RECORD, Volume 24, No. 2*

Maui II Spring Meeting
June 22–24, 1998

Session 9I
Believe It Or Not!

Track: Education and Research
Key words: Education, Financial Reporting, Research

Interviewer: STUART KLUGMAN
Interviewee: THOMAS N. HERZOG
Moderator: STUART KLUGMAN

Summary: This session introduces credibility theory. It answers the five W's of good journalism: who, what, when, where, and why. "How" will be covered in a follow-up session. The questions answered include:
- *What is credibility?*
- *Who should be interested in credibility?*
- *Why should credibility be important to actuaries?*
- *When should actuaries think about credibility?*
- *Where is credibility important?*

Mr. Stuart Klugman: This session is supposed to be conducted somewhat like a talk show. I'm your host. I'm the principal financial group professor of Actuarial Science at Drake University. Tom is my guest this evening. Tom is the chief actuary for the Federal Housing Administration (FHA) in the Department of Housing & Urban Development. He will be playing not only himself, but also many other historical figures.

As is appropriate for all talk shows, the only reason someone would come to such an affair is to promote something they've done, so Tom, our guest, is here to tell you something about his book, *Introduction to Credibility Theory*, second edition, available at quality bookstores everywhere or direct from Actex publications. What is the big deal, Tom? Why should we be interested in credibility and what's it about?

Mr. Thomas N. Herzog: Before I answer your question, let me give you some background. I started getting interested in credibility in the early 1980s when I was working briefly for the state of Florida as a worker's compensation actuary. There was a lot of reading material from diverse sources required for the casualty exam on credibility. There was a lot of overlap and not much common notation, so my idea was to write something to eliminate the duplication, create a common notational scheme, and make it a lot easier for the students. That's how I got started in this, and my book grew out of that.

When we're talking about credibility, what we have is a problem in statistical estimation. Typically we're given some past data and we want to know the expected value of the next observation, but the problem can be a lot more complex. We may have prior information and collateral information, or we may need to estimate several means simultaneously. The prior information may be from past experience with this or similar insurance or industry data or it may be something totally different. There's a famous problem in a *Scientific American* paper that deals with baseball players. One year, 15 or 20 baseball players had a certain average after 45 at-bats, and the question was, what was their average going to be at the end of the season? It's a very similar problem here. You may have collateral information or, when estimating several means, we may decide it's acceptable to be biased on each one, provided the overall estimate isn't biased. As a result, we may be able to reduce variability. Another issue is dealing with loss functions. Do you want to use a squared loss function on many different baseball players' averages?

Mr. Klugman: I've studied statistics a fair bit and I expect almost everyone in our audience has also done some statistical studies. One of the things I teach my students is that, when it comes to estimating the mean, the most appropriate way to do that is through use of the sample mean. We learned that a sample mean is unbiased and consistent. It's linear, unbiased, and has the smallest variance. Why would we want to do anything other than use a sample mean?

I have an example of why the sample mean might not be the best thing to do. Suppose we have two groups of workers covered by worker's compensation. Group A had an exposure of 20 life-years with zero claims. Group B had 5,000 life-years of exposure and produced 75 claims. We have collateral information on the two different groups covered by worker's compensation and some industry data that says groups like these have generally experienced about 0.01 claim per life-year. From my statistical training, I know the SOA statistics exam says the best estimate of Group A's frequency is zero and for Group B it would be 75 over 5,000. Can you tell us what credibility has to say about that and help us perhaps understand why the sample mean might not be the best thing to use?

Mr. Herzog: I think Stuart and I are in agreement that estimating the Group A mean at zero makes no business sense. It does make sense that Group A deserves some reduction in premium for the favorable business it had in the past. At the same time, Group B is claiming that its mean of .015 is an unlucky aberration, and the true mean continues to be 0.01. What credibility tries to do is to strike a balance between the deviation attributable to bad or good luck and the deviation attributable to real differences.

I had a similar problem that dealt with some data quality issues at FHA. We had some miscoded data records, about 25, and they all had the same problem. The question was, if we looked at 200 more records that had similar characteristics, would they have the same problem as the first 25? It's just another way of restating the same problem basically.

Mr. Klugman: Tom, over the years, a number of different approaches have been developed for credibility. In the introduction to your book you discuss some of those distinctions. Again, at this point, we're just thinking of credibility perhaps as a way of combining data, in particular to improve upon the sample mean as an estimator. Can you tell us about how you've categorized some different approaches to credibility?

Mr. Herzog: The two approaches we're talking about here are the Bayesian approach and the frequentist approach. The Bayesian approach is named after the Reverend Bayes, who developed Bayes theorem. This approach allows you to explicitly quantify prior information. Here you treat parameters as random, and the data is fixed.

In the frequentist approach, the parameters are fixed and the data is assumed to be random. It's difficult at best to bring in prior information. Some of the credibility analysis uses parts of each of these approaches. The Bayesian approach allows you to model and reflect prior and collateral information, but it may contain nuisance parameters. By nuisance parameters we mean that they're not necessarily of primary interest. We then could estimate some of these parameters by frequentist methods if we wanted to.

Mr. Klugman: Perhaps, in a taxonomy of credibility methods, this first set of distinctions isn't as critical as the second one, which does lead to substantially different formulas and ways to calculate them. That's the distinction between the limited fluctuation approach and the greatest accuracy approach—some terms that came into use in the last 30 years, although they're now applied to methods that were developed much earlier. Tom, could you tell us about the two methods and maybe draw some of the distinctions between them?

Mr. Herzog: The limited fluctuation approach is a frequentist approach. It doesn't use any collateral information and the answer, in the sense of the credibility estimate, is in the form of ZX plus (1 minus Z)M, where X is the data, M is the prior mean, and Z is the credibility factor based on the quality of X. How much weight should we give to X? For the greatest accuracy approach, the linear version of this looks the same as the limited fluctuation approach, and the estimate satisfies the same equation, but Z also depends on the quality of M. The Bayesian version uses a full statistical model, and the result need not be linear, but under certain conditions it is. The greatest accuracy, the Bayesian, and the limited fluctuation approaches may all give the same answer.

Mr. Klugman: In preparing for this interview, I noticed that the word "credibility" isn't always used by the same people to mean the same thing. If you had to give a concise definition of a credibility approach or when you can tell that someone has used credibility, what would you say defines an analysis by credibility? Is there a distinguishing feature?

Mr. Herzog: If you're going to be a frequentist and use credibility then you end up with this equation: ZX plus (1 minus Z)M. You just use the base formula and get your predictive distribution. And, with your loss function, you get your point estimate that way.

Mr. Klugman: I think there's some disagreement that, for some people, credibility is only the formula ZX plus (1 minus Z)M, however you get to that point. It appears that you want to take, at times, a broader view that it's really the posterior mean. It's estimating the expected value using all information, including prior information, to get to wherever that takes you.

Mr. Herzog: I'd like to take a broader view and say you have your posterior distribution and your predictive distribution. When you specify a loss function, you then get the appropriate point estimate based on the loss function, which may not be the mean. I'd like to stress the importance of the loss function.

From the Floor: In the formula, the ZX plus (1 minus Z)M, where X is the data and M is the prior mean, can you describe what X and M are? Is M last year's net single premium rate, pure premium, or something like that? What would X be then—the average loss it had in the past year or something to that effect?

Mr. Herzog: That's right.

From the Floor: Could you elaborate?

Mr. Klugman: In the limited fluctuation approach, M is whatever it used to be. Often M, because it's a manual rate, could be an expected value that came from elsewhere. It's the thing you do if you don't have any data. That's why it's also a prior value. X is the appropriate thing from your data that would typically be a sample mean, but for this approach it wouldn't have to be.

We're going to begin our historical tour with our first individual, Albert Mowbray, who wrote a key paper in the first volume of the *Proceedings of the Casualty Actuarial Society* (CAS) in 1914. Your paper was called "How Extensive a Payroll Exposure Is Necessary to Give a Dependable Pure Premium?" and you introduced us to the concept of full credibility. A quote from your article: "A dependable pure premium is one for which the probability is high that it does not differ from the true pure premium by more than an arbitrary limit." In the context of the taxonomy we've discussed, how did this get credibility going in 1914?

Mr. Herzog: This is the basis of the limited fluctuation credibility approach. No statistical basis was presented at the time. An ad hoc basis was developed later, but this did not indicate what to do when the data was not fully credible. There are no set criteria for full credibility. I guess it was close to 100 years ago. This was quite a good paper for the time in that sense there wasn't a lot of statistical development going on. And there certainly wasn't the high-powered computing capability we have now.

Mr. Klugman: You're saying that, at some point we can just let Z be 1 and say that when you're thinking about experience rating, what you did and what that group of policyholders did should be sufficient to determine what their next premium will be. Other industry averages or the experience of other groups isn't needed when Z is 1. Any notion of how much? Your title says how much is necessary?

Mr. Herzog: What some of my colleagues have done is to manipulate some of the parameters so they always get to full credibility.

Mr. Klugman: Interestingly, the two types of credibility that you mentioned earlier, limited fluctuation and greatest accuracy, were developed about the same time. We'd now like to move ourselves to 1917 when our next guest, Albert Whitney, wrote in the fourth volume of the *CAS Proceedings* a paper called "The Theory of Experience Rating" and was able to solve the problem, as you noted, by not worrying about statistics. To quote you: "From the standpoint of equity to the individual risk, of striking a balance between class experience on the one hand and risk experience on the other, the problem is, therefore, to find and apply a criterion which will give each its proper weight. The balance depends on the exposure, the hazard, the degree of concentration within the class, and the credibility of the manual rate." What were you talking about?

Mr. Herzog: Let me try to relate each of those terms to current terms. The exposure is just the sample size, by the hazard we mean process variance, and by concentration we mean the variance of the hypothetical means.

Mr. Klugman: We might back up a moment in case some of our audience is in between your original 1917 words and the words "process variance" and "variance of the hypothetical means," which are the current names. Could you clarify for us just what those two phrases mean—the process variance and the variance of hypothetical means?

Mr. Herzog: We may have a number of different groups, so we can estimate a mean for each of those groups. Then the hypothetical means is just the variance of those means across all of the groups.

Mr. Klugman: How do you get them?

Mr. Herzog: You'd use standard statistical procedures, perhaps the method of moments, maximum likelihood, or something like that.

Mr. Klugman: You're asking the wrong guy because, in 1917, Mr. Whitney really didn't have a clue, but we'll meet some of his intellectual descendants later on who did work on this problem. In 1917, there was a theoretical notion that it was important in doing your credibility balance to look at how different one rating group was from another and how to measure that. I think we still want to mention something about process variance and what that is.

Mr. Herzog: Later we're going to talk about the expected value of the process variance, but to self-define it, it's just the variance of the process generating the claims or whatever the risk is. Maybe you want to comment further on that.

Mr. Klugman: It's the within-group variance—how much one policyholder in the group varies from another—and that's the variability you're insuring against, why the person is buying insurance, because he or she might be different from someone else in that group. But your different groups don't really care about. They don't want to pay for the fact that they're different from the other groups, particularly if they're better.

Mr. Herzog: One of the things I came up with was the credibility formula that the rate is equal to Z times the risk experience plus 1 minus Z times the manual rate. A novelty here was to get the credibility factor Z as the ratio of the exposure N and divide it by N plus K, where N is the sample size or the exposure and K is a constant. As we go on, we'll talk about how some of the people in the future try to estimate K.

Mr. Klugman: Now, you and Mr. Mowbray, who unfortunately had to leave or we could confront him face-to-face, came up with essentially the same formula, ZX plus (1 minus Z)M, yet you did it by different routes. Are there any fundamental differences between your view of how to do things and his view?

Mr. Herzog: Not that I'm aware of, but perhaps you'd like to make some observations.

Mr. Klugman: Well, Mr. Mowbray talked about Z being 1.

Mr. Herzog: That's right.

Mr. Klugman: And I notice as I look at your formula, Z could never be 1, unless K is zero. My guess is, whatever that appropriate constant is, zero is probably seldom the appropriate value. So, it looks like your Z can never be 1. Is this just a consequence of your mathematics or, as indicated, of your support for the notion that no group or set of data, however large, should ever deserve full credibility?

Mr. Herzog: Yes, I support the notion that no group should ever get full credibility because you can always get more data. So, as much as I don't want to disagree with my friend, Mr. Mowbray, I feel compelled to do so.

From the Floor: However, market forces might force you to give full credibility to a group because it demands it or its members will go across the street to insurer Y.

Mr. Klugman: The question was, would market forces lead you to a Z of 1, even though your theory says not to? I guess I'll respond for Mr. Mowbray and say you'd better quote his article.

Mr. Herzog: But also, if N is large enough relative to K, while it may never equal 1, it may be, for practical purposes, very close to 1.

From the Floor: My point is, what happens if you are dealing with a large group, and they have enough clout to say, "We want 100% credibility for our data or we'll take our business across the street?"

Mr. Klugman: I suspect the question still is, if you're close to 1, is it OK to let it be 1? Again, we note that in your paper you're more concerned with the business relevance of striking various balances. If you came out with a 0.995, would you accept that?

Mr. Herzog: Yes, but I think the point here is somewhat different. I think that the actuary is only one player in the insurance company. There are underwriters and various other people. I don't think actuaries necessarily have 100% of the say, but that's not a statistical question.

Mr. Klugman: Moving ahead, it's now 1932, and we have Francis Perryman in our studio. Writing in the 19th volume of the *CAS Proceedings,* you titled your article "Some Notes on Credibility," in which you had four notes. Note 1 is a switch from a binomial to a Poisson approximation. Using the Poisson approximation, you are able to do a couple of things. One is to find a way to remove exposure bases from the full credibility standard and move toward claim counts and base full credibility on the number of claims, rather than the number of exposures or the sample size. Also you can extend credibility to the compound Poisson model and, again, get a standard of full credibility.

Perhaps the most remarkable result is in your Note 4, where you did something your predecessors were not able to do—help them figure out what Z should be when it was not 1. Again, you're arguing from the earlier paper, where, if you had enough information, Z could be 1, but at that time the question of what to do if you did not have enough information for Z to be 1 had been left unanswered. In your groundbreaking 1932 paper you were able to solve that problem. Could you tell us about your solution and maybe go through a numerical example with us?

Mr. Herzog: It goes back to Mowbray's paper, where we just dealt with Z being equal to 1. Using the square root approach, we were able to deal with partial credibility problems. We can illustrate Note 4 with this example. Consider the case where we have 75 claims in 5,000 policy life-years. We then assume that we have the Poisson distribution applying with a normal approximation and the sample mean is 0.015. It should be within 10% of the true mean 80% of the time. When we do that, the formula for full credibility indicates that 164 claims are required.

Mr. Klugman: So this is basically the earlier result, that if we had 164 claims, we could use full credibility and announce 0.015.

Mr. Herzog: Right. This is the limited fluctuation approach that we're using and, when you crank out the formula, you get 164 claims that are needed for full credibility. But to get the partial credibility, we're using an estimate based on 75 claims. We calculate Z as the square root of the ratio of the 75 claims that we've observed, divided by the 164 that we need for full credibility. When we run this through the hand calculator, which of course we didn't have in 1932, we get 0.67 for the limited fluctuation estimate of credibility factor Z. Then, when we make the substitutions in our credibility formula, we get an estimate of $.67(0.015)+.33(0.01)=0.0134$. That's our estimated relative claim frequency.

Mr. Klugman: I notice when we apply your idea to the other group, the one that had no claims, zero over 164, square root or not, is still zero, so the other group gets no credibility and winds up getting no recognition of its excellent experience. It gets a Z of zero and will continue to have premiums based on the industry rate of 0.01. Does that make sense?

Mr. Herzog: That was the best I could do at the time.

Mr. Klugman: We're now in 1950, and our next guest is Arthur Bailey writing in the 37th volume of the *CAS Proceedings*. Your paper has a long title, "Credibility Procedures: Laplace's Generalization of Bayes' Rule and the Combination of Collateral Knowledge With Observed Data." Quite the name dropper there. It is at this point you wrote: "The ordinary individual has to admit that, while there seems to be some hazy logic behind the actuary's contentions, it is too obscure for him to understand. And the trained statistician cries 'absurd, directly contrary to any of the accepted theories of statistical estimation.' The actuaries themselves have to admit that they have gone beyond anything that has been proven mathematically, that all of the values involved are still selected on the basis of judgment, and that the only demonstration they can make is that, in actual practice, it works. Let us not forget, however, that they, the actuaries, have made this demonstration many times, and it does work."

So, we find that credibility methods that were developed by our previous guests, despite the fact that they were not based on sound statistical reasoning, were used a lot. They were used a lot because they worked. Companies used them, made money, and their customers were happy. The fact that the statisticians couldn't explain it didn't seem to bother anybody. But Mr. Bailey has set out to do something about that, so we wonder, looking at his title, how the old probablists Bayes and Laplace could help put the credibility ideas onto a mathematical footing. Just as a reminder to the audience, you might want to recall that Bayes' theorem was published posthumously by Richard Price, sometimes regarded as the original actuary, and we tried to have Mr. Price here for this presentation, but he had a prior commitment.

Can you tell us about how Bayes' theorem, prior probabilities, and the like, can help make statistical sense out of the credibility methods that we've been talking about?

Mr. Herzog: We're going to use Bayes' theorem. That's what we're talking about. When we do that, it turns out that the data in the sample is only part of the story. We can also use the manual premium or, alternately, information from other risk classes. We can do better in the Bayes model than assume such information is worthless, which is what's done in classical statistics. In Whitney's model, we're using a least squares approximation and best linear approximation to the posterior mean. What I've done in my paper is shown specifically how this works when the observation is assumed to be either from the binomial or from the Poisson distributions.

Mr. Klugman: We have a modern audience here who knows most things you didn't know about statistics. Have you learned how Bayes' theorem works and how it's used for statistical estimation? It's not surprising that statisticians and actuaries are surprised when they find out that the sample mean is not the best estimator because, once you adopt the Bayesian approach and agree that that's the way to do your analysis and follow it through, you come out with the credibility formula, a weighted average of the sample mean and something else. Mr. Bailey, you said you showed it for two cases?

Mr. Herzog: Right, I showed it for two distributions of the observations. One was the binomial, and the second one was the Poisson.

Mr. Klugman: Looking ahead, because you're a smart guy, do those two distributions have anything in common. Are they the members of some larger family?

Mr. Herzog: Yes, they're both members of linear exponential families, and I guess this probably works for all the linear exponential families. Under the Bayesian model, if you assume that the prior distribution is from the class of distribution called prior family, then this will work all the time.

Mr. Klugman: I read your paper, and it's a remarkable piece of work, for the times and, in particular, for what statisticians knew, to get to the right answer with the tools that were there. Some 17 years later, our next guest, Hans Buhlmann has the benefit of much better developments in probability and mathematical statistics. Writing in the *Astin Bulletin* in 1967, you essentially reproved Bailey's results in greater generality. You didn't limit yourself to the Poisson and the binomial distributions. Where your predecessor, Bailey, took page after page of manipulations to get to the result, your paper is just two or three mathematical parts. It's very brief, elegant, and simple.

To quote from your paper: "Insurance people are practically minded." Notice that theme seems to be reported by all of our guests. "Their goal is to get an answer that works, and if they can bend the mathematics to it, so much the better." You were the one who clearly, maybe for all time, did the best job of making the mathematics correctly agree with what we knew for some 50 years was the right thing to do. "Insurance people are practically minded," you wrote, "and so they have in many circumstances come out with pragmatic solutions, even if the theoreticians were not able to provide them with full theoretical justification for doing so. In the United States market—and this is a remarkable fact—the credibility formula has been found to do an excellent job."

You introduced some of the modern terminology and really nailed down what that K was all about in Whitney's formula. Tell us more about K and how the various factors interact to increase the value of Z, the credibility weight, or to decrease the value of Z, and how the various properties of the statistical process affect Z as derived through your formulas.

Mr. Herzog: What I did was to come up with K as the quotient of the expected process variance divided by the variance of the hypothetical means. So, if you remember, what we had was that Z is equal to N divided by N plus K.

Mr. Klugman: If the formula for Z is N over N plus K, and the sample size increases, what happens to Z?

Mr. Herzog: Z goes up.

Mr. Klugman: And would we agree that that's sensible, that the mathematical result that you've proven--remember, Mr. Whitney just kind of talked around it and came up with a good idea—makes sense? As you get more data, more weight should be put on the data that you have?

Mr. Herzog: Right.

Mr. Klugman: Another component is this process variance. Which part of K is that?

Mr. Herzog: That's the numerator.

Mr. Klugman: So, as the expected process variance goes down, Z goes up.

Mr. Herzog: Is that right?

Mr. Klugman: That's right. It goes up. Does that make sense? Can you help us on that? As the process variance goes down, do we assign more credibility?

Mr. Herzog: As the process variance goes down, everything looks more and more the same.

Mr. Klugman: Within our particular group we're trying to set a premium. If everybody looks the same, should they get a lot of credibility? That is, if everyone in the population looks pretty much the same, does it make sense that our sample should get a lot of weight? Sure. So you're two for two.

Mr. Herzog: Now, the hypothetical means. That's in the denominator of K, so the credibility should be directly proportional to that. Right?

Mr. Klugman: So, as the hypothetical means become more variable, Z goes up.

Mr. Herzog: We then want to give more weight to the observation.

Mr. Klugman: So, for an experience rating group, a health plan, or an individual policyholder, if our group is different from everyone else, it deserves more credibility?

Mr. Herzog: Right.

Mr. Klugman: This is a new thing that Mr. Perryman and Mr. Mowbray missed. This relates to the quality of the group. The sample size and the process variance are talking about our sample, its mean, and how much weight to put on it. If everybody else could be a lot different from us, does it make much sense to put weight on everybody else? Is that right? The process variance is measuring how useful all the other information is. There's a lot of variability from group to group.

Mr. Herzog: We're talking about the hypothetical means.

From the Floor: I've often wondered, though, in your N over M formula, how do you define choice N?

Mr. Klugman: That was the hard part.

From the Floor: Yes, but N is the sample size. Is the sample size 1 for one year's worth of data, and the sample size 2 the past two years' worth of data? Is that how the N worked, or is N the number of people in your group? You have 20,000 employees in a group, so N is 20,000.

Mr. Klugman: What's N in this formula? When we talk about the sample size, is it one group, one employee, or one year? The answer is, it could be.

Mr. Herzog: It could be, yes. The answer is yes.

Mr. Klugman: One way to think about it is for the formula ZX, where X is your standard statistical estimate of the mean. Typically, X is a sample mean, and N is the number of things that were averaged, whatever those things might be, to produce the sample mean.

From the Floor: So, if you were looking at five years of data, N would be five? If you're looking at the average claim per group, which has 10,000 employees, N would be 10,000.

Mr. Klugman: That's right. And if it's five years and the things being averaged are the loss ratios for each year, you only have one of each to average. But, if in each year you have the individual payments on thousands of individuals and you're averaging those larger numbers, then N is this large number, but it all relates to those variances. Is the process variance the year-to-year variance? If so, N is years. If the process variance is the person-to-person variance, then N is people.

From the Floor: Could N be exposure years as opposed to just a head count at that point in time? I have an observation on what you both just talked about in regard to which one it would be. It might be if you're calculating the premium rates on the group, the observation is the number of individuals in the group, but if you're calculating the refund formula for the whole group, then the observation is the total loss ratio, as you were talking about; therefore, in the observations, it is 1 for every year.

Mr. Klugman: Right. You could also be asking two different questions in using credibility for two different approaches. If we're trying to experience rate an individual and give that figure for that person's premium, then we'd want to count individuals. If we want to do a rate adjustment for a group, then N is the one group it might represent. We have several years of data on that group, but we may just have one observation, and that's a particular case where Mr. Buhlmann's formula becomes much more useful. Even one observation could still have high credibility if what you're putting the 1 minus Z on isn't very reliable. Again, your key breakthrough is putting it on a mathematical foundation but, through the variance of the hypothetical means, making it clear that the quality of what 1 minus Z multiplies is just as important as the quality of what Z is going to multiply. It's striking that balance that produces a good result.

Mr. Herzog: Again, what we've developed here is the linear approximation to the Bayesian solution. In certain conditions, these are equal and in other conditions they're not, in which case the Bayesian solution is the optimal solution, at least in theory.

Mr. Klugman: And, if I'm correct, in your paper you demonstrated your formula is always the best linear approximation to the exact Bayesian formula. I'm going back to the point when you, whoever you were at the time, made that all-assuming squared error loss.

Mr. Herzog: That's right.

Mr. Klugman: Our goal is that the best estimator is the posterior predictive mean.

Mr. Herzog: Under the squared error loss?

Mr. Klugman: I believe in your paper, squared error loss was crucial to developing the result.

Mr. Herzog: Right.

Mr. Klugman: It is very easy for me to read your paper because I can differentiate something squared, a critical step in the analysis, and that was not too hard to follow. You're applying it in the Bayesian context, recognizing that a linear approximation was needed to produce the time-honored answer.

From the Floor: On the expected process variance, let's say we have several years loss ratio for a group. Would that expected process variance be taken from looking at those loss ratios or would it be taken from a comparison between those and the manual rating?

Mr. Klugman: That goes back to your earlier question of how can we put numbers to these things because process variance is the within variance. It's just within our group. There wouldn't be a reference to a manual premium or any other experience. The process variance is how that group varies from year to year, or, again, if it's people, how people in that group vary from one person to the next.

From the Floor: So, if we had a group that went, say, for five years and had no claims, that would give us a process variance of zero?

Mr. Klugman: Let's be very careful. It's an estimated process because what we have is a sample of five.

From the Floor: Right.

Mr. Klugman: You've led us into the next question: How do you estimate these variances, the process variance and the variance of the hypothetical means, because in theory they're not zero? Am I right that the true variance is probably not zero. Unfortunately, in your example we can only see five observations from that population, and, therefore, we do have to address the statistical estimation error problem of putting a number to these things. That's where our next guest will provide us with some insights. Erwin Straub, you wrote a paper with Mr. Buhlmann in the *Actuarial Research Clearing House* in 1972. Actually, you wrote it earlier in German, and the English translation of the paper is appeared in *ARCH* in 1972.

You stated your objective very well: "It is the object of this paper to present a means by which such individual risk premiums can nonetheless be precisely defined and calculated. This is possible if the behavior of the underlying risk is described by a suitable mathematical model." The issue that you addressed very successfully in a formula that's still in common use today is called the Buhlmann-Straub formula, which addresses the issue of how you do sample data to put numbers to the things that we wish we knew but don't. Could you talk a little bit about how that works and what you came up with in your paper?

Mr. Herzog: The key assumption is that the variance per exposure unit is proportional to the reciprocal of the number of units. This differed from the earlier model that Mr. Buhlmann did by himself, which assumed that all the weights were the same. So, the variance or the weighting factor is reciprocal to the number of units, which allowed us to bring in the whole concept of different exposure units.

Mr. Klugman: But, aside from getting a slightly more complex refinement to the model, I've always thought was that the real contribution of the paper was the estimation part, trying to estimate the process variance and the variance of the hypothetical means. What sort of statistical techniques did you use to make that happen?

Mr. Herzog: We're taking the elements of K and estimating them by the method of moments, which still persists to the present time. The advantage is that we don't need any probability distribution, and the formulas are relatively easy to work with. The result is to get Z in the form of N over N plus K.

Mr. Klugman: If you look at your formulas for the two variances, they are exactly the same as the formulas for the within and between variances in an analysis of variance problem. Many of the members of our audience did learn analysis of variance at one time. The youngest folks, of course, know that that's been removed from the syllabus, but with credibility getting a more prominent place in our syllabus, it'll return in a different disguise as the process variance and variance of the hypothetical means.

Those two kinds of variances, the within group and between group variances, are exactly the elements of K, and they're derived by the same method-of-moments approach that is producing unbiased estimators that you found when you studied analysis of variance. In that setting, you might also have noted that they were maximum likelihood estimators, but that's only if you assume that everything has a normal distribution. It's probably not true for our insurance data, but we'll note it as a method of moments and unbiased estimators with no distributional assumptions. This distribution-free, moment-matching approach makes it very flexible. You did show that they're unbiased, right?

Mr. Herzog: Right. Going back to what you were saying about the analysis of variance, the advantage of today's computing world, looking forward here 30 years, is that you can use existing software to calculate these within and between variance components. For instance, if you have SAS or SPSS software, you can do that and a paper describing just that has been submitted to the *Journal of Actuarial Practice* by Dennis Tolley and others at Brigham Young University.

Mr. Klugman: Since 1972, most of the work in credibility has been refinements and not breakthroughs. The basic ideas of Buhlmann and Straub have persisted, and what people have done since then is made them a little more accurate and extended them to new settings. We're not going to have time to introduce all of the authors who have done something novel by way of extension, but we have a few of them here today. One is Charlie Hechemeister. From the proceedings of the Actuarial Research Conference that was held in Berkeley in 1974, you wrote one of the papers that appeared in *Credibility Theory in Practice* called "Credibility for Regression Models With Application to Trend." You made the observation that "inflation has moved from a minor annoyance to a major element" and that "no standards have been specifically developed for evaluating credibility of trend lines for states versus countrywide trend lines." It looks like your problem was that we were going to assign credibility weights to data from different states, a complement of credibility going to national data, but in both cases we wanted to project future losses in a linear fashion from the past. Can you tell us about incorporating trend lines and some of the challenges that might create for credibility analysis?

Mr. Herzog: The problem was that we had the energy crisis in 1973, so we had a lot of inflation, and this was becoming a real problem in property/casualty insurance. We put in some trends to take care of this inflation, but I was running into difficulties when trying to estimate both the slope and the intercept simultaneously. So, I used a transformation of the data that eliminated the need to estimate the value at intercept. All I had to do then was estimate the value of the slope. This was another example of how the models of Bailey and Buhlmann have been used and extended.

Mr. Klugman: As I mentioned, since 1972 we've seen a couple of different trends. One is to extend the ideas to new problems. Another has been to make refinements in the formulas of Buhlmann and Straub. Our next guest, Charles Stein, worked on the second idea. The statistics community thinks he's the inventor of this idea. Not so much anymore, because the statistical community has been reminded that actuaries knew about this stuff for a long time. The question was, how could the sample mean not be the best estimator? How can that be? The conclusion was that it's not. The mathematics were correct, as Buhlmann had definitely showed earlier, and we were told by another guest that credibility is astonishing. You did, through your work and as developed by others in the statistical community, one refinement, and it had to do with the estimation of those two variances. What was the problem with the estimating of variances that you worked on and what did you do to correct that?

Mr. Herzog: The problem is that the estimate of Z is biased. We introduced a correction factor involving the term N minus 3. When we did that, the credibility factors became unbiased, as long as the underlying populations have the normal distribution. If we go back to the baseball example that we talked about before which was the example that Efron and Morris used in their paper in *Scientific American*, when you're estimating four or more means simultaneously, the sample mean is no longer the best estimate, I think that's part of the point here.

Mr. Klugman: This is not unlike some of the degrees of freedom issues that you've spotted in other statistical work you've done, that, because we're estimating the variances from the data, you lose something. Therefore, you need more than one, two, or three simultaneous means before this all starts to work. If N is three or less—the N here is the number of groups, not the sample size—Z comes out to be 1. You get full credibility. You can't start doing this magical adjustment to a set of simultaneous sample means unless you have at least four of them. Again, your motive was to produce an unbiased estimate of Z while our friends, Buhlmann and Straub, produced unbiased estimates of each of the numerators and denominators of Z. If you want to show yourself to be smarter than many of my students, you can affirm that the expected value of a ratio is not equal to the ratio of the expected values. So, the ratio of two unbiased estimators is not itself unbiased, and this was your adjustment to make things a little bit better.

Mr. Herzog: Yes, and we should remind all our friends here that we're using a squared loss function.

Mr. Klugman: We mentioned earlier that the separate estimates were unbiased no matter what the distribution, but to get the ratio to be unbiased we needed the normal distribution to hold. Does this work? Does it make sense? Can somebody do it? As I read papers and I try to learn things, I always wonder whether anybody could actually do what the author's talking about. And, when I referee papers for journals, if they didn't do one, I tell them they have to because, if you can't do it, how do you expect us to read your paper and do it? I'd at least like to know that you could do it.

Our next guest, Glenn Meyers, writing in the *CAS Proceedings* in 1984, applied the Stein correction of the Buhlmann-Straub formula, to a worker's compensation problem. Why don't you quote from your own paper to give yourself a little publicity here?

Mr. Herzog: In summary, what we've done is present a credibility formula whose parameters are derived entirely from available data, and stated the assumptions that are used in deriving this formula. As is often the case in actuarial science, the model associated with these assumptions is necessarily simpler than the real world. That goes back to the famous quote from George Box at the University of Wisconsin that "all models are wrong, but some are useful."

We haven't mentioned the term empirical Bayes. Here, we're using the available data, rather than subjective prior information, to estimate some of the parameters of our model. The term empirical Bayes was formulated by Professor Robbins, in the early 1950s at Columbia University, where he is a member of the statistics department. I was at an American Statistical Association meeting where there was a big discussion on who did empirical Bayes first, Professor Robbins or Carl Morris, who was one of the people who wrote the paper on Stein's estimation.

Mr. Klugman: That's a good point you've made. The approach you use, Mr. Meyers, is a hybrid approach. Bayes formula was used by Buhlmann and by Buhlmann and Straub to get the N over N plus K, to understand what K should be, but the elements of K are estimated from data by classical statistics' method of moments, not a Bayesian approach, so empirical Bayes is often used to categorize methods that are hybrids that try to be as Bayesian as can be. We were asked earlier how you get numbers. With the Bayesian approach, you have to pull them out of your head, and that's sometimes difficult, particularly in a business, regulatory, legal environment where you have to be able to justify what you're doing.

It's a lot easier to justify a formula you can pull out of a book than a number you create out of your head, so the empirical Bayes approach solves that problem by making it all a little more concrete. It removes the subjectivity from the Bayesian approach. Although, as a good Bayesian, George Box, I don't have a famous quote from him, but I know what he'd say, and Meyers pointed this out: Everybody makes subjective judgments, Bayesian or frequentist, because the moment you postulate a model and make assumptions, you're being subjective and bringing your own decision-making ability to bear on the problem. That's what you point out in your paper. The method makes explicit all the assumptions that you're using so you have some feeling for how much confidence you can have in the results that you've put forth.

I'm literally beside myself with excitement over the next guest. One of the things you can ask as an academic is, what if you take things all the way? Nobody might care what happens. But we have the time and the luxury to ask these questions. One I thought about was, what happens if you try not to be empirical Bayes at all but take a Bayesian analysis all the way to the end? There's a way to do that. It's something called a hierarchical linear model that was brought out in 1972 by Lindley and Smith. What you try to do is move the subjective part as far away from the problem as you can. So, instead of letting those two variances be the subjective component, you put prior distributions on those two variances, abstract it to a higher level, and then use what's called a noninformative prior—which says I don't know anything about it—and a mathematical way to express it. Then the question is, can you do it? Can you get an answer? Do you want to speak for me, Tom, or should I just do this one myself?

Mr. Herzog: No, I think you should do that yourself.

Mr. Klugman: One of the ideas that you can address with a full Bayesian approach is that you never get a funny answer. With the empirical Bayes approach, you can obtain negative variance estimates. That has always bothered me a great deal. But if you look at the Buhlmann-Straub formulas and any refinements that have tried to fudge on those, you still always have the possibility of getting negative variance estimates. The idea then is that you make it zero, the closest legitimate variance estimate to it. Other folks have tried some iterative methods that you might think can never be negative, but they can all give you zero as an answer, so in that sense it's still not a useful solution. Any Bayesian analysis has the advantage in that the answer is always a member of the eligible space of values. The parameters are considered random. If I say a variance is a positive random variable, any estimate of a variance has to be positive. When you do a full Bayesian analysis, though, you have to do a lot of integration, and they tend to be integrals that aren't doable.

Yes, we all learned how to do numerical integration, yet I'm probably the only one in here who's ever actually done one. And it works real well. You may remember, though, that you could make it as accurate as you like by doing more and more subdivisions. And the computer round off in that never would get in the way, so we can do that. The problem with most full Bayesian analyses is that you need high-dimensional integrals numerically, and two-dimensional numerical integration is the square of the effort, the one-dimensional one and so on.

A lot of work has been done by Bayesian analysts in the last 15 years to improve our ability to approximate integrations in high dimensions and make some references about it. Brad Carlin had a paper on our transactions probably about four or five years ago where he does a full Bayesian analysis. Brad is an expert in what's called Markov Chain Monte Carlo methods of high-dimensional numerical integration, and you can read his paper to get some insights into what the high-level Bayesian analysts are doing numerically. I got lucky. When you take the standard Buhlmann-Straub problem—and in my paper I looked at the same data that Glenn Meyers looked at in his—and do a full Bayesian analysis, it comes down to a one-dimensional integral. It starts higher, but you can do by hand the antiderivatives until you get it down to a single integral. Then you can do that one numerically, and it takes very little computing time.

These are some of the things that people are working on, whether it's new models or methods of making the current model a little more accurate. To close things off I'm going to have Tom Herzog return and ask him about some recent activities, particularly in the health insurance area, in applications of credibility.

Mr. Herzog: There's currently either a task force, a Society committee, or some group of people who are trying to get together to bring credibility and health insurance together so we can get some applications of credibility to live data. There are four projects either currently going on or getting started in the next few months. There's some medical pricing being done with data originating from Trustmark, and the analysis is being done by Dennis Tolley and some people at Brigham Young. Jim Robinson has some data, and he's doing some work on cancer pricing. He's also anticipating doing some work on medical evaluation and individual disability income and disabled life reserves.

Two other recent papers, unrelated to this health insurance group, have been written by Virginia Young. They both have to do with credibility functions. The first paper is in the *Scandinavian Actuarial Journal,* and the second came out in the *North American Actuarial Journal* in January 1998.

Mr. Klugman: Credibility continues to be an interesting topic. I believe these are all sponsored by the Education & Research Section. At our annual meeting in October in New York City we have three sessions scheduled.

Mr. Herzog: That's right. Edward Frees and Virginia Young are talking about credibility theory and longitudinal data models that are used by survey statisticians, among others. Gary Venter, who's a member of the CAS, is going to talk about advanced concepts in credibility. And Dennis Tolley and I are going to try to put something together on credibility in health insurance, probably going back to his medical pricing paper.