RECORD OF SOCIETY OF ACTUARIES
1993 VOL. 19 NO. 1B

CREDIBILITY THEORY

Instructor: CHARLES S. FUHRER

MR. CHARLES S. FUHRER: My first topic is a general discussion or introduction to credibility. The first question is, "What is credibility?" The concept of credibility, in insurance, goes back quite a few years. If you are interested in some of the early history, you can look through some of the early journals of the various actuarial societies. Also, in our previous teaching session at the Society meeting in Las Vegas, Arnold Shapiro gave a concise history of credibility in insurance. See the *Record* 18-1B (1992): 659-62.

The concept of credibility can be confusing. I am going to use the term *credibility* as simply meaning the amount that we believe the experience of a particular risk when we're setting the rates for that risk. The common conception of insurance is that you cover a large number of people or risks and charge them all the same rate, and because you are the insurance company, the claims will match the average rate for that group of people. Somehow the large number of risks protects you.

Basing rates on a risk's own experience seems to contradict this conception of insurance. In my opinion, this common conception of insurance is flawed, and therefore there is no contradiction with insurance principles. The flaw is that the concept of probabilities, as well as the law of large numbers, do not necessarily depend on all the risks being identical. In fact, the statistical theorem, called the law of large numbers, only requires that the variance doesn't go up too quickly. This is not a subtle point, but I've seen many people, even in the insurance industry, miss it. For example, they will want to add a new insurance product, but then wonder if they can handle the risk of that new product. Because the new product starts out with a small number of insureds, it could easily have a 200-300% loss ratio. The insurer will decide either to not write the new product or to reinsure most of it. (I'm not against reinsurance. It just is not needed in this case.) This reasoning is flawed. If you add an unrelated risk, even a small one, to an already risky environment, you still get the safety of the law of large numbers. It is not any worse than adding another similar or identical risk to the portfolio. In any case, if you look at risks as individual entities, then you want to determine the best rate for each particular risk.

Most of what I'm going to talk about is aimed at group insurance, particularly group medical. Most of the concepts are readily applicable to the other group coverages, and the more general credibility material would also apply to nongroup coverages.

Typically, insurers experience rate (base the rate on a particular risk's experience) using a linear formula. In fact, the concept of credibility seems to relate to the linear formula. A linear formula may seem too simple, but is certainly a good starting point. Here is a simple example to illustrate how we normally think linearly about experience rating. If the credibility is 50%, then a risk with zero claims will have its rate cut in half. A risk with double the expected claims will get a 50% load. Thus the rate depends linearly on the experienced claims. In other words, the actual less the expected claims multiplied by the credibility is added to the rate. A nonlinear formula would be the equivalent of varying the credibility percentage based on the claim experience of the risk.

There are two ways of setting credibility. One way, which I'm going to present here, is used for experience rating. The other one is mainly used in setting manual rates for various classification criteria. An example of such a classification manual rate would be the rate for a particular industry. The credibility would determine how much belief would be given to the experience of the groups within that industry versus the overall average. The two ways of setting credibility are different with some similarities. The manual classification credibility typically is set by first determining the size of experience needed for 100% credibility. This is done by determining the amount of experience that would yield a rate whose 95% confidence intervals are within 5% of the rate. The 5% and the 95% were selected arbitrarily. The credibility for less experience is then determined according to one of many formulas.

The credibility for experience rating is done without confidence intervals. This kind of credibility uses a least squares criteria rather than confidence intervals. Suppose there are two random variables: X_1 and X_2, where X_t is the claims for a risk for year t, or the t insurance period. We typically use a year, so I'll just say year from now on. Now the real problem in credibility is to estimate the conditional expected value of X_2 given X_1. If we want to use a linear formula, assume this conditional expectation is approximately equal to the credibility Z multiplied by the claims X_1 plus a constant: $E(X_2|X_1) \approx ZX_1+C$ Now the reason I use "Z" is that typically in the casualty literature, credibility is called Z. I'm not sure that many people in the Society are used to it. Perhaps a C would have been more natural, but C is already used for a constant or for claims.

We want to use the least squares criteria to pick Z and C, that is, to minimize the "expectation of the square" of the difference between the conditional mean and the linear formula. That is minimize: $E[E(X_2|X_1)-ZX_1-C]^2$. On examination 110 the solution to this problem is presented:

$$Z = \frac{Cov(X_1,X_2)}{Var(X_1)}$$

and $C = E(X_2)-ZE(X_1)$, where Cov is the covariance and Var is the variance. C is just a mean correction term, so that if you have X_1 equal to its mean then X_2 will equal its mean. Most of the time, for ease in writing the formulas, we assume that the two means are equal. The means can be made equal by dividing by the trend.

This formula has a couple of nice intuitive explanations. I presented these at the Dallas Society meeting: *Record* 16,1 (1990) 55-72. First, if we assume that the variances of the two random variables are the same $(Var[X_1] \approx Var[X_2])$, then Z equals the correlation coefficient (ρ) in linear regression:

$$Z = \frac{Cov(X_1,X_2)}{Var(X_1)} \approx \frac{Cov(X_1,X_2)}{[Var(X_1)Var(X_2)]^{1/2}} = \rho$$

So this kind of credibility is linear regression and with X_2 regressed on X_1.

There was quite a discussion, I would not call it a controversy, at the Dallas meeting concerning linear regression and credibility. Some of the actuaries said that because the correlation coefficients were well under 50%, there was no credibility of claim experience. I disagreed with this. Typically, when linear regression is taught on the Examination 110, the model that's being looked at is where X_2 is actually a linear function of X_1. These are not random variables except that the observed values of X_2 are subject to a random error term. In this model, linear regression is used to discover the unknown parameters of the linear function. If the correlation coefficient is much lower than 70%, then there is little evidence of a linear relationship, and the parameters are so obscured by the errors that they cannot be reasonably estimated. In experience rating, there is no assumption of a linear relationship, and both X_1 and X_2 are random variables. There's really no reason not to use the credibility, no matter how low it is, as long as it is positive. Of course, if the credibility is very low, below about 1%, it wouldn't be worth the trouble of doing the calculation. There is really no reason not to use a 5% or 10% credibility level.

The second explanation is: let $Y_1 = X_1 - E(X_1)$ and $Y_2 = X_2 - E(X_2)$ so $E(Y_1) = E(Y_2) = 0$.

$$Z = \frac{Cov(X_1, X_2)}{Var(X_1)} = \frac{E(Y_1 Y_2)}{E(Y_1^2)} = \frac{E\left[Y_1^2 \dfrac{Y_2}{Y_1}\right]}{E(Y_1^2)}$$

Thus Z is the weighted average of Y_2/Y_1, where the weights are Y_1 squared. Since the credibility is sort of a ratio of X_2 to X_1, it is natural to use such an average. The Y_1 squared weighting results from using the least squares criterion. If you were to use a different criterion, then the weights would change.

This gets us into another subject, which I want to discuss briefly. Lately, the value of using the least squares criterion has been questioned in some of the statistical literature. Most of the theory of least squares statistics was developed by some important researchers in the early part of this century. They did not have computers. One of the advantages of least squares is that the calculations are relatively short. Covariance and variance are easy to calculate. Other criteria generally requires electronic data processing machines. Thus some of the modern statisticians have suggested that we should now move to these other criteria and that the early researchers only used least squares because it made the calculations easier. They've often suggested using the least absolute deviation criterion as an alternate, which they claim is more natural. Nevertheless, I think that there are still many good reasons to use least squares. In fact, least squares has some intuitive appeal. For example, the mean or simple average is actually the least squares location parameter. The least absolute deviation value is the median. Most people are used to mean as an average, although medians are a little better for some things. An outlier, a numerical value that lies far from the others, can have a great influence on the mean. In fact this is the essential problem with least squares. In doing credibility experience rating, one outlier can lead to ridiculous answers for credibility. Below, I will discuss some ways of handling this problem. In summary, if you do not like least squares, try to balance an object at its median. There is some research work that could be done to use the other criteria in credibility theory. This talk will stick with least squares.

The standard (casualty) formula has been developed by the casualty actuaries where they have used it in, for example, automobile insurance. It is almost certainly the credibility formula that you would use for group life insurance. Assume that, instead of just X_1 and X_2, we have n years of data, X_1 through X_n, and we're trying to estimate the rate for year $n+1$. So we write exactly the same formula as we did before, that X_{n+1} is approximately equal to the mean of the n years, multiplied by the credibility plus a constant:

$$X_{n+1} \approx Z\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] + C$$

Once again, Z will merely be equal to covariance, between what we're trying to estimate and what we're going to multiply the credibility by, divided by the variance of that same second term:

$$Z = \frac{Cov\left[X_{n+1}, \frac{1}{n}\sum_{i=1}^{n} X_i\right]}{Var\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right]} = \frac{\frac{1}{n}\sum_{i=1}^{n} Cov(X_{n+1}, X_i)}{\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} Cov(X_i, X_j)}$$

One of the better properties of covariances and variances is that they're linear. Therefore, in this last equation I have moved the covariance and variance operators under the summations. I have also used the fact that $Var(X) = Cov(X,X)$.

Now if we assume that $Cov(X_i, X_j) = w$ for all $i \neq j$ and $Var(X_i) = v$ for all i:

$$Z = \frac{(nw)/n}{\left[nv + (n^2-n)w\right]/n^2} = \frac{w}{v/n + (1-1/n)w} = \frac{nw}{nw + v - w}$$

$$= \frac{nw}{nw + v - w} = \frac{n}{n + \frac{v-w}{w}} = \frac{n}{n+k}$$

Where $k = (v-w)/w$. Note that $k \geq 0$ since $v \geq w$ and therefore $0 \leq Z \leq 1$. This formula: $n/(n+k)$ is well-known and often used. The quantity w is sometimes called the process variance and $v-w$ the variance of the hypothetical means. Thus k is the variance of the hypothetical means divided by the process variance. The rationale for this terminology is all covered in one of the study notes on the part 422 exam (the paper by Philbrick), so I will not go into a lot of detail on it.

This casualty formula is actually the basis of the formula that I developed in my 1988 paper ("Some Applications of Credibility Theory to Group Insurance," *TSA* XL) designed for group health insurance. The major theme of that paper was to develop the parametric method discussed later. This resulted from my dissatisfaction with a method presented by Margolin (discussed later), which I've called naive. I apologize for using that word. Actually, I'm extending what Margolin did to deal with how to change the credibility by size of group. He only derived a formula that gave the credibility for a particular size group and left it up to the reader to determine how to fill in the credibility for sizes of groups where there wasn't enough data. So, I'm presenting a method that one might naively infer from his paper. I'm not saying that his method is naive.

The underlying assumption in the standard (casualty) formula is that the covariance between any two of the $n+1$ years is constant. Some of the group insurance actuaries have used this standard (casualty) credibility formula by treating n as also referring to the number of individuals in the group. Now number of individuals can be expressed as number of employees, employees plus dependent units, or actually insured persons. In any case, they treat n as not the number of years, but as the number of people years exposed. In my opinion, this is not very good. It ignores the fact that each individual's claims tends to be more highly correlated with that individual's own prior years claims, than with everybody else's claims in the group. In particular, when I've looked at data, I found that for a one-life group, or an individual policy, the credibility would be about 20-25%. Now this may seem a little strange, since individual health rating seldom uses credibility, but I think we need to understand what we're talking about. What I mean is that, if an individual's health history is completely unknown, then how credible is that individual's claims? In the absence of individual underwriting data, the rates would vary considerably based on just last year's claims data. If the individual had a large hospital claim last year, there is a very high probability that the individual will repeat it. On the other hand, if the individual was healthy enough to not have any claims, then that individual is probably healthier than the average.

If the credibility of one person is greater than 20% and $Z = n/(n+k)$ with n the number of individual exposures in the group, then $1/(1+k) > 0.2$ and therefore $k < 4$. If $k < 4$ then a 50 member group has a credibility of over 92% and 80 members over 95%. This is much too high. When I wrote my 1988 *TSA* paper on credibility, I derived the formula for size of group completely theoretically. Now I'm seeing that it wasn't so much a new development, as much as finding a formula to get a better fit by size of group.

The Margolin paper (*TSA* XXIII-1971) on the naive method basically just derives the formula, the covariance divided by the variance. He then goes into a lot of detail, as to why the author felt we should all use that formula. To me, that was sort of a wasted effort, because people were already using that formula, and in fact, his derivation of it was a little bit long and hard to follow. Although there were many negative discussions of the paper, there would not be much disagreement on the concept of using the covariance divided by the variance as the basic credibility factor for linear least squares credibility. As we have seen, this is the basis of the traditional casualty formula. Margolin never actually tells us how to go about getting an answer by size of group. He merely estimates the covariance and variance for each particular size bracket. So I decided that his method is to put all of the group data into various size ranges and then calculate the credibility for each range.

Here is some simulated group data that are not real data, but provide a good illustration. I have done this on real data, and the results are similar. We have g groups with n_k individuals in group k. Let $x_{i,k,t}$ $(1 \le i \le n_k)$ be the manual loss ratio for individual i of group k in year t. Then let:

$$y_{i,k,t} = x_{i,k,t} - \frac{\sum_{r=1}^{g} \sum_{j=1}^{n_r} x_{j,r,t}}{\sum_{r=1}^{g} n_r}$$

The y's are the mean adjusted x's. Also define:

$$y_{\bullet,k,t} = \sum_{i=1}^{n_k} y_{i,k,t}$$

and

$$\bar{y}_{\bullet,k,t} = \frac{y_{\bullet,k,t}}{n_k}$$

The simulated data consists of 200 groups with 33,060 members.

$$\sum_{j=1}^{g} y_{\bullet,j,1}^2 = 3{,}395{,}165, \quad \sum_{j=1}^{g} \sum_{i=1}^{n_j} y_{i,j,1}^2 = 713{,}214.6,$$

$$\sum_{j=1}^{g} y_{\bullet,j,1} y_{\bullet,j,2} = 2{,}836{,}590, \quad \sum_{j=1}^{g} \sum_{i=1}^{n_j} y_{i,j,1} y_{i,j,2} = 170{,}287.10,$$

and

$$\sum_{j=1}^{g} n_j^2 = 9{,}970{,}220.$$

You can determine credibility from the simulated data using Margolin's method. Make a partition of the size of groups $0 = t_0 < t_1 < t_2 \ldots.$

Define $J(k) = \{j \mid t_{k-1} \le n_j < t_k\}$. Then for m with $t_{k-1} \le m < t_k$ set:

$$Z_m = \frac{\sum_{j \in J(k)} \bar{y}_{\bullet,j,1} \bar{y}_{\bullet,j,2}}{\sum_{j \in J(k)} \bar{y}_{\bullet,j,1}^2}$$

Here are some values from the simulated data:

k	t_{k-1}	t_k	Z_m
1	0	20	-0.027
2	20	40	0.039
3	40	60	0.141
4	60	80	0.753
5	80	100	0.294
6	100	120	0.275
7	120	140	0.210
8	140	160	0.496
9	160	180	0.557
10	180	200	0.370

The data did not consist of very many groups, about 200, which may be a little unfair to this method, but the results are typical of what can happen. The credibility is negative in that first bracket, because there were one or two groups that had high claim experience in one of the two years, and very low claim experience in the other year. That added a large negative term to the correlation or the covariance, and thus, created a negative credibility.

The credibility in the fourth bracket, 60-80, is much higher than one could conceivably want to use for that size group. I think there were only 20 groups in this bracket. It just so happened that all the groups that had high claim experience in one year, also had high claim experience in the other year, therefore the credibility that was calculated was extremely high. I would not suggest that anybody use a table like this for experience rating. We need to do some kind of smoothing.

At the time that I wrote the parametric paper (*TSA* XL-1988), I didn't say it was smoothing. I actually derived the formulas based on the linearity of the covariance. I will not repeat that derivation here. The derived formula is:

$$Z = \frac{k_1 + (m-1)k_2}{1 + (m-1)k_3}$$

In this formula, k_1 is actually the correlation between an individual's successive claim years and tends to be around the 20-25%. The k_2 and k_3 are correlations between different people's claims (in successive years and the same year, respectively), who are in the same group. I've estimated these two constants from data a number of times, and it tends to be at the 1-2% level.

There are many different ways you can estimate the k's from data. Here is one method:

$$k_1 = \frac{\sum_{k=1}^{g} \sum_{i=1}^{n_k} y_{i,k,1} y_{i,k,2}}{\sum_{k=1}^{g} \sum_{i=1}^{n_k} y_{i,k,1}^2}$$

$$k_2 = \frac{\sum_{j=1}^{g} y_{\bullet,j,1} y_{\bullet,j,2} - \sum_{j=1}^{g} \sum_{i=1}^{n_j} y_{i,j,1} y_{i,j,2}}{\left[\sum_{j=1}^{g} n_j^2 - n_j \right] \left[\frac{\sum_{j=1}^{g} \sum_{i=1}^{n_j} y_{i,j,1}^2}{\sum_{j=1}^{g} n_j} \right]}$$

$$k_3 = \frac{\sum_{j=1}^{g} y_{\bullet,j,1}^2 - \sum_{j=1}^{g} \sum_{i=1}^{n_j} y_{i,j,1}^2}{\left[\sum_{j=1}^{g} n_j^2 - n_j \right] \left[\frac{\sum_{j=1}^{g} \sum_{i=1}^{n_j} y_{i,j,1}^2}{\sum_{j=1}^{g} n_j} \right]}$$

Estimating k_2 and k_3 in this way is subject to the same kind of problems that we had in the first method. A correlation between groups is used. Particularly, k_2, although it is defined as the correlation between different individual's claims in the same group, it turns out that it's almost the same as the group-to-group correlation. The problem is that one group, even with a large amount of data, that has high claim experience in one year and low in the next, can actually drive that k_2 value down below zero. But at least the advantage in this method is that we use all of our data together, and thus, it's less likely that one group of that type will upset it. There have been times when I've obtained very reasonable looking answers from actual data. You could do an analysis of how accurate the estimate was, based on the size of the data. I haven't done that. These formulas are complicated so I'm not going to go through how they were derived. The data in the paper is eight or nine years old already, but I got extremely lucky, the two values came out close together and were about 2%. Since then, I've had trouble getting as nice a result. Two percent is probably a little on the high side, but once again, it depends on how we measure the size of groups. At that time, I used each dependent unit as a count of one and each employee as a count of one. If you use just employees, you obviously get a different answer. If you use the total number of covered people, which is often not available, then you're going to get another set of answers. One of the things I've found is that k_3 tends to be more stable than k_2. Since generally you want the credibility to approach one as m increases, you might decide to set k_2 equal to k_3.

In the paper, I went on to apply this method to a number of other problems. One of the things that is often done in group health insurance is to pool claims over a certain amount. This means that large claims are limited to a fixed maximum when calculating the experience rate for the group. For example, the fixed maximum, called the pooling point, is $50,000 or $100,000 depending on the group. If you do use a pooling point, you will get higher credibilities when calculating these k constants. Throwing out the high claims may not help though for getting a reasonable answer for k_2 or k_3. So, this is not necessarily a panacea, but it may help.

Here are the values based on the simulated data:

$$k_1 = \frac{170{,}287.1}{713{,}214.6} = 0.239$$

$$k_2 = \frac{2{,}836{,}590 - 170{,}287.1}{(9{,}970{,}220 - 33{,}060)\left[\dfrac{713{,}214.6}{33{,}060}\right]} = 0.0124$$

$$k_3 = \frac{3{,}395{,}165 - 713{,}214.6}{(9{,}970{,}220 - 33{,}060)\left[\dfrac{713{,}214.6}{33{,}060}\right]} = 0.0125$$

A nonparametric topic is the only one that has not appeared someplace in the Society literature. I did try this method on some real data a few years ago. It worked quite well. In this method we're going to use a nonparametric approach to smoothing that uses kernel functions. The concept is that to calculate the credibility for a particular size group, we'll use a weighted average of the credibilities for various groups in our data. The weight will be the distance between the size of the group in the data and the size of the group for which we're trying to estimate the credibility. Assume that we have the simulated data and we calculated a credibility for each group of the data. Now, for example, we want to determine the credibility for a 24-life group. In the naive method we just calculated the credibility for all the groups from 20-40, and we use this for the 24-life group. In this method, we take a weighted average of the credibility of all of the data groups that are near 24. A 24-life group will have the highest weight and the 23- and 25-life groups a little lower weight and 22 and 26 a little lower yet. To calculate these weights, we use some sort of smooth function of the distance that has a peak at zero. This is called the kernel function. The function that I'm using here is the normal density, which is a nice bell-shaped curve:

$$K(x) = \frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}}$$

The function is called K for kernel. The general formula for smoothing a function f defined on a set of values $k\,(1 \le k \le g)$ with smoothing parameter h:

$$f(m) = \frac{\sum_{k=1}^{g} f(k)K\left(\left|\frac{m-k}{h}\right|\right)}{\sum_{k=1}^{g} K\left(\left|\frac{m-k}{h}\right|\right)}$$

This general formula says that to estimate a function f at the point m, we take a weighted average of the function at various points k, where the weights are the kernel function applied to the difference between m and k. We divide by the sum of the weights.

The nonparametric credibility with smoothing value h and normal kernel is:

$$Z_m = \frac{\sum_{j=1}^{g} \left[\bar{y}_{\bullet,j,1}\bar{y}_{\bullet,j,2} K\left(\frac{m-j}{h}\right)\right]}{\sum_{j=1}^{g} \left[\bar{y}^2_{\bullet,j,1} K\left(\frac{m-j}{h}\right)\right]}$$

This is the formula that I used for credibility. I chose to use covariance in the numerator and the variance in the denominator. Since I applied the same weights to each, there was no need to divide by the weights.

Chart 1 is the parametric curve estimated from the same simulated data set. Of course, it's smooth. It is also very similar to what you would get if you were to fit a curve to the data points, using least squares.

CHART 1
Parametric Curve

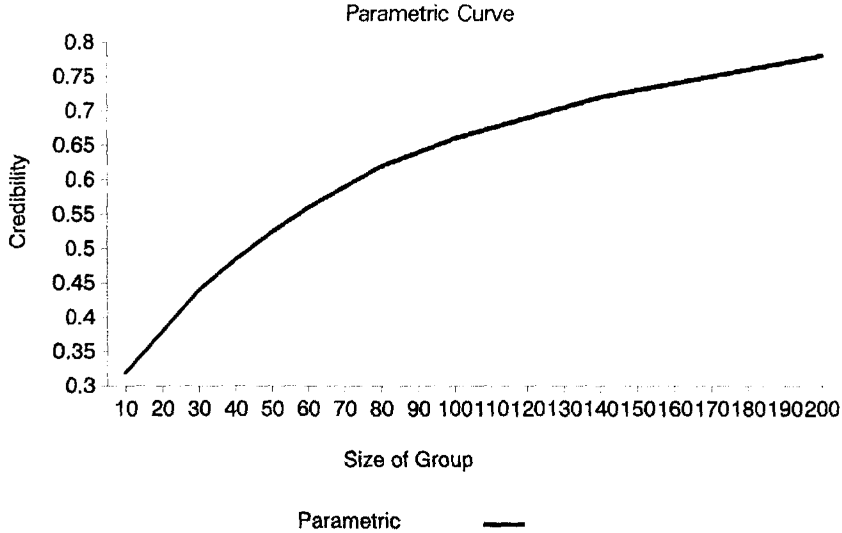
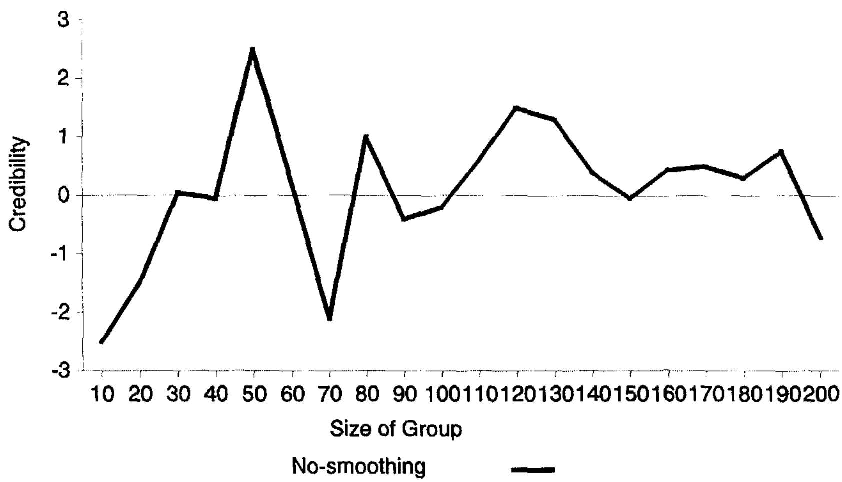


Chart 2 is the credibility that we get when we don't do any smoothing but instead, put it into the short buckets. No smoothing would not be a very good way of doing credibility.

CHART 2

Nonparametric No-Smoothing Curve



One of the problems with this method of smoothing is that we need to decide on how wide the kernel function should be, which is controlled by the smoothing parameter. If we're using a normal density for the kernel function, then the standard deviation is the smoothing parameter. The wider the kernel function, the more smoothing you're going to be doing. The narrower, the less smoothing but the closer you will be to the data. It's analogous to the process of graduation; really, it's almost the same as a moving average graduation, except this is continuous. Chart 3 shows a standard deviation of 15. The curve is smooth; it has a nice sort of gentle change in slope. However, it has a major drawback in that it has two places where the credibility starts back down again. This is not what we want.

So, one might then be tempted to do more smoothing. Chart 4 shows the smoothing parameter at 70. Everything looks fine, and now we have a curve we could use. I would maintain that this is too much smoothing. We've done more than smooth out the bumps in the curve. We've actually sort of pushed all the data together and lost what the data told us about the slope of the credibility. The shape of this curve is largely a byproduct of the groups that we had in the data. When I did this for some actual insurance data, I got a curve (with the smoothing about 15) that was very close to the parametric with k_2 and k_3 about 1%.

Chart 5 compares the last two smoothed curves to the parametric curve. I believe that the parametric is generally the best. Those formulas for k's are a little compli-cated but they're not all that hard to calculate. I think it works well. I also think that, if you can't get a reasonable answer for k_3 from the data, it would be better to set the value at about 1%. I think you'll find that these credibilities will make more sense for the marketplace, which, of course, is what this is all about anyway.

CHART 3
Nonparametric: 15 Smoothing Curve

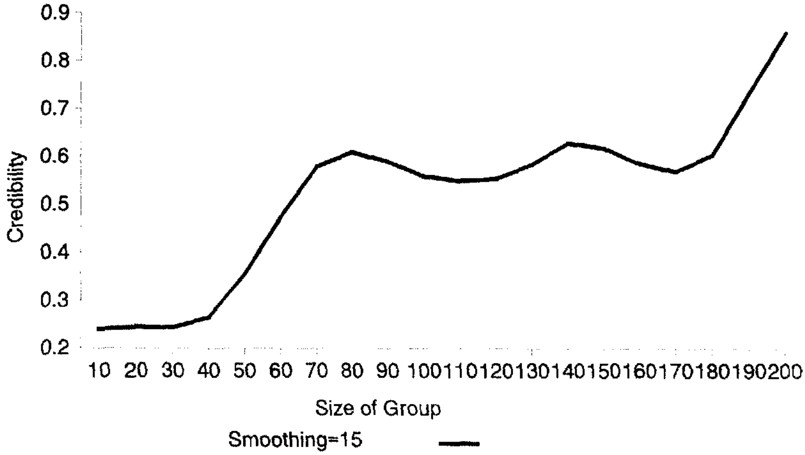


CHART 4
Nonparametric: 70 Smoothing Curve

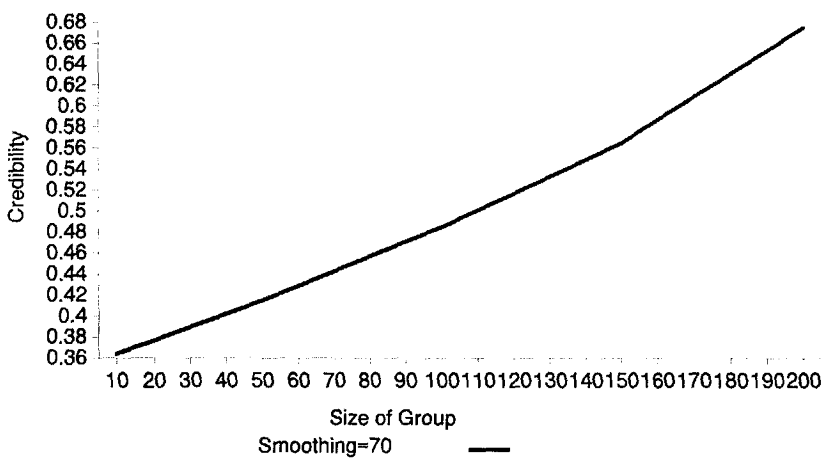
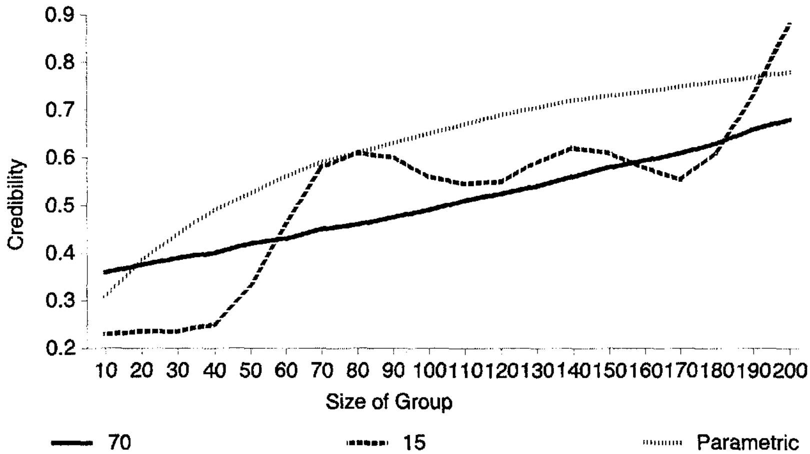


CHART 5
Comparison of 3 Curves



MR. FRANCIS G. MOREWOOD: Our company did a great deal of work on credibility in the mid-1960s, primarily to develop a formula for group life insurance. We quickly concluded that the credibility was more related to the number of claims than to the number of lives in the group. A very young group would get very low credibility and an older group would have high credibility with the same number of lives exposed in the group. I think you had mentioned a group or an individual who had no claims during the preceding year, as opposed to someone who had a heart attack during the preceding year. I'm quite convinced that credibility is more related to the number of claims than to the number of lives exposed, yet all of your work seems to relate to the number of lives exposed. I wonder if you have ever considered working with the number of claims, or even better, the number of expected claims according to a standard table?

MR. FUHRER: I completely agree with the suggestion that you use number of expected claims instead of number of lives for doing credibility in life insurance. Practically all of the remarks I made earlier concerned group health. The difference is that in life insurance we don't really get a chance to find out that somebody is sick. Usually once they have a claim, that's the end of it. I didn't really mean that to be funny, there are disability claims. If you look at the way the $n/(n+k)$ formula was derived, you will see that n should be the number of expected claims for group life. But, I don't think this is at all the same as medical.

MR. TIMOTHY M. ROSS: Regarding the last comment, in the $n/(n+k)$ formula, the process variance for a younger group would be quite a bit higher and that would lower the credibility.

MR. FUHRER: That is correct.

MR. ROSS: But the point is that then your credibility table would not be fixed. So maybe the best approach is indeed an expected claim approach.

MR. FUHRER: I think that you are suggesting using expected claims for medical credibility. Possibly you could do something similar on the medical where you would use the expected claims instead of the number of lives. I didn't get into that because I guess I didn't think it was too important. Usually the number of claims doesn't vary that much. You would not want to use the expected dollars of claim, because that would depend on the cost of medical care too much. Maybe age variation would make sense. Maybe I'll try to look at this idea.

MR. ROSS: Okay, I guess my comment then is with respect to the Bühlman approach: the ratio of the variance of the hypothetical means to the process variance. I'm not convinced that we're really fully aware of or we're fully testing the variance of the hypothetical means. Because if I understand correctly, that relates to essentially understanding just how good our manual rates are. In medical, we're looking at age, sex (where it's allowed), industry, and our plan rating values. So I guess I'm wondering to what extent have people looked at that explicitly. The only example I know of is in Roy Goldman's study note, (422-27-91) now a paper ("Pricing and Underwriting Group Disability Income Coverages," *TSA* XLII: 1990, p. 171) on LTD underwriting where he talks about variations from actual to expected by group size and that variation decreases for some of the larger groups.

MR. FUHRER: I'm not sure if that's a question. I think there is some question as to whether the quality of the manual that you're using affects the answer. The thing that I didn't say here is that I would prefer using manual loss ratios instead of actual claim data. When I say manual loss ratios, I'm talking about the claims of a particular individual divided by the premium for that individual. Effectively, we do have an individual manual premium, but we very seldom calculate it in group insurance. We calculate a single age factor for the whole group, but if you were to apply your underlying age table to your base rates, you could calculate individual rates. When I did this study in the paper, I did use loss ratios, and in fact, I came up with a slightly different formula that lowered the credibility for greater variance of the individual manual premiums.

The other thing is that you mentioned the Goldman paper. I did write a discussion of that paper, a good part of which dealt with the credibility of LTD experience. LTD presents another problem because, although you do have claim data that unlike life insurance can vary in size, it's like life insurance in the sense that the same individuals don't give you repeat claims, at least not very often. So it's a different thing, and I think there are some methods that would work with LTD that are different from both life and medical. I've been meaning to sit down and work through those. But I haven't had a chance yet.

MR. ROBERT E. COHEN: I definitely agree with the first individual's perspective that credibility is more related to number of expected claims. A very obvious example would be for a given size group, you're going to have much higher credibility for medical surgery claims, which are frequent and smaller in dollar volume, than you would for hospitalization claims.

MR. FUHRER: I think the number of claims could be incorporated into this parametric model of mine for medical also. That may be an improvement; I don't disagree with the concept at all.

MR. COHEN: Theoretical considerations aside, what shape of curve works best in the marketplace? It strikes me that your parametric approach produced an increasing but concave down curve, which strikes me as one that actually works in the marketplace.

MR. FUHRER: The whole issue of what works in the marketplace is extremely important. I believe that the parametric curve has the best place for the marketplace.

Let me speak about a related marketplace issue. I think that a lot of the companies were using a curve that was more linear than the parametric curve. That is they gave a lot less credibility to the under-200 life groups. I think they heard a lot of arguments from their sales people, who particularly wanted to see a lot more credibility in the 50-200 life market. They wanted to push the center of that curve up. Typically, what I saw happening in insurance companies was that the sales people make this argument to their underwriters. The underwriters would respond by saying that the actuaries have worked out the theory and the sales people were wrong. I think that when my paper came along, it hit some resistance because it actually said that the sales people were right. Nobody wanted to hear that. On the other hand, I think the marketplace wants to use as much credibility as possible. From a practical point of view, a group doesn't really know how you obtained your manual rates, doesn't trust them, and probably thinks they're conservative, (i.e., too high). They would just as soon go with 100% credibility all the time. So there is always a push towards higher credibility. Nevertheless, I think the sales people were correct. Often, actuaries would do well to listen to the sales department.

In one of the sections of my paper, I dealt with the problem that you might want to have more credibility for your bad or high claims groups than your low ones. If the marketplace is not using much credibility and so is using close to manual rates, the insurer will be able to keep the low claim groups at higher rates. So from a practical point of view, you might want to do that. This may get you into trouble with your brokers who think you should have the same credibility for high and low claim groups.

MR. THOMAS L. LUCERO: First, at the far left end of the curve, you were saying 20-25% credibility, assuming all you knew was the claims. If you know something more about the individual, you still get a residual of say 7-10% credibility. Second, at the high end, cases of 10,000 or more, I found that I couldn't get a credibility over about 80 or 85% from the experience. Obviously cases like that are always getting 100% credibility in the marketplace. Third, I notice your parametric curve is almost uniformly higher than the two nonparametric curves. I was wondering if there was some explanation of why that's the case?

MR. FUHRER: Let's see, working backwards, I think it was only higher because of the size of groups that I had. Also, as the smoothing goes up the curve tends to level out.

The second question was what to do about the fact that very large groups still don't have 100% credibility. You found that it peaks at about 85%. I suspect that is true. I think that to the extent that it is true, it may be due to a lot of the other uncertainties in the marketplace. Maybe new entrants to the group, changes in the hospitals in the region, or a number of other things that can have some effects that are unpredictable. In fact, the way the formula works, if you had 100% credibility, then it would imply that you could predict exactly what the group's claims are going to be. This would be based on the experience, and you can't really ever do that, of course. None of this is really very important because you're not going to ever convince a 10,000-life group not to use its experience alone, at least for the claims under the group's pooling point. And so, it really doesn't matter a whole lot what we do. In most cases, when I've used my formulas, when I've gotten up over about 90%, I've just rounded it to 100% and let it go, knowing full well that nobody would mind.

Your first question was concerning the one-life group. Your contention was that, if you did some individual underwriting, you could refine the rate and the residual credibility to the claim experience would still be about 7%. I've never done that. I have no idea if that's right. My guess is that it might even be lower than that. If we had actual health data, we'd probably know a lot more than just from the claim data.

MR. BRADFORD S. GILE: I just have a quick comment that I think might be of some interest. Regarding this $n/(n+k)$ formula, if you substitute premium in there for n, you get a credibility formula that's typically used in commercial fire insurance. In that situation, in fact, you can vary k by the nature of hazard, so that the more hazardous exposures would have a larger value of k. This might be appropriate for the varying types of group insurance. It wouldn't surprise me if that, in fact, is where it came from.

MR. FUHRER: I think you're right. That's where it came from. I also think that using the premium instead of n is more or less exactly what the first speaker mentioned. Using the number of expected claims in life insurance is effectively the same thing. So at least to the extent that premium is the same as expected claims, which in the case of fixed claim size, it's the same thing.

MR. GILE: If I could add one other thing. It seemed to me that maybe if you were using, say, a $P/(P+k)$ formula, where P was the premium, you could do a couple of things in group insurance. First, if you were dealing with a particular type of coverage that was highly experimental, you may want to use a large value of k. In group life insurance, if for some reason you were dealing with some exposures that were unusually hazardous, you might want to have a different value of k in the credibility table.

MR. FUHRER: Okay, that's possible, but generally, the less accurate your manual is, sometimes the more credible you want the experience to be. So, sometimes the less you know, the more credibility you'd want to have.

MR. MARK H. JOHNSON: Because of the nature of this smoothing function you've chosen, it's not surprising that with the higher smoothing constant, there's less credibility at the higher group sizes. Your credibility converges to a constant with increasing smoothing parameter. The parametric estimator of the credibility will always be higher at the higher group sizes, in the limit, because of how this smoothing parameter works. Also, there's been some discussion about the desirability of certain convexity features to these curves, yet the nonparametric curve converges to a flat line. Credibility should increase with group size, which in the limit doesn't happen with this smoothing function. So I wondered if you could comment on the desirability of such a smoothing function in light of those features?

MR. FUHRER: I think it's completely wrong when you're doing kernel smoothing to use too big a smoothing factor. Therefore it's almost irrelevant what happens when you approach the limit. Kernel smoothing should only be used with a relatively low parameter, to even out the ups and downs on a local level. There are some methods that actually give you a way of calculating what an optimum parameter is. The graphs showed that 70 was too high.

MR. JOHNSON: I thought you expressed some surprise that with the higher smoothing parameters, the credibility was lower at the higher group sizes. I think that you wouldn't want to use the extremes, but the fact that with the same set of data, at the higher group sizes, you will have decreased credibility with even a slightly increased smoothing parameter, I think is an undesirable characteristic. You might want to choose a smoothing function that would look more like a moment generating function or the value of the variable times the density function. Because that would at least give you an increasing credibility function with group size, which I think would be desirable.

MR. FUHRER: I think that's a really good point. I picked the normal because it was easy to work with. Also, the whole concept behind nonparametric smoothing is to not allow your preconceived notions about the curve influence the results. Thus symmetric kernels are usually used. This is one of the reasons that I prefer the parametric method.

MR. LAURENCE R. WEISSBROT: Has anyone already tested these credibility formulas? You get this request from the field. They say, "Show me that this really works." Has anyone gone back and said, "Okay, here's what our formula predicted for this year, here's what really happened. Had we used this credibility, we would have had a different result that would have been closer or further." Is there anyone else who's done this kind of testing? It seems to me that you could get an optimum credibility formula this way.

MR. FUHRER: Maybe somebody else could volunteer, if they've done anything. The only thing I've done is looked at it on an ad hoc basis. It should be worth pointing out though that the formulas here are basically asking, what would have happened if we had used them to predict the claims in the year two, based on year one? The credibility formulas are the optimum for those two years. So in that sense I have done what you suggest. The process of going back in statistics and seeing how well you did is really important, and I think this is something that we all should do.

MR. ROSS: The question was raised about the 10,000-life group and that it may not have theoretical credibility of 100%, yet it generally insists on 100% credibility. I think that you should switch over to the other model of credibility, based on confidence intervals, where with high percentage probability, we're confident that we're going to be within a certain percentage of where they're actually at. We apply some margin, and we're happy with the result. The client is happy with the result. That maybe gets into another issue of credibility, which I don't want to get into, but it has to be somewhat goal-oriented. How accurate do you want to be? What are you trying to avoid: lapses or losses? How do you balance those issues?

The other question that I wanted to raise goes back to what I was discussing before: I've been playing around with the approach of doing a true estimation, as opposed to a credibility approach. It seems to me in that kind of a method, you can take into account explicitly the measurement error of your claim data. Because generally you're going from paid claims, which you somehow adjust to a projected incurred loss. It's not clear how variations in measurement error affect the paid claims. For example, changes in claim lag, etc., are not taken into account in these credibility formulas. So I'm wondering what comment you might have on that aspect of it?

MR. FUHRER: First of all, I assume that when you did the estimation approach, you were effectively not using a linear formula. It's clearly superior if you have the data. The question of using paid claims, which is all that is available, versus incurred, is certainly also a good idea. Certainly when you use my parametric formulas, if you treated the second year as incurred claims but the first as the paid, then you could just directly get the answer, and presumably, there would be lower credibility. You could adapt the method to doing that, without making any explicit assumption on the relationship between paid and incurred.

MR. DAVID A. HILBRINK: I was wondering if you could explain briefly how you went about simulating the data used in your examples?

MR. FUHRER: I just assumed a credibility level in the underlying data, and used a random number generator to come up with the numbers, based on claim size distribution.

MR. ROBERT M. DUNCAN, JR.: Could you comment briefly on other than group health? What sort of considerations might one go through, given that things like underwriting and selection and turnover are probably much different than they are in the group market? What kind of relevance do you put to lives versus numbers of claims, or dollars in establishing scales or measuring experience?

MR. FUHRER: You're talking about individual health? Is that right? I think that if you're looking at individual, and if you haven't underwritten the individual or the family recently, that it would be appropriate to look at the claim experience, and I think you would get the credibility levels that push up into the 20% area. That probably gets into all kinds of ethical questions, as to what you're covering, and what the guarantee renewability option of those individual policies demand that you do or not do in terms of the rerating. My assumption here was that there was a separate term rate that was being given for each separate year. There was no extended renewability or anything on these things. That's probably all I know about individual health. I haven't worked in it, and so I'm probably not answering your question.

MR. DUNCAN: Well, I'm looking more for what approximates the group approach, once you have a block of business. When you're looking at the different cohorts that may come in or whether the block is closed, at what point do you get to a period of comfort with the data you have that you can assign some credibility to it?

MR. FUHRER: Okay, now we're talking about the other kind of credibility, where we're talking about setting up rates based on how big a cohort is, as opposed to experience rating an individual policy.

MR. DUNCAN: Yes, I'm not talking about the individual. I'm talking about the block or the form that people are on.

MR. FUHRER: I have done very little with that, also. I think that probably the confidence interval method is as good as any other. But once again, I don't really know the answer.

MR. DUNCAN: Are there any study materials or texts out on this?

MR. FUHRER: I do not know of any specific to health insurance. I might look at the casualty literature.