricing, and they return to actuarial theory.

Quite the contrary is true. The traditional (classical) credibility figures are plucked out of the air. We say things like: "The full credibility standard is 1,024 claims, which gives a 95% probability that the historical claim frequency is within ±5% of the true claim frequency. If there are fewer than 1,024 claims, then the credibility assigned to the historical experience is determined by the square root rule, and the complement of credibility is assigned to the trended expected loss ratio."

There is no doubt that this impresses the layman. But what does "a 95% probability … " have to do with a firm that is trying to maximize profits? What relation does it have to pricing in a competitive market? The actuary is using cost-based pricing when the actual prices will be set by marketplace forces. No credibility formula will be correct, since the actuary has not asked the right questions.

The actuary should be asking: "If the indicated rate from my own experience is $1,000 per car for a certain classification and territory, and the corresponding average rate of my peer companies is $1,100 per car, what rate should I use?" This is the proper question, and this is a statistical question. The answer depends on (i) the price elasticity of demand, (ii) the persistency rate of insureds at different cost differentials, and (iii) the discount rate for future profits. At one extreme, with (i) a high price elasticity of demand, (ii) a low persistency rate of insureds at high cost differentials, and (iii) a low discount rate for future profits, it is wise to price below the competition (as long as one can do so profitably), pick up market share (both in new business production and in transfers from peer companies), and accrue the long term profits from the expanded book of business. At the other extreme, with (i) a low price elasticity of demand, (ii) a high persistency rate of insureds even at high cost differentials, and (iii) a high discount rate for future profits, it is better to move towards the market rate and to take the current profits from the redundant price.

This is a credibility question. At any given price elasticity of demand, persistency rate, discount rate, and differential between one's own indications and the rates of peer companies, there is a theoretically optimal credibility to assign to one's own experience. Of course, price elasticities are difficult to measure, and some companies do not keep track of persistency rates, but at least the pricing actuary is asking the right questions. Once casualty actuaries are turned in the right direction—that is, once they have formulated the questions correctly—they will make rapid progress on the solutions.

We have come full circle. Readers who skim lightly over Boor's paper receive the impression that the estimate of credibility is the crucial question, and the secondary consideration is to know what will receive the complement of credibility. On the contrary: until we define the purpose of the credibility procedure, we cannot know what should receive the complement of credibility. And until we know what will receive the complement of credibility, we cannot know the amount of credibility to assign to the experience.

11. CONCLUSION

Boor's paper leads in many directions, continually circling back to his thesis.

There are four rationales of credibility procedures: (i) limited fluctuation, (ii) proxy for past experience, (iii) greatest accuracy, and (iv) marketplace pricing tool. Each of these rationales implies a different formula for calculating the credibility, and each of these rationales implies a different set of data that should receive the complement of credibility.

Limited Fluctuation

Credibility may be used to limit the fluctuation in rate levels from year to year. This is particularly important in a regulated industry with great public concern about price increases and about alleged rate redundancies in some lines.[23] This rationale leads to the classical credibility procedures. The parameters of the full

[23]The author of this discussion, like most casualty actuaries, would dispute these allegations. Nevertheless, they continually recur, and they have great influence on many state legislators and regulatory officials.

credibility standard—that is, the size of the confidence interval and the probability constant—depend on how strictly one wishes to limit the fluctuation in rate levels.

The complement of credibility should be assigned to the "current rates:" i.e., to the underlying pure premium or to the expected loss ratio. The figure receiving the complement of credibility should first be adjusted for all factors other than random loss fluctuations, such as loss cost trends and changes in the insurance compensation system.

We may term this the Venter view of classical credibility. Some pure actuaries look with disdain upon this procedure, as a relic from unsophisticated actuarial practice. Nevertheless, it remains the prevailing standard in most lines of business.

Proxy for Past Experience

The credibility weighting procedure may be used as a proxy for the historical experience of older years.

The credibility assigned to the historical experience depends on the rapidity of shift of risk parameters over time. In more formal actuarial terms, it depends on the covariance structure of these risk parameters along a time dimension.[24]

(B) The complement of credibility should be assigned to the "current rates:" i.e., to the underlying pure premium or to the expected loss ratio, after adjustment for any part of the most recently filed rate revision that was not approved by the state insurance department. In addition, the figure receiving the complement of credibility should be adjusted for all factors other than

[24] See Mahler [6], pp. 261–263, for a full explanation. Based on Mahler's analysis, which examined baseball win-loss statistics, not insurance losses, a wide range of credibility figures may give equally good results. Mahler's sports results are probably valid for insurance experience as well, since they stem from the stochastic characteristics of random variables, not from any peculiarities of baseball. However, it is difficult to prove this assertion.

random loss fluctuations, such as loss cost trends and changes in the insurance compensation system.

This use of credibility is discussed by Mahler, and the adjustment to the expected loss ratio is documented by Graves and Castillo. The procedure is used by ISO for general liability ratemaking.

Greatest Accuracy

Credibility may improve the predictive accuracy of cost-based pricing.

This rationale is the underpinning for Bayesian or Bühlmann credibility methods. The credibility equals $M/(M+K)$, where M is a measure of business volume and K is proportional to the "within variance" divided by the "between variance."' This procedure is not concerned with deviations from the current rate.

In the Bayesian perspective, there is no conceptual difference between the credibility amount and the complement of credibility amount. There are as many estimators that may receive some credibility as there are ratemaking data sets. Ratemaking data sets are more useful to the extent that (i) they are accurate predictors of future experience, (ii) they are practical, and (iii) they are unbiased. Independence of these data sets avoids the costs of extra analysis that may have little benefit.

The Bühlmann credibility formula is commonly used in experience rating plans, though the K values are not always chosen by rigorous statistical analysis.[25] Bayesian credibility procedures have often been explored for territorial ratemaking, and K values have sometimes been estimated for these applications. Bayesian credibility analysis is not commonly used in practice for standard statewide ratemaking (class ratemaking), though many casualty

[25] In the past decade, there has been a trend toward a more actuarial selection of the K constants, particularly for rating bureau pricing procedures, such as those at ISO and at the NCCI.

actuaries and insurance companies have explored this topic and are using some of these procedures on a limited basis.

Pricing

Credibility may be used to combine cost-based and market-based pricing indications.

The goal of pricing is not to estimate the costs of the product but to optimize the long-term profits of the firm, or to meet other objectives of the firm. The credibility to be accorded to the company's historical experience depends on the price elasticity of demand, the persistency of insureds at different cost differentials, and the discount rate for future profits.

The estimate that should receive the complement of credibility is the marketplace price, for which the rates of major competitors (or peer companies) is often substituted. Adjustments must be made for underwriting differences among the peer companies.

The actuarial literature, which is replete with papers on ratemaking, is almost devoid of material on policy pricing. In practice, senior company actuaries provide both ratemaking and pricing recommendations for their employers.

Policy pricing is generally learned on the job, not from books and papers. Policy pricing is learned *from* experience; the price is not found *in* the experience.

The extension of actuarial expertise to real-world pricing problems in competitive markets is one of the most alluring tasks for the future casualty actuary. One of the primary questions is how much weight should be accorded to one's own indications. Boor's paper awakens us to the other, equally important question: to what information should we give the remaining weight?

Solutions to these two questions will help move the actuary's backroom desk to the forefront of insurance company operations.

REFERENCES

[1] Conger, Robert F., "The Construction of Automobile Rating Territories in Massachusetts," *PCAS* LXXIV, 1987, pp. 1–74.

[2] Cook, Charles F., "Trend and Loss Development Factors," *PCAS* LVII, 1970, pp. 1–26.

[3] Graves, Nancy C. and Richard Castillo, "Commercial General Liability Ratemaking for Premises and Operations," *Pricing*, Casualty Actuarial Society Discussion Paper Program, 1990, Volume II, pp. 631–696.

[4] Harwayne, Frank, "Use of National Experience Indications in Workers' Compensation Classification Ratemaking," *PCAS* LXIV, 1977, pp. 74–84.

[5] Longley-Cook, Lawrence H., "An Introduction to Credibility Theory," *PCAS* XXXIX, 1962, pp. 194–221.

[6] Mahler, Howard C., "An Example of Credibility and Shifting Risk Parameters," *PCAS* LXXVII, 1990, pp. 225–308.

[7] Rodermund, Matthew, *Foundations of Casualty Actuarial Science*, Second Edition, New York: Casualty Actuarial Society, 1992, pp. 1–24.

[8] Stern, Philipp K., "Ratemaking Procedures for Automobile Liability Insurance," *PCAS* LII, 1965, pp. 139–202.

[9] Venter, Gary G., "Credibility," *Foundations of Casualty Actuarial Science*, Second Edition, New York: Casualty Actuarial Society, 1992, pp. 375–483.

[10] Woll, Richard G., Review of Cozzolino and Kleinman: "A Capacity Management Model Based on Utility Theory," *Pricing, Underwriting and Managing the Large Risk*, Casualty Actuarial Society Discussion Paper Program, 1982.