RECORD OF SOCIETY OF ACTUARIES
1994 VOL. 20 NO. 1

MEASURING UNCERTAINTY IN LOSS RESERVES

Moderator: ROGER M. HAYNE
Panelists: SPENCER M. GLUCK
THOMAS S. WRIGHT*
Recorder: ROGER M. HAYNE

This panel will consider the question of uncertainty in loss reserves including discussion of sources and measurement. The panelists will analyze the same data set from different perspectives and discuss the strengths and weaknesses of the approaches.

MR. ROGER M. HAYNE: I am an actuary with Milliman & Robertson in Pasadena, California. The panelists are Spencer Gluck and Tom Wright. Spencer is also a consulting actuary with Milliman & Robertson in the New York/New Jersey office. Spencer has many years of experience in loss-reserve analysis for primary insurers and reinsurers, large and small. For a number of years, Spencer has been working with stochastic models for loss-reserve analysis and the measurement of uncertainty in loss-reserve projection. He has made numerous presentations regarding these models and serves on the Casualty Loss-Reserve Subcommittee of the Actuarial Standards Board (ASB) and was active in drafting the standard of practice on discounting reserves. He chairs the Special Subcommittee of the Actuarial Standards Board to write the standard of practice on reflecting risk and uncertainty in loss reserves.

Tom Wright is a chartered statistician in the U.K. and has worked as a statistical consultant in commerce and industry since 1984. He's been in property/casualty, or as they say in the U.K. general insurance, since 1988 including three years as a senior statistician in a large U.K. consulting firm called Bacon and Woodrow. He's now a partner in a smaller firm of consultants, English, Wright & Brackman which was formed in 1993.

We're going to discuss measuring uncertainty in loss reserves. The way we're going to structure this presentation is I'm going to lay the groundwork by identifying some of the issues, and maybe defining a few of the buzzwords. Then Tom is going to make a presentation on one set of methods that he has developed to estimate the variability. Spencer will make a presentation on another set of methods that are somewhat akin to Tom's. Then I'll come back in the end with a short presentation on a third way of looking at these things.

The first question that you should ask, is what are you measuring the variability in? Are you talking about variation or variability in the unpaid amounts? Is that the expected value of the unpaid amounts? That's different. Are you talking about the mean or are you talking about the entire distribution? Are you worried about the uncertainty in a particular parameter for a particular distribution? These things are important and this comes into one of my biggest bugaboos which is the use of the term confidence intervals when you're talking about loss reserves. Normally a

---

*Mr. Wright, not a member of the sponsoring organizations, is Partner of English, Wright & Brackman in Epsom, Surrey, England.

statistician is going to talk about a confidence interval as an interval about a parameter. What we're going to be addressing here are not intervals about parameters but distributions for the entire aggregate reserves. There's a big difference, and you'll want to be clear that you know what you're talking about when you mention these.

Generally, we identify three different sources of uncertainty. One is called process uncertainty which involves purely random fluctuations. Whether a claim is a dollar or $10,000, and whether a claim happens or not, this area of uncertainty is totally inescapable. You will always have it. The next level of uncertainty is a question of whether the parameters that you have in your models really are the right parameters. This is a little bit more difficult to measure. A third area of uncertainty is the question of whether or not the models that you're using are the right models. So we're going to adopt the terms process uncertainty, parameter uncertainty and specification error or specification uncertainty for those three sources of uncertainty.

The first one is easiest to measure. You set up the models by throwing a die or with random picks out of claim distributions. Once you have all your assumptions, you can turn a bunch of cranks and get a handle on that. Parameter uncertainty is a little bit more difficult. Specification uncertainty, or specification error, is even more difficult.

Generally I've seen a couple of major ways to approach quantifying uncertainty in loss reserves. One is a macro approach. The macro approach will look at the aggregate data and the triangles alone. Both Spencer's and Tom's approaches end up being macro approaches. They'll assume some sort of a curve or some sort of a runoff surface on the loss data and try to fit some kind of a model to that. Oftentimes, they're going to be stochastic. They're going to make some assumptions as to the underlying distributions so that you can derive some information about not only the process but the parameter risk and hopefully something about the specification error.

I would like to call the second genre of approaches micro because they kind of start off at the lower, smaller level and looking at the individual claim, the phenomenon, and the processes that do generate the losses. Here we'll take a look at models based on the collective-risk approach and these are the models that I tend to like a little bit more. They do address the actual uncertainty that arises out of claim count and claim size. There are ways that you can build in parameter uncertainty. It turns out that Tom, at least in his first paper, starts off with a collective-risk approach, makes some assumptions, and ends up coming into a macro approach. Tom creates a bridge between the micro and the macro approaches.

I sent both Tom and Spencer a data set. The data set was composed of several loss-type triangles: one triangle of paid losses, one triangle of incurred losses, a triangle of claims closed with payments, and I believe I sent a reported claim triangle along with a vector of exposures. These data were for a reasonably short-tailed liability line. We're not talking about heavy tail environmental or anything like that (not even as long-tailed as medical malpractice). There has been much happening with the data. There have been changes in the rates at which claims were being closed. There have been changes in the relative adequacy of the case reserves. It is something that's close to a real data set, but it's also highly pathological. So I threw down the gauntlet to both Spencer and Tom to try to estimate the uncertainty present in this data set. With that, I think I'll turn it over to Tom.

MR. THOMAS S. WRIGHT: I'm going to talk about two stochastic methods, which for the purpose of this talk, I'm calling Method A and Method B. Method A uses the aggregate incurred data. By incurred I mean aggregate paid plus outstanding. So of the triangles that were available, I'm using the aggregate-paid-loss triangle and the outstanding-loss triangle. Adding the two together gives what I call aggregate incurred.

The second method, Method B, is an average-cost-per-claim method, so it uses the aggregate-paid-loss triangle and the triangle of nonzero claim numbers. It uses operational time. The average cost per claim is modeled as a function of operational time instead of development time. Operational time is defined as the proportion of claims closed at any point. It is the cumulative number closed at any point divided by the ultimate number; it starts off at zero at the beginning of each origin year and eventually reaches one.

Roger said that he regards my approaches as bridging the gap between the micro and macro approaches. Both these methods sort of do that. They are macro in the sense that they're methods that are meant to be applied when you only have aggregate data available. But the methods are developed on the basis of mathematical models of the underlying claim process for individual claims, so they start off as micro models in Roger's terminology.

Both methods have a certain amount of leeway with regard to making assumptions within the framework, so I've applied both Methods A and B with two different sets of assumptions. The key assumption of Method A is the assumption about what bias exists in case estimates. Generally you find that the amount that is eventually paid is usually less on average than case estimates at any point. You have to make an assumption about that, but the data does obviously give you some information about what the appropriate factor is.

So, in Table 1, I've run Method A with two values of this bias factor and the results give you an idea of the sensitivity to that assumption. Under one assumption, I have $202,000 for the total reserve and another gave me $213,000. The standard error in the right-hand column includes both process uncertainty and parameter uncertainty.

The process uncertainty is the uncertainty due to the inherent random nature of the claim process. The parameter uncertainty is the uncertainty due to the fact that we have a finite amount of data, so the parameters of any model are not going to be 100% reliably estimated. Both those components are included in the final column.

For Method B, the main parameter that you have to specify is something called the variance index which I'll explain later. You can see it's a more objective method in the sense that it's not so sensitive to the value specified for this parameter. The reason I've applied both these methods is because the available data included all three of these triangles—aggregate paid losses, outstanding losses, and the number of nonzero paid claims. I don't have a single method that uses all that information simultaneously, so I've applied both methods and then performed a sort of judgmental averaging of the final results. This is how I allow for model specification uncertainty.

TABLE 1
RESULTS

Method	Assumptions	Reserve	Standard Error
A	o/s bias factor = 1.5	$202,000	$12,000
A	o/s bias factor = 2.0	213,000	12,000
B	variance index = 2.0	213,000	14,000
B	variance index = 1.6	210,000	13,000
Final Estimate		210,000	13,000

What do the methods have in common? They're both based on a mathematical model of individual claim payments. They're both fitted by iterative-weighted least squares. That's because the mathematical model gives both an expression for the expected liability and a variance. It turns out that the variance is a function of the mean. Normally, to fit a model by least squares, the variance would specify the weight given to each data point, so a point with a high variance which is relatively uncertain would get relatively little weight in fitting the model. Such a point has relatively little influence in determining how the model is fitted.

The purpose of fitting the model is primarily to estimate the mean. If the variance is a function of the mean, obviously you must proceed iteratively. You have to make initial assumptions and specify the variance so you can fit the model that estimates the mean; then you get new estimates of the variance. Those estimates specify the weights for a second fit, so you have to proceed iteratively. That process is equivalent to minimizing a function known as the deviance or maximizing something called the quasi-likelihood which is a generalization of maximum likelihood estimation. This sort of technique (iterative-weighted least squares) can be done using GLIM which is a specialist package for doing precisely this. GLIM stands for generalized linear interactive modeling. It can also be done in SAS which is a very well established statistical software package. The facility to fit generalized models has only been introduced into SAS in the last year or so. I haven't actually used it, but I'm told it now can be done using SAS. Both methods are published in the *JIA* and *PCAS* references I mentioned, and both have been applied many times in practice.

An informal survey was done by the General Insurance Study Group Meeting in 1993 at the Institute of Actuaries. The actuaries present were asked whether they used stochastic methods for claims reserving. Forty percent of those respondents who were involved in claims reserving said that they had used a stochastic method in practice at one time or another. Table 2 shows the number using various classes of stochastic methods. The first one is log incremental static methods where you take logs of aggregate payments and fit a model by ordinary general linear modeling. The second one is a dynamic log incremental model that uses the Kalman filter-type approach. The third one is bootstrapping which is a method Spencer has used in the past. Finally, operational time stochastic methods such as those used in my Method B are in the last column. So they have been quite widely used in the U.K.

TABLE 2
NUMBER OF ACTUARIES WHO HAVE USED
STOCHASTIC METHODS IN PRACTICE[a]

	Log Incremental Static	Log Incremental Dynamic	Boot Strapping	Operational Time
Consultants	3	3	1	7
Lloyds/RI	2	0	0	0
Insurers	7	1	2	3
Total	12	4	3	10

[a]Based on informal survey taken at Institute of Actuaries in 1993 by the General Insurance Study Group.

I'm going to talk about Method A, which is the aggregate-incurred-loss method. It's based on a mathematical model of the underlying claim payment process, but it could be regarded as a macro model, it's meant to be applied when you don't have data on individual losses, but you do have an aggregate-paid triangle and an aggregate incurred. Even so, the model is arrived at by considering individual payments (see formula below). The number of individual payments is N_d. It's a number occurring in development period d. The size of an individual payment is X_d. In the model, the expected number of payments occurring in any given period follows a curve of the form shown in Chart 1. It's proportional to the development period to the power of some parameter multiplied by e to the power of some other parameter times the development period. That's a gamma-type curve. The rationale for using that, apart from the fact that empirical studies seem to show that it's a realistic shape, is that a claim is paid when a number of independent processes are completed. First the claim has to be reported; then it has to go through various stages of processing. Each is regarded as an independent waiting process which can generally be well modelled as a negative exponential of the time for an event like that to occur. The sum of a number of independent negative exponentials is like a gamma distribution. So that's where this form of equation comes from. That gives the expected number of claims occurring in any development period and the variance of the number in any development period is proportional to the mean; that is we have Poisson claim numbers. This is a standard model for that type of data.

METHOD A: UNDERLYING MODEL

Number of Claims: $E(N_d) \propto d^a \cdot e^{-bd}$

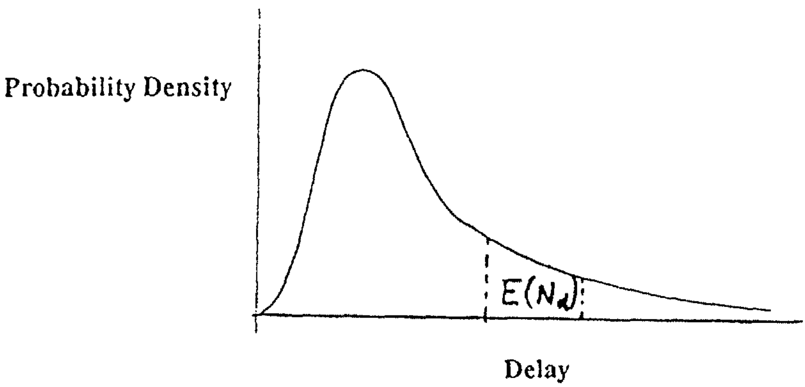
$\mathrm{Var}(N_d) = E(N_d)$ (Poisson)

Size of Claims: $E(X_d) \propto d^\lambda$

$\mathrm{Var}(X_d) \propto E(X_d)^2$ (constant c of v)

For the size of claims, X_d is the size of an individual payment made in development period d. The model allows for the fact that payments frequently tend to increase in size as development progresses because larger claims tend to take longer to settle than smaller claims. So the equation is in the form: expected value of individual claims is proportional to d^λ to the power of some parameter. For the variance of individual claims, the coefficient variation is assumed to be constant. The percentage variation in claim sizes is constant, or in other words, the variance is proportional to the mean squared.

CHART 1
METHOD A—MODEL OF MEAN NUMBER OF CLAIMS



First, Chart 1 is a model for the number of claims. This is a gamma type curve for the probability density for the delay until payment. In a finite time interval representing a particular development year, the expected number of claims closing in that development year is represented by this area, the expected value of N_d.

For the size of individual claims, the model in Chart 2 shows that the expected value is proportional to d^{λ}. So the case when payment sizes don't depend on delay, is given by $\lambda=0$ shown by the straight line. Normally you have something like the curved line where λ takes some value between zero and one so claim sizes tend to increase but not proportionately to the delay.

Combining those two components of the underlying process, claim numbers and claim size, and by doing a little bit of mathematics, you can arrive at the equations shown below. The Y_d here is the aggregate amount paid, the total of payments made in development period d. It turns out that the expected value of that is again a gamma-type curve as a function of d and the variance is proportional to the mean multiplied by d^{λ}. The parameter λ describes how claim size depends on delay.

METHOD A: AGGREGATE PAID LOSS Y_d

$$\mu_d = E(Y_d) \propto d^{B1} \cdot e^{-B2.d}$$

$$Var(Y_d) \propto d^{\lambda} \cdot \mu_d$$

That model looks like Chart 3. The curve is the expected value of Y_d of the total payments made in development period *d*. The other curve shows the variance of Y_d in a case where λ is greater than zero. That is when claim sizes tend to increase with delay. If λ is equal to zero, then claim sizes don't tend to increase with delay. You simply get the variance proportional to the mean so the region enclosed by the dotted lines would be widest where the curve is highest. Because the width of that region is two times the standard deviation, it would be proportional to the square root of the curve. Where λ is greater, then there is a bulge to the right because the claims tend to be larger.

CHART 2
METHOD A—MODEL OF MEAN SEVERITY

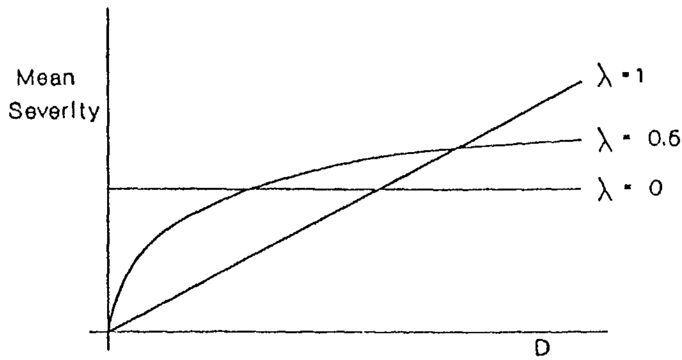
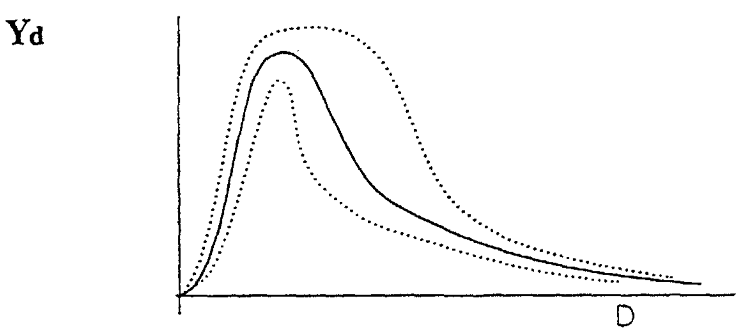
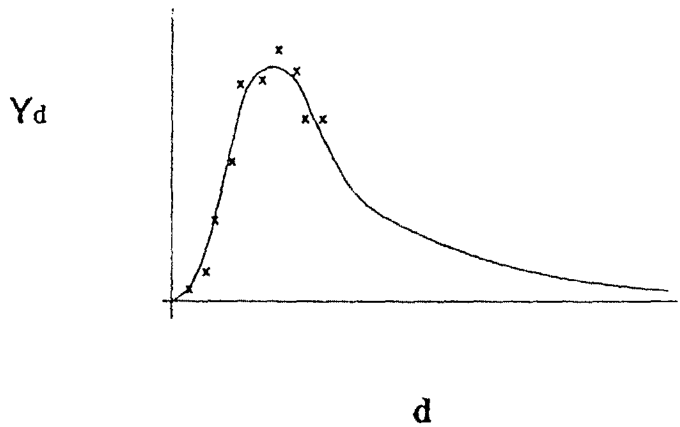


CHART 3
METHOD A—MODEL OF AGGREGATE PAID LOSSES (Y_d)



That model for the variance determines how the gamma curve for the mean should be fitted to the data. Because the variance of each data point can be specified in terms of the mean, you have to use iterative-weighted least squares to fit the model. The weight given to each point is inversely proportional to the variance. Chart 4 shows a curve of some incremental aggregate paid losses and a fitted curve.

CHART 4
METHOD A—AGGREGATE PAID LOSSES: DATA AND FITTED METHOD



There's one other point I should mention. How can we determine a suitable value for λ? The value of λ describes where the variation in the data is greatest and therefore how much weight should be given to the data in fitting the model. We look at standardized residuals which are the differences between the data and the model scaled down according to the variance assumption coming out of the model. If the mo del is right, when you do that you should have constant variance for all the development periods. So you can look at the plots of these residuals to determine if the variance is constant (Chart 5).

So far I've assumed that our data is aggregate paid amounts. You can also use this approach with incurred data and it's generally better to do that if incurred data is available (as it was in the example data set provided by Roger). Claims are reported before they're paid so if you use that data, there's less projection to be done. This can improve the estimates quite substantially. To do that we add outstanding to cumulative paid and then make an adjustment to the outstanding component to allow for the fact that case estimates are biased in general.

CHART 5
METHOD A—STANDARDIZED RESIDUALS VERSUS DELAY (λ=0.0)

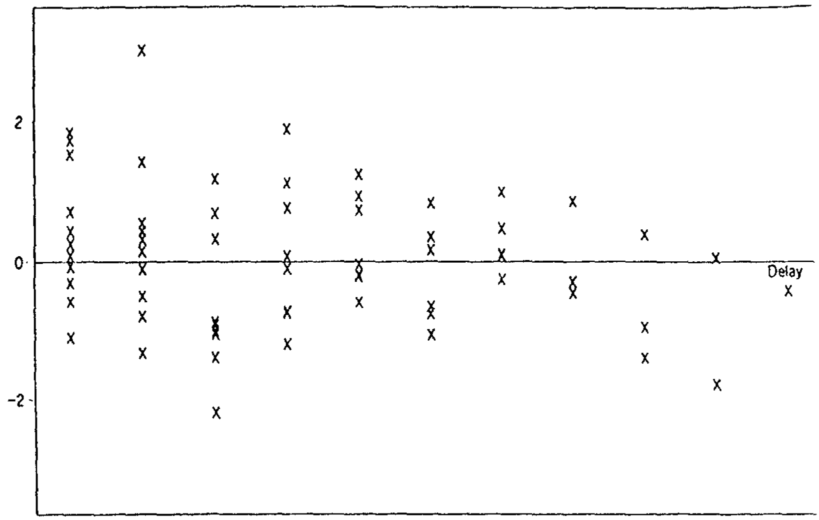


Chart 6 illustrates what I mean. In Chart 6 the bottom curve is the cumulative paid amount for a single origin year. The top curve is paid plus outstanding or what I call incurred. You can see that it reaches a peak at year 5 and then decreases; that's quite common. It starts to decrease when case estimates tend to be replaced by smaller actual paid amounts. Because the model is applied to incremental data (the total amount paid in each development year), and because you're using a gamma curve, you can't have decreases. It has to be positive. We'd like to have the total of all claims reported in that development year, i.e., the total amount actually paid rather than case estimates. To get at that, a bulk adjustment is made to this outstanding component. The middle curve is arrived at in this case by assuming that the outstanding amounts on average have a bias factor of 2. This might seem to be a high factor, but that includes the fact that there will be case estimates set up for claims that are eventually settled for no cost. So a factor of two is not unusual. By assuming a factor of two we get the middle curve which is halfway between the bottom and the top curves. You can look at that middle curve incrementally rather than cumulatively and make a judgement about whether it seems to be following a gamma curve. If it does, then it's appropriate to fit the model to that data. It has a shorter tail than the bottom curve, so you generally get more reliable results.

That's all I have to say about Method A in general terms. Now let's look at the example data set. I used a bias factor of 2 which I've already shown you. In Chart 7 we used a bias factor of 1.5, so the middle curve is arrived at by knocking off a third of the difference, or a third of outstanding.

CHART 6
METHOD A—ORIGIN YEAR 1981 (o/s bias=2.0)

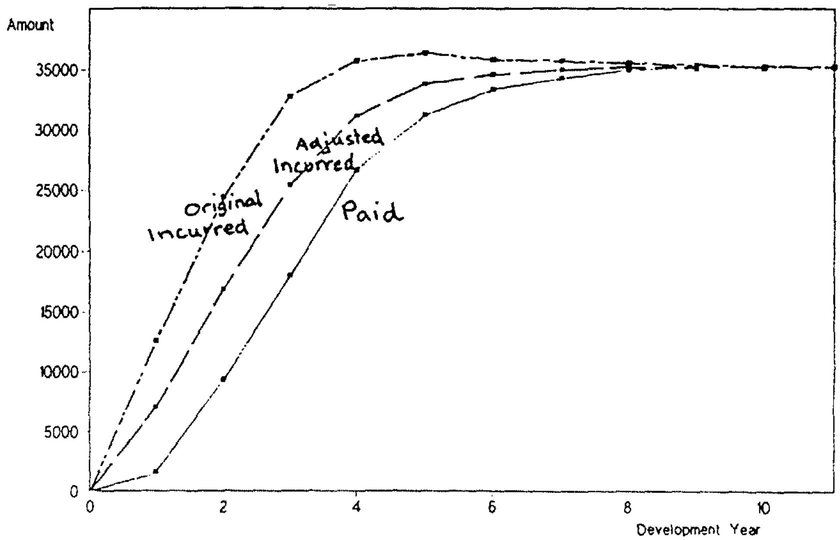
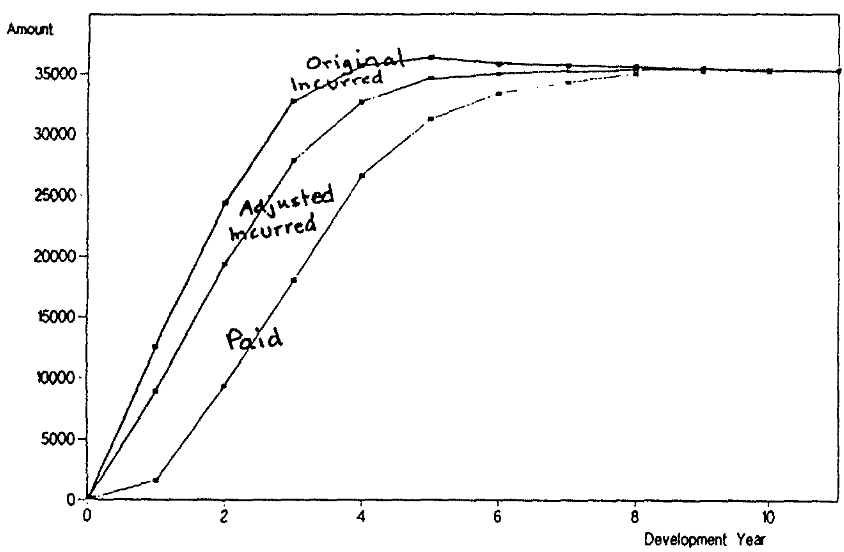
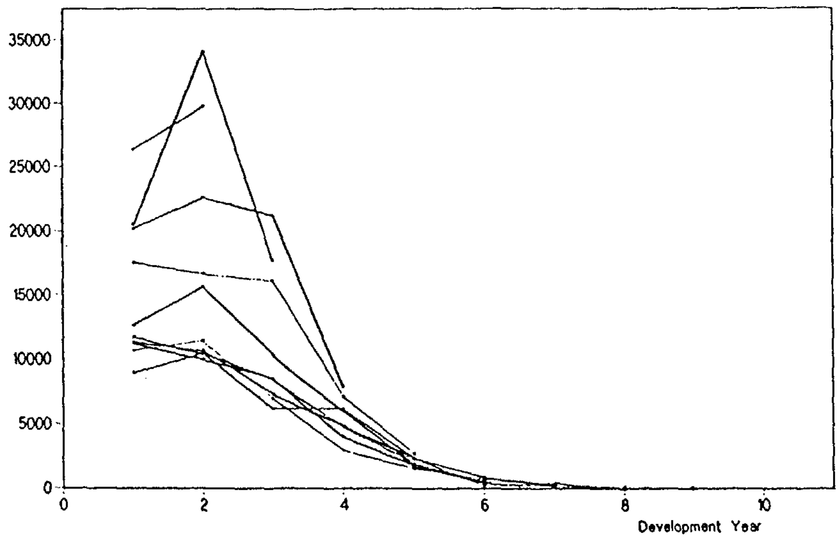


CHART 7
METHOD A—ORIGIN YEAR 1981 (o/s bias=1.5)



Once completed, you can look at the bias adjusted incurred data on an incremental basis for each origin year (Chart 8). It appears to follow a gamma curve. So it does seem to fit the model. This graph is for one origin year, and shows the adjusted incurred data and a fitted curve.

CHART 8
METHOD A—BIAS ADJUSTED INCREMENTAL INCURRED CLAIMS



Then we must decide what parameters are appropriate to describe how the size of losses depends on delay; we look at the standardized residual plots in Chart 5. These plots show the difference between the data and the fitted curve divided by the standard deviation as assumed in the model. So if the model is correct, the standard deviation of these should not depend on the delay. We're looking for a constant variance, the same variance at both ends. Because there are always more points at the left end than at the right end, you would expect a higher spread of points at the left even if the variance is constant. This is a case where I thought that there was an increase in the variance as delay increases, so I refitted the model with the perimeter of λ equal to 0.5 which makes these points come in a bit (Chart 9). The model is assuming that there's a higher variance in the right tail. So when I adjust for that, the variance becomes less; however the results were not very sensitive to the value used for that parameter.

Table 3 is a complete set of results of the example data set using this method. The standard error for each origin year includes both parameter uncertainty and specification and process uncertainty. I'm not going into more details as to how those components are calculated. I've already given you a reference to the paper. For your information, it would be possible to separate those two components.

CHART 9
METHOD A—STANDARDIZED RESIDUALS VERSUS DELAY (λ=0.5)

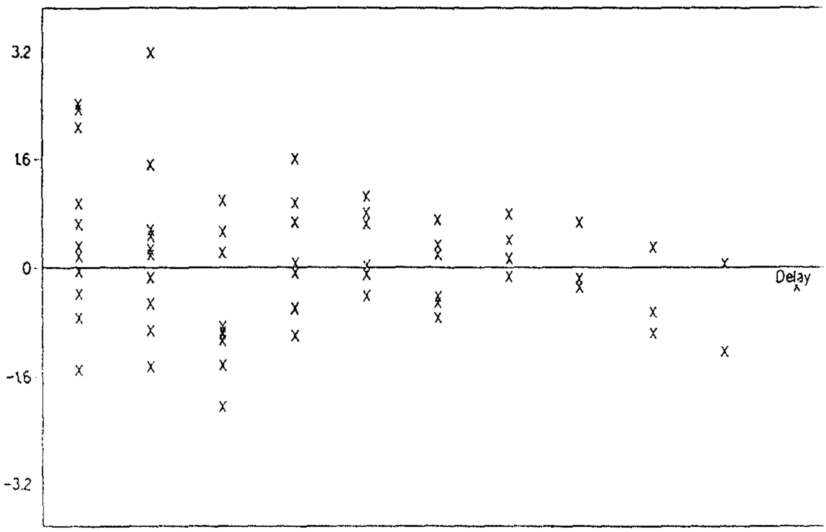


TABLE 3
METHOD A—RESULTS WITH BIAS FACTOR = 1.5 (λ=0.0–0.5)

Origin Year	Reserve	Standard Error
1981	$63	$13
1982	76	26
1983	228	50
1984	133	96
1985	650	184
1986	1,276	374
1987	3,771	750
1988	11,290	1,483
1989	29,840	2,704
1990	64,805	4,877
1991	90,236	7,579
Total	$202,368	$11,629

Method B is the mean claim amount as a function of the operational time method. Again it is based on a model of the underlying payment process. X_d again is the size of an individual payment. The expected value of X_d is some function, any function, of operational time at all. Operational time goes from 0 to 1 and sort of follows up. I don't have a graph of operational time against development time; it is just the cumulative distribution function of delay to payment.

METHOD B: UNDERLYING MODEL

- Mean Claim size is a function of "operational time:" $e(X_d) = m(\tau)$
- Variance of claim size is a power function of the mean: $\mathrm{Var}(X_d) \propto m(\tau)^{\alpha}$
- $\alpha = 2$ for constant c of v

Now the reason for modeling mean claim size as a function of operational time, is because this automatically takes account of changing settlement rates. If claims are settled more rapidly in late origin years, then we don't have to worry about that. If we were modeling against development time, this would be a complication that would have to be allowed for in the model. The other element of this model is that the variance of individual payments is proportional to some power function of the mean. If you want to assume constant coefficient of variables for payment sizes, the index should be set to 2.

Operational time, as I said, is the proportion of claims closed. To calculate the operational time at any stage of development you must have an estimate of the ultimate number of claims. Here's a simplified example of this method. In Table 4 we have just three origin years, and this shows the number of claims closed in each of three development years. In the final column there is an estimate of the ultimate number of claims, obtained by projecting that triangle to each development year. Table 5 shows how these claim numbers translate into operational time for each development period. Our first figure for 1989 is 0.1 because in the first development year of 1989, ten claims were closed out of an estimated total of 50 so at the end of development year one, the operational time is ten divided by 50 or 0.2. So the mean operational time in that development year was 0.1. At the end of development year two, for 1989, a total of 30 claims have been closed out of 50 which is operational time 0.6. We've gone from 0.2 at the beginning of development year two to 0.6 at the end of development year two, so the mean operational time for the second development year of 1989 is 0.4.

TABLE 4
INCREMENTAL NUMBERS CLOSED

Year	1	2	3	Total
1989	10	20	10	50
1990	20	20		100
1991	30			45

TABLE 5
AVERAGE OPERATIONAL TIMES

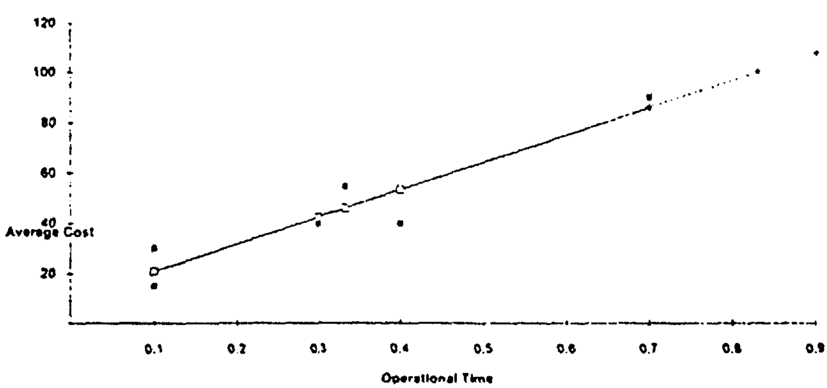
Year	1	2	3
1989	0.10	0.4	0.7
1990	0.10	0.3	
1991	0.33		

That was claim numbers. We also have aggregate paid losses shown in Table 6. This is also an incremental triangle, not a cumulative one. By dividing those amounts by the number of claims closed, we get this triangle of the average cost per claim or per payment. From the two triangles, you can plot the average cost per claim against the average operational time for each cell (Chart 10). Our model is that the average cost per claim is some function of operational time so we have a curve. In this case a straight line seems to fit reasonably well, and we can project that curve to calculate the reserves. For any claim not yet paid, we can calculate its operational time and just read off the expected amount of the payment from this fitted model. Of course there is also a variance assumption in the model so it also gives you the variance of each payment yet to be made. We can calculate the expected amount and the variance of all outstanding payments.

TABLE 6
AGGREGATE PAID LOSSES

Incremental Loss Amounts			
Year	**1**	**2**	**3**
1989	150	800	900
1990	600	800	
1991	1,650		
Average Costs			
Year	**1**	**2**	**3**
1989	15	40	90
1990	30	40	
1991	55		

CHART 10
AVERAGE COST AGAINST OPERATIONAL TIME



This is a very simple example. That process of estimation is very simple as shown in Table 7. For 1989, we have 40 claims closed so far, and an estimate of 10 outstanding claims. The average future operational time is therefore 0.9. Because a straight line model was used, we can just use the average, rather than calculating the operational time for every future payment. The average future cost of the 10 outstanding claims can be read off of the previous graph showing the fitted model. It's the fourth column. Then just multiply the number outstanding by average future cost to get the reserve. It's a very simple example, but it gives you some idea of how the method works.

TABLE 7
RESERVE CALCULATION

Year	Number of Claims		Average Future Operational Time	Average Future Cost	Reserve
	Closed	Outstanding			
1989	40	10	0.90	107.50	1,075
1990	40	60	0.70	85.85	5,151
1991	30	15	0.83	99.91	1,499

The assumptions about the mean and variance of individual claim payments translate through a bit of simple mathematics into a model for the sample mean for the data actually available—the total aggregate paid divided by claim numbers. You can fit that model by iterative-weighted least squares and then look at the residual plots again to see whether the variance assumption is about right; adjust the index in the variance model if necessary. First, you get the variance assumption right through that process and then you can try fitting different models for the mean as a function of operational time such as a cubic or some other polynomial function or whatever other function seems to fit.

You can use standard statistical tests to compare different models. Once the variance assumption is correct, you can calculate ratios of the residual sum of squares and do F-tests to determine which model most closely follows the data. Then you can use the best model of mean as a function of operational time to calculate the total mean of all future claims by totaling the prediction for all future claims.

Fitting by iterative-weighted least squares (also known as maximum quasi-likelihood estimation) gives you the standard error of each parameter as well as giving you parameter estimates. You can use those to calculate the parameter uncertainties in the final estimates. So that's the parameter risk in the three components of risk that Roger was talking about.

For the second component, process risk, we have a model for the variance of each individual claim amount. It is proportional to the mean to the power of alpha. The constant proportionality can be determined from the magnitude of the standardized residuals. Having done that you can sum this over all future claims to calculate the future process variance attributable to claim sizes. There is also another component of future process variances to deal with; the number of future payments is uncertain, and that can also be taken into account. It's a lot more complicated; you will have to look at the paper to get the details.

Here are the final results of applying this Method B to the example data set Table 8. Reserves and standard error of prediction include both parameter uncertainty and process uncertainty. The model allows this standard error to be broken down into its component parts as shown by the four columns of Table 9. The first one is the parameter uncertainty. The second one is the standard error due to uncertainty about future monetary inflation, something I haven't discussed. Again, you'd have to read the paper to get more information. There are two components of process uncertainty in these final two columns: the uncertainty due to variation in the size of individual claim payments and the uncertainty due to variation in the number of future claim payments. For the overall standard error on the previous table, those four elements are all mutually independent, so the overall standard error the square root of the sum of the squares of the four components.

TABLE 8
METHOD B RESULTS

Origin Year	Reserves	Standard Error of Prediction
1981	$ 58	$ 135
1982	117	198
1983	255	303
1984	549	452
1985	1,238	690
1986	2,993	1,106
1987	7,120	1,788
1988	16,407	2,881
1989	33,468	4,443
1990	63,493	6,708
1991	84,611	8,007
Total	$210,307	$13,273

Briefly I'll go through graphs of how this method was applied to the example data set. Chart 11 shows just the data. It's a graph of the mean claim amount against operational time. In this example, we had some years that were fully developed so the data goes right up to operational time one. There's one line for each origin year, and you can see a clear pattern.

Chart 12 is a deliberately overparameterized model fitting that data. The idea is to quantify the magnitude of the random variation in the data in order to get a correct variance assumption, and to see how the variance depends on the mean claim size. The variance appears to be increasing at this end, but that's partly due to the fact that the numbers of claims are smaller. Each of the data points is a sample mean. It's the total amount paid in a particular development period divided by the number of nonzero payments. The variance is greater near operational time one partly because there were smaller sample sizes—a smaller number of claims in each development year for the later operational time. That's taken into account in the model.

TABLE 9
METHOD B RESULTS CONTINUED

Origin Year	Parameter Uncertainty	Inflation Variation	Severity Variation	Claim No. Variation
1981	$ 5	$ 1	$ 89	$ 104
1982	10	1	121	156
1983	22	3	180	242
1984	47	6	264	363
1985	102	14	397	554
1986	227	35	618	888
1987	473	84	955	1,434
1988	850	197	1,456	2,328
1989	1,236	407	2,108	3,689
1990	1,757	781	2,980	5,694
1991	2,130	1,151	3,443	6,811
Total	$6,236	$2,680	$5,373	$10,061

CHART 11
METHOD B—INFLATION ADJUSTED MEAN CLAIM AMOUNTS

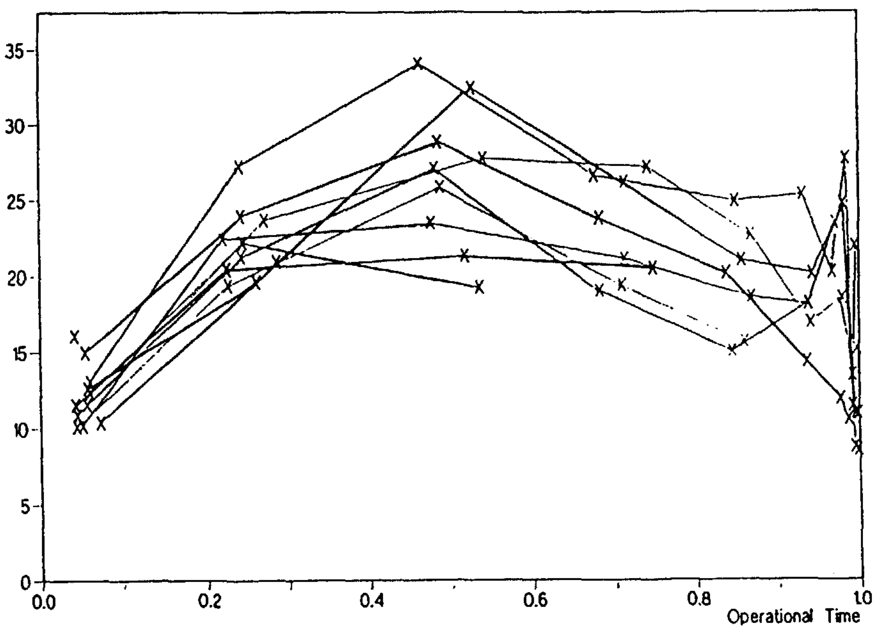


CHART 12
METHOD B—DATA AND FITTED MODEL ZERO

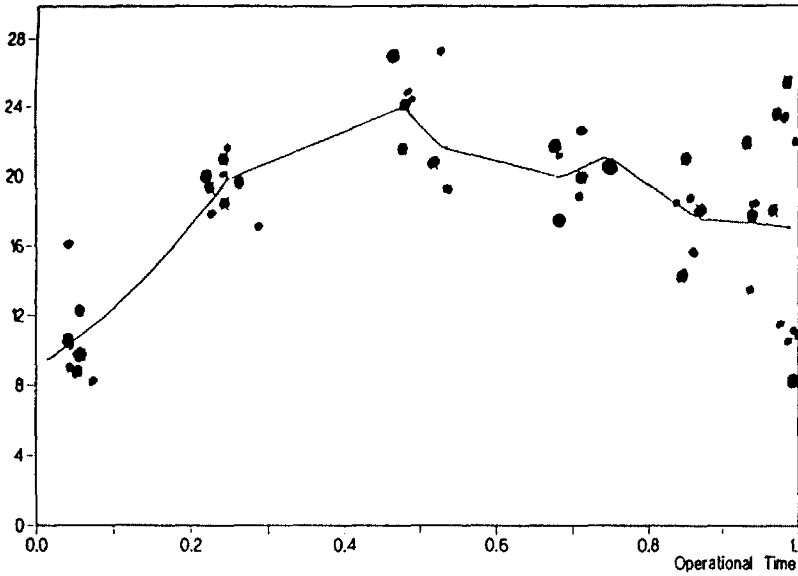


Chart 13 is a graph of standardized residuals versus operational times. We see that when the residuals have been standardized (scaled down to allow for the variance as dictated by the model), the spread is less at the top end than it was on that graph of the data. That's because standardization allows for the smaller sample sizes in each sample mean for the later operational times.

What we're really interested in is looking at the standardized residuals against the fitted mean, because the assumption is that the variance depends on the fitted mean, (proportional to the fitted mean to some power alpha). So to verify that alpha is correct, we should see constant variance going from left to right in Chart 14 where alpha equals 2.0. I thought there was decreasing variance as the mean size increased, so I tried it again with alpha equal to 1.6 which helped a bit (Chart 15). The results, as I said at the beginning, were not very sensitive to the value used for that variance index. So you can see from the data that the range 1.6–2 is about right and as every value in that range gives a similar final result, it's not worth worrying about exactly where the value is in that range.

Chart 16 shows the data with the initial, deliberately overparameterized model. The smooth curve is the final curve that I fitted to a cubic equation. The initial curve had about eight parameters; the smooth one had about three. The variation of the initial curve around the smooth one is quite small compared to the size of the random variation that's going on in the data. This is formalized by the statistical tests. I obtained a small F-statistic showing that the smooth curve fitted well compared to the initial one, so I used that curve to calculate the results.

CHART 13
METHOD B—STANDARD RESIDUALS VERSUS OPERATIONAL TIME ($\propto = 2$)

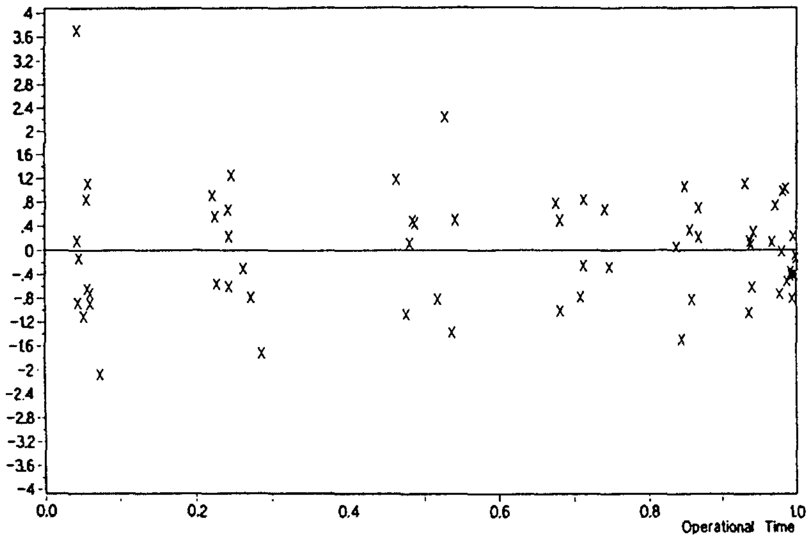


CHART 14
METHOD B—STANDARD RESIDUALS VERSUS FITTED MEAN ($\propto = 2$)

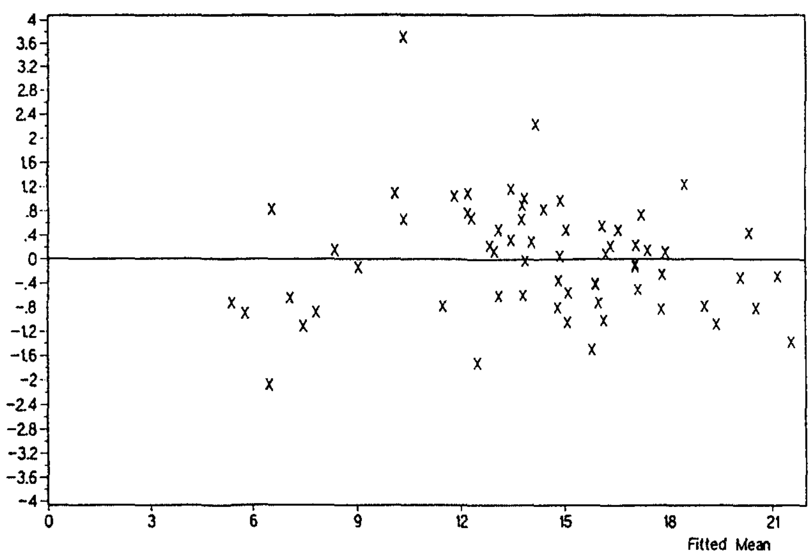


CHART 15
METHOD B—STANDARD RESIDUALS VERSUS FITTED MEAN ($\alpha = 1.6$)

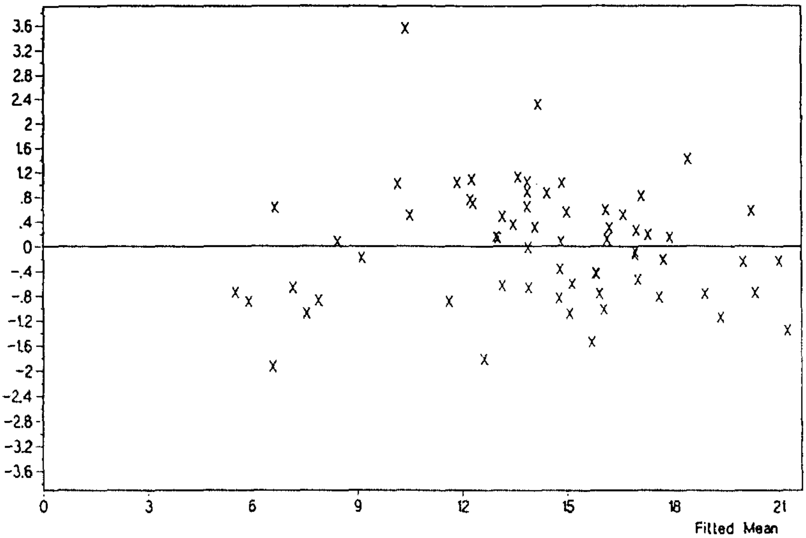
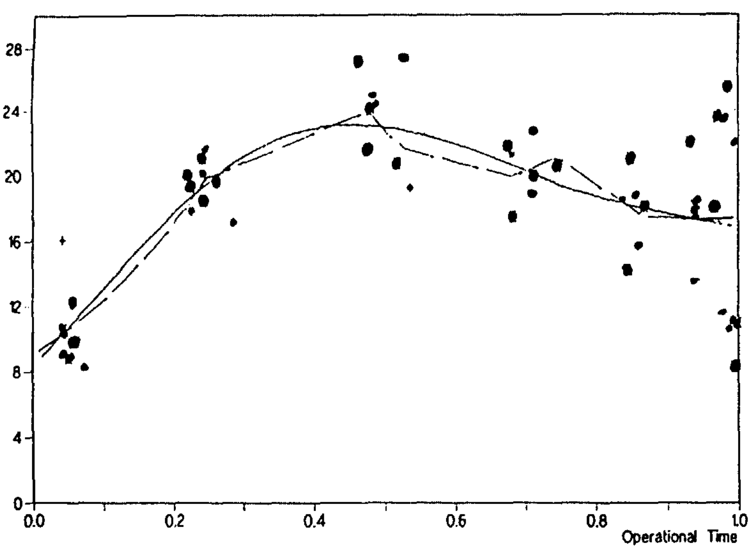


CHART 16
METHOD B—FITTED MODEL ZERO AND FINAL MODEL



MR. SPENCER M. GLUCK: One of the disadvantages that I have is that Tom and I have been talking about these models for years. We have relatively similar models, so I'm going to have to try to gloss over the points that have already been made and I'll emphasize some of the differences and show you how I do my model fitting. When I emphasize the differences, of course it's necessary for me to talk about Tom's models.

None of us are measuring specification error. In the U.S., many people talk about process and parameter risk as if they are the only two risks. I think that custom developed in the context of ratemaking where the thing you're projecting is a future mean described as a parameter. So you just distinguish between the variance around that mean, the process risk, and the possibility that the mean is wrong which is then described as the parameter risk. But in this case, when we're talking about a modeling framework, we're talking about the parameters of the models themselves and the projection could be wrong because the model structure is wrong and because the parameters themselves are wrong. That's why we have the specification error. The difficulty in specification error is that you can't measure it directly because all the means that we have of measuring error presume the model structure to be correct. I think specification error may well be very large. One of the reasons that it's good to have a number of different models is to give you some feel for how large the specification error might be. If you have a number of different models giving you the same results, then you appear to be well specified. If they give different results, that gives you some measurement or at least some feel for specification error.

The disadvantage of regression models is they are not specifically connected to the collective risk model, although Tom has made a connection. We measure process risk in the regression model by just looking at the model, looking at the summarized data and determining how much variation there is in the summarized data. Of course, because we have a whole model structure involved, we have a difficulty with specification error.

The advantage is that we start with data in a triangular form; we model it in a triangular form, and therefore we have methods for the whole distribution of errors, not just the standard error. We can actually look at that distribution at any given point for every individual cell in the future projection period. For example, in the typical reserve analysis framework, we're really interested in the error of the total reserve. But if it's a payment model, you'll have a projection of the payments in the next calendar year. For example, if you were trying to use those results in an investment strategy, you might be interested in the distribution of likely results by calendar year of payment and you can get that. In another case, it might be particular accident years only. If you were in a ratemaking context, you would be more interested in the recent accident years.

The second one is very significant in that, yes, we want to know the projection intervals for their own sake to know what their uncertainty is, but here the projection intervals that you come up with are also a diagnostic on your model. It's directly tied to how successful your model has been. Then of course when we're in a modeling framework we can directly measure the parameter estimation error.

My models are going to look like Tom's models, but there are some differences. The differences are more in the variance structure. The basic model is for in-period payments. This formula is basically the same as Tom's. The first term, the alpha, is simply a scale parameter. Then we have power term and an exponential term and then finally the error term which is multiplicative in this case because I'm going to use a log transformed model. So in this case, I'm talking about percentage errors log normally distributed. The scale parameter, α_{ij}, means there's a different scale parameter for every point in the triangle. Of course, if that were true, then you'd have as many parameters as points before you even reached the other parameters. So we're going to simplify that structure to some extent. You can't really use all those parameters.

$$P_{ij} = \alpha_{ij} \cdot (1 + j)^{\beta} \cdot e^{\delta j} \cdot \epsilon_{ij}$$

i = Year of Origin (74 . . . 91)

j = Delay (0 . . . 17)

Calendar year of payment $= i + j$

α_{ij} are scale parameters

$\beta + \delta$ are shape parameters

ϵ_{ij} are multiplicative errors

The formula below shows the scale parameter structure so that I don't have as many different parameters as it appears. This is again more parameters than I'll tend to use in the real model. I can have a free scale parameter for every accident year, and I can have a free parameter for every calendar year of trend. You notice the calendar-year parameters start in 1975 rather than 1974 because the 1975 parameter is the percentage trend from 1974 to 1975. If each of those parameters were different, then the scale term for any one cell would be the accident-year parameter times the product of all prior calendar-year parameters up to that point.

The starting point model is the simplest one. I assume that one calendar-year trend can account for all the variations in scale. In that case, I just simply have a single calendar-year trend parameter. I drop the subscript so the scale term associated with the single point is the product of the one accident-year parameter (it's really just a general scale parameter for the whole model now), and the trend parameter for the appropriate number of years. I will always divide the paid data by either an exposure count or an ultimate claim count, so it becomes reasonable to adjust only for trend in a good stable data set. This was not a good stable data set. Roger made sure of that for us.

SCALE PARAMETER STRUCTURE

Accident Year Scale Parameters:

AY_{74} .... AY_{91}

Calendar Year Trend Parameters:

CY_{75} .... CY_{91}

$$\propto_{ij} = AY_i \cdot \prod_{k=75}^{i+j} CY_k$$

If one AY parameter and one CY parameter:

$$\propto_{ij} = AY \cdot CY^{i+j-75}$$

In the formula below I've inserted the scale parameter structure where I previously just showed $\propto_{ij}$. Then we have the shape parameters to the right. Then I take a log transformation so this all becomes linear in the parameters. You see the error term now additively, but because it's a log the $\in_{ij}$'s are log normally distributed errors, meaning there is no probability of readings of zero or less. So if you have a data set with readings of zero or less, you have some trouble and you have to deal with it. It's usually reasonable to do that for in-period payments. This is in contrast to Tom's model in which he's starting from the collective risk model and coming up with a specific error structure. Here we're just presuming percentage errors. Because the model forms generally don't go to zero or negative, his model can't handle zero or negative payments that are systematic. Because he's not using log normally distributed errors, it will handle a zero or negative payment as a random event but, he's still constrained to a form if the mean that he's fitting is strictly greater than zero.

Combined Model:

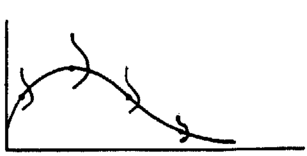
$$P_{ij} = AY_i \left[\prod_{k=75}^{i+j} CY_k \right] \cdot (I + J)^{\beta} \cdot e^{\delta j} \cdot \epsilon_{ij}$$

Log-Transformed:

$$\ln P_{ij} = \ln AY_i + \left[\sum_{k=75}^{i+j} \ln CY_k \right]$$

$$+ \beta \ln (I + j) + \delta_j + \ln \epsilon_{ij}$$

ϵ_{ij} are % errors, log-normally distributed



MR. WRIGHT: It is true that the fitted mean in my method is strictly greater than zero, but there is not any assumption of normally distributed errors in method A or method B. Both methods give first and second moments for the reserve, and to do this, only first and second moment assumptions are necessary for the data. There are no higher order distributional assumptions.

MR. GLUCK: I'd like to be able to model, as Tom does, not only paid data but incurred data. When you look at it incrementally, incurred data often will have negative or zero amounts in it. If it's systematically going negative or zero, I basically use the same approach that Tom did. I multiply all the case reserves by a factor between zero and one sufficient to eliminate the systematic negative development so that I can then model it.

In using an incurred data triangle, we make an assumption of constant relative adequacy in case reserves. Obviously if I multiply all the case reserves by the same factor, I have not interfered with that assumption. So I don't think I've necessarily added a bias of any kind by putting that factor in. I've only moved toward the paid data analysis. If you put in that factor at zero, you would simply wind up with the paid data and you'd still have a valid basis to model.

MR. WRIGHT: Spencer seems to have misunderstood what I mean by *bias adjustment* of incurred data. The intention is not to introduce bias, but remove it. Cumulative incurred tends to decrease in the right tail of the run-off when case estimates tend to exceed the amount actually paid. In statistical terms, case estimates have a positive bias. The purpose of the bias adjustment is to remove this bias.

MR. GLUCK: I also apply a number of hybrid models. Rather than modeling the data directly, I apply the same model to the development factors minus one. Then, of course, that means that I have to go through another step to get the ultimate losses. There I'll use deterministic methods. I use something I call the generalized Cape Cod method once the development factors are modeled, to get to the ultimate losses.

The Cape Cod method, shown below, is reasonably well known. You're projecting a single value of the ultimate losses per exposure. (This would only make sense after the losses have been corrected for trend.) You calculate it by the ratio: a sum of the known losses is in the numerator and a sum of the exposures divided by the appropriate development factor is in the denominator. The bottom equation is the same equation.

I've just expanded it because I think it's a little more descriptive. It really shows a weighted average of the losses developed to ultimate for each accident year divided by the exposure. The weights are proportional to the exposures and inversely proportional to the development factor. They're proportional to the exposures because more data deserves more weight. They're inversely proportional to the development factor because the development projection is less reliable if the development factor is larger.

CAPE COD METHOD

Notation

LTD_i = Losses to date for year i

DF_i = Development Factor to Ultimate

ULT_i = Ultimate Losses

EXP = Exposures

$$E\left[\frac{ULT_j}{EXP_j}\right] = \frac{\sum LTD_i}{\sum \left[\frac{EXP_i}{DF_i}\right]}$$

$$= \frac{\sum \left[\frac{EXP_i}{DF_i}\right] \times \left[\frac{LTD_i \times DF_i}{EXP_i}\right]}{\sum \left[\frac{EXP_i}{DF_i}\right]}$$

This particular approach is analogous to using a single-scale parameter to describe the pure premiums for all accident years. Now sometimes that one scale parameter idea doesn't seem to work that well. This applies not only to the stochastic model, but any time you're using the Cape Cod method. You frequently find that if you're applying it over a long period of time, you wind up giving too much weight to out-of-date data in order to project your recent years.

So I've accounted for that by adding another term to the weight: a "decay" factor (between zero and one) which is then taken to the power of $i-j$, which is the lag *from the accident year that you are measuring the pure premium from to the accident* year that you're projecting to. This winds up allowing the value of the expected pure premium now to be a function of the year. It's not constant for all years, and it allows it to drift to some degree. This is very much analogous to what's done in dynamic modeling where a parameter estimate is allowed to drift.

GENERALIZED "CAPE COD" METHOD

$$E\left[\frac{ULT_j}{EXP_j}\right] =$$

$$\frac{\sum \left[\frac{EXP_i}{DF_i} \times \text{Decay}^{|i-j|}\right] \times \left[LTD_i \times \frac{DF_i}{EXP_i}\right]}{\sum \left[\frac{EXP_i}{DF_i} \times \text{DECAY}^{|i-j|}\right]}$$

$$0 \leq \text{Decay} \leq 1$$

Special Cases:

Decay = 1 ⇒ "Cape Cod" Method

Decay = 0 ⇒ Development Method

Basically, if you set the decay factor to one you're using the standard Cape Cod method and if you set the decay factor to zero then you're using a straight development approach which is analogous to every accident year having a free parameter.

When you use something in between zero and one, you wind up in effect with a fractional number of parameters.

I also apply an operational time model where everything is exactly analogous to Tom's method. Operational time is the number closed divided by the ultimate number. T_{ij} is the operational time at the midpoint of the period. As Tom said, you could use any number of curves to relate the mean claim size to the operational time. I use the same curve that we use for in-period payments—the one with the power term and the exponential term. Again I allow this factor ($\propto_{ij}$) so I can make as complex a scale parameter structure as I need to fit the data. In general, I'd like the scale parameter structure to be as simple as possible. The more complex I have to make it, the less confidence I can have that my model is really valid. If I've had to use many parameters and do a lot of monkeying with it to make it fit, then I may get something that looks like it fits well at the end of the day, but I'll have less confidence in it.

Operational Time Model

Operational Time $= \tau$

$= \#$ Closed/Ultimate $\#$

$\tau_{ij} = \tau$ at midpoint of period ij

Mean claim size in period ij
$= \propto_{ij} \cdot \tau_{ij}^{\beta} \cdot e^{\delta \tau ij} \cdot \epsilon_{ij}$

There are some the diagnostics to look at. The scatter plots are very important. I look at the scatter plots in three dimensions: along the accident-year axis; along the development-year axis (to see if the development pattern is fitting well), and along the calendar-year axis, which is probably the most important one, to see if my model appears to fit well over time. The scatter plots give you something that you don't get out of any of the fit statistics. Sometimes you can have a very tight fit, but when you look at the scatter plots you see a clear trend in the residuals that indicates there is something going on that you didn't model. You might have a great R^2, but you really don't have a good result if you don't have at least a reasonable belief that the model is valid. That validity means you need random looking errors, especially errors that are random over time. I do look at fit statistics, R^2. The adjusted R^2 is an adjustment for the number of points you have to model relative to the number of parameters.

The more parameters you use, the more you penalize the R^2. I look for heteroscedasticity. Is the relative variance constant in the delay direction? If it is not, I make a correction for it. I have a particular diagnostic statistic. I also take a look at autocorrelation coefficients. The problem is if they don't look good, I don't know how to fix it. But I guess if they don't look good it does give me some indication that again the validity of the model is in question. On selected models, I go through some process to look at projection intervals and standard errors and that's of course a diagnostic of the model fit as well.

The adjusted R^2 formula is relatively simple. You take the difference between R^2 and 1 and you penalize it by this value N plus P over N minus P where N is the number of points and P is the number of parameters. I still feel that it doesn't penalize it enough. This is just a general thing on number of parameters, but it doesn't really take into account the structure of the parameters. For example, if you set a free parameter for just the latest accident year alone, that only counts as one parameter. The reality is you have a great deal of uncertainty in that latest accident-year projection. If you have to set that parameter free it creates a lot of error in the projection for the latest accident year. We do pick that up in our standard errors.

Let's discuss the models I fit to this data. The first one is a model of the in-period payments. Tom didn't model the in-period payments. I like to model them without the outstanding losses as well. More models gives you some better feel for specification error. The hybrid models are the ones where I modeled the development factors. I did that once with the paid, and twice with the incurred. In one, I used the incurred data as is. I had positive development in only four columns. After the fifth column, everything turned negative, but it was really slightly negative so I ignored that. That might give me a biased high answer. In the fourth one I hit the case reserves with a 0.7 factor. That gave me positive development for I think it was seven columns, so I had more data to model. Finally, there's the operational time model.

Basically I have found that those curves don't fit that well to in-period incurred data. Even if it's strictly positive, they drop down to the axis a little too fast for the curve. So I have a little more success when I model incurred development factors, because then I pick up that extra scale parameter for shape.

Shown below is an example of some of the first basic output I get. This tells me, this is model one, the paid one. Here I switched the trend after seven years because this data set was not consistent at all. In the last eight or nine years of the data set, we fit a trend of -1.5% which is odd.

Notice that I freed up many accident years. This is because the paid models really didn't fit very well and these would not be models that I chose. The only way to make the model fit was to allow the accident years to go free which leads to a large standard error. Also the negative calendar-year trend is not as disturbing as you might think when you hold it up against the accident years. You'll see that the parameters are negative because we modeled logs of data. They are increasing so that you're picking up some of the trend in the accident year; that's why the residual calendar year trend is negative. The R^2 on this model is very high, but many of the other diagnostics and the number of parameters I had to use would lead me to conclude that that model was really not performing terribly well.

Statistical Outcome of Regression
Model in use: HOERL2
Calendar parameter coefficients and t-statistics:

A	0.460	14.312
B	−0.015	−1.218

Accident-year parameter coefficients and t-statistics:

1	−3.053	−26.333
2	−3.361	−25.109
3	−3.458	−22.095
4	−3.933	−20.703
5	−4.319	−20.539
6	−4.364	−19.938
7	−4.228	−17.706
8	−3.900	−15.916
9	−3.802	−15.120
10	−3.819	−14.758
11	−3.529	−12.538

Coefficients of curve:

Values	3.992	−1.401
t-statistics	33.601	−41.355
R^2	0.995	
Amemiya's Adjusted R^2		0.994

Shown below is one of the incurred models. The R^2s are lower, but you're modeling a smaller piece of data on the incurred. We're modeling incurred but not reported (IBNR) reserve, the unemerged losses. So you could have a larger percentage error, which you're going to see in these lower R^2s. When the error is expressed as a dollar amount, you may still have a more accurate projection.

Statistical Outcome of Regression
Model in use: HOERLCC
Calendar parameter coefficients and t-statistics:

A	−0.026	−2.364

Accident-year parameter coefficients and t-statistics:

1	0.111	0.899

Coefficients of curve:

Values	0.568	−1.484
t-statistics	1.297	−7.125
R^2	0.948	
Amemiya's Adjusted R^2		0.939

I also have to go through a conversion process. Each model has a conversion process. When it's just incremental pay data, the conversion process is simple. It takes the logs. When I'm modeling the development factors, the conversion process is more complex. I'm calculating development factors, taking the logs of those. Then the deconversion process puts the logs back into development factors and uses the Cape Cod model. So I really have to go through the two processes if I want to put these things on a comparable basis. I go through the whole deconversion process to look at the actual data, deconvert it, and recalculate the R^2s based on those data. Also here I can make a correction for more parameters. So in this particular development factor model there were nine parameters, but then the Cape Cod method is outside the regression, and its additional parameters have been added. Here I use the

Cape Cod decay of 0.75 which can be calculated as the equivalent of 3.375 additional parameters. So I adjust the R^2, accounting for the parameters that I introduced outside of the regression model as well.

Statistics Relating to Logs of Deconverted Data	
Model in use: HOERLCC	
Weighted Mean Error:	−0.003
R^2:	0.952
Amemiya's Adjusted R^2:	0.943
Number of points:	138.000
Number of regression parameters:	9.000
Number of Cape Cod parameters:	3.375

My best performing model was the operational time model:

Statistics Relating to Logs of Deconverted Data	
Model in use: OPTIME1	
Weighted Mean Error:	−0.005
R^2:	0.981
Amemiya's Adjusted R^2:	0.979
Number of points:	156.000
Number of regression parameters:	8.000
Number of Cape Cod parameters:	0.000

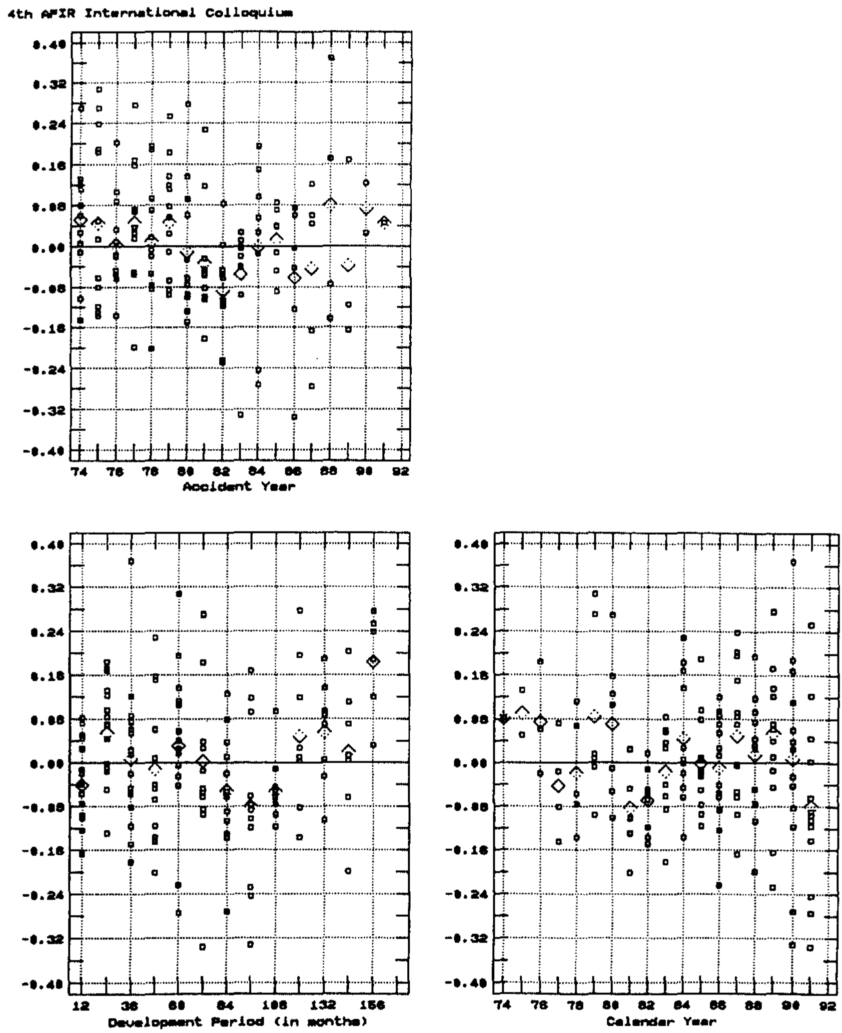
Let's discuss the fit statistics on the competing models. The first two models shown in Table 10 are models of the paid. As you can see by the large errors, I had the least success with them. I also think the other diagnostics, and the way I had to overparameterize them to make them fit, casts doubt on them. The next two are the incurred models which both fit well. The operational time model also fits well. It had the best R^2 statistics and one of the lowest standard errors.

TABLE 10
COMPARISON OF FIT STATISTICS

	1	2	3	4	5
Converted Data					
R^2	99.5	98.1	94.8	96.9	82.2
Adjusted R^2	99.4	97.8	93.9	96.3	80.2
Deconverted Data					
R^2	97.7	95.2	82.5	95.0	98.1
Adjusted R^2	97.2	94.3	76.6	92.1	97.9
Standard Error	21.1	17.7	10.0	13.4	10.5

Chart 17 is the scatter plot of the paid model. First is the scatter plot against the accident year. This is not bad performance, but I did have to use many accident-year parameters to make it do this well. The diamonds are at the weighted mean at each point. These are weighted errors. We weight for two reasons. I weight every point by the ultimate exposures or reported counts for the accident year; it's just a size weight. Then I may put in a heteroscedasticity model. That becomes an element of the weight as well. The square error should be inversely proportional to the weight. So in looking at these plots, and to put the errors in comparable terms, we have to look at weighted errors. The chart shows the accident-year direction. The development direction chart shows whether I fit the development pattern reasonably well. In the calendar-year direction those diamonds, especially as we look in the more recent calendar years, are pretty close to the line. On the other hand, at least by eye, it looks like we have some spread this way so that we've got more variation in the fit for recent calendar years.

CHART 17
RESIDUAL PLOT VERSUS TIME CONVERTED DATA
BODILY INJURY LIABILITY IN PERIOD LOSS PAYMENTS



In Chart 18 the fit is to the development factors. These are just the pure regression fits and only tell you how well the development factors are fitting. It's not telling you how well the ultimate losses are fitting after you run them back through that Cape Cod model.

CHART 18
RESIDUAL PLOT VERSUS TIME CONVERTED DATA
BODILY INJURY LIABILITY IN PERIOD LOSS PAYMENTS

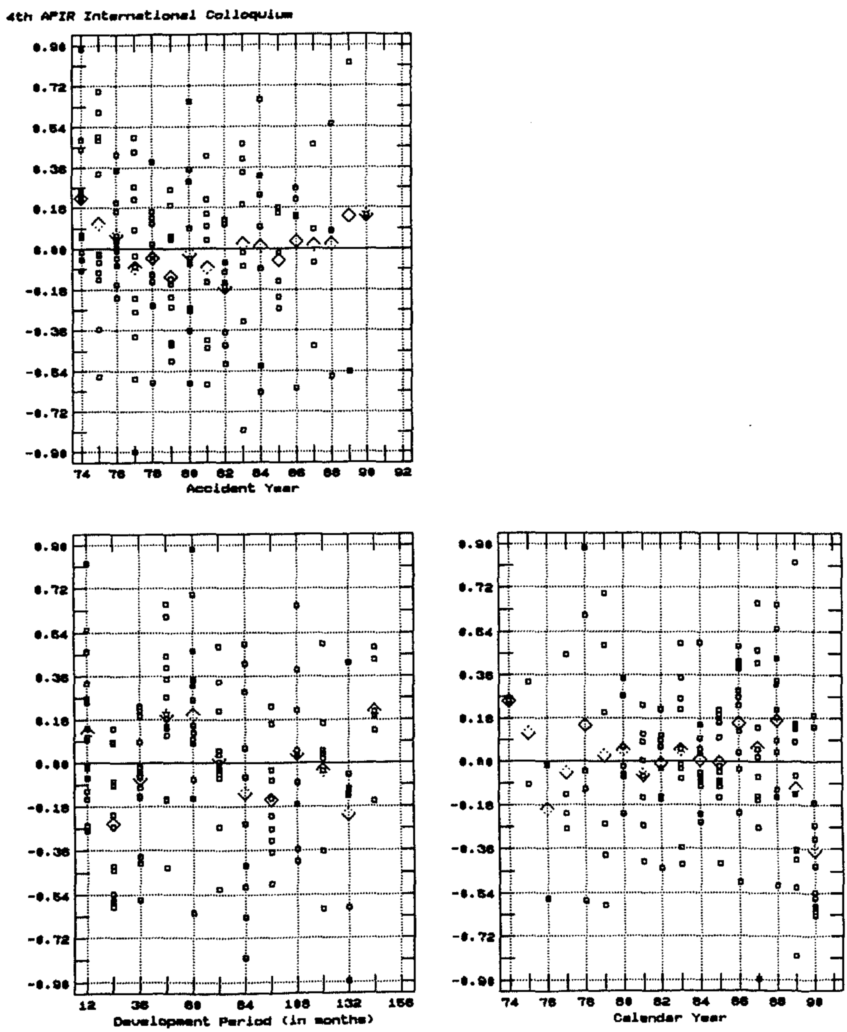
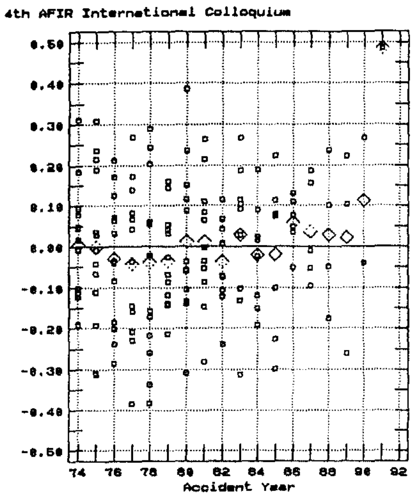
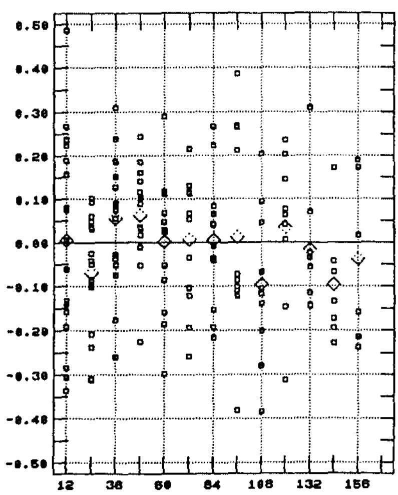
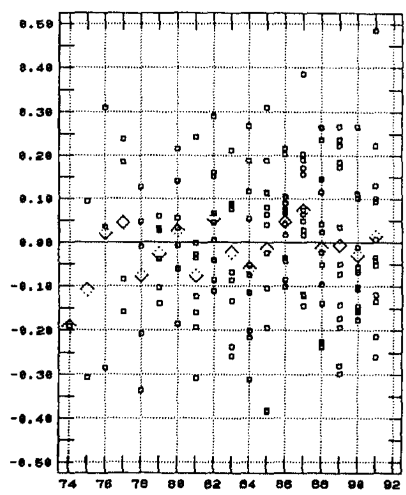


Chart 19 is the best model, the operational time model. These all look reasonably good and reasonably random, which is what you're looking for.

CHART 19
RESIDUAL PLOT VERSUS TIME CONVERTED DATA
BODILY INJURY LIABILITY IN PERIOD LOSS PAYMENTS

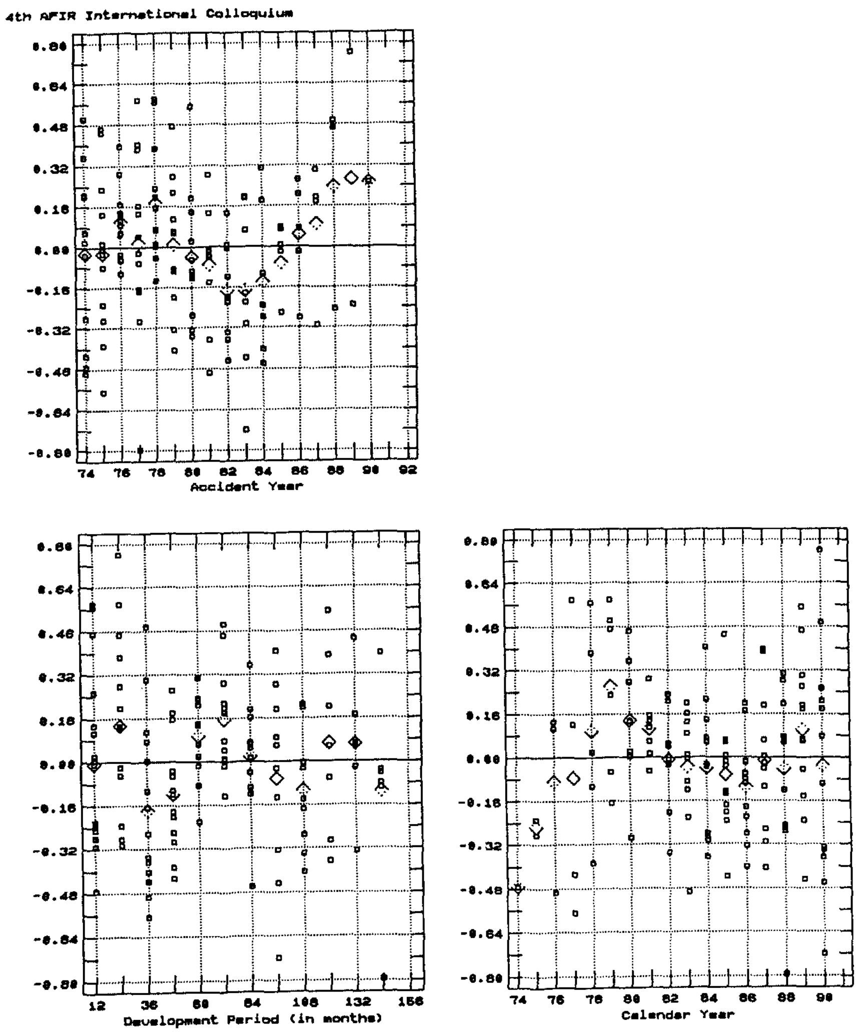






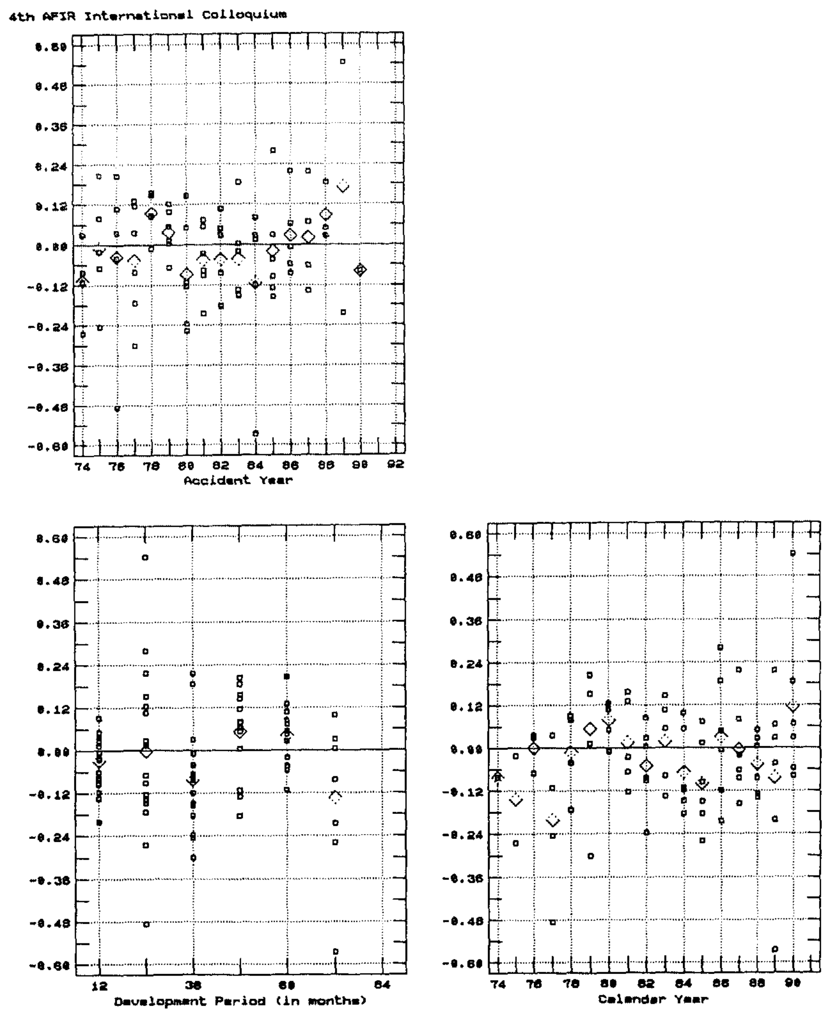
In Chart 20, I have the deconverted data. These are especially important for the models with development factors if you really want to see the whole process. On the accident year chart, you can see that this one really didn't perform very well; if you look at the pattern of those diamonds, you'll see that we didn't have a good fit to the accident year on a pure-paid development basis.

CHART 20
RESIDUAL PLOT VERSUS TIME CONVERTED DATA
BODILY INJURY LIABILITY IN PERIOD LOSS PAYMENTS



In Chart 21 the Cape Cod decay factor is at 75% which ties the years together closely. I also ran off one at 25% which allows each accident year to find its own level much better. If you just looked at the accident year scatter plot for that, all these diamonds now came close to the line. But then the standard error of that projection was much larger, because on a projection with a reasonably long tail; you get a point right near the line for recent accident years, but you get more error allowing the recent accident years to find their own level.

CHART 21
RESIDUAL PLOT VERSUS TIME CONVERTED DATA
BODILY INJURY LIABILITY IN PERIOD LOSS PAYMENTS



Now we can look at the various models and determine how well the curve fits the actual data. The solid line is the actual data and the dotted line is a fitted curve (Chart 22). This is a reasonably old accident year. You have lots of actual data points.

Chart 23 is the M1 model of a more recent accident year where there are just a few actual data points. Chart 24 is an incurred data model so the curve starts to drop much sooner because of the shorter tail on the incurred data. The fit on the incurred models was substantially better. Chart 25 shows one more incurred model. In this case it's always dropping, but there is a good fit.

In addition to plotting how the paid-loss is projected in the operational time model, I can also show you a plot of the actual average sizes which are modeled within the regression model. Chart 26 is an example of that. I used that same Hoerl curve, the curve with both the power term and the exponential term. The curve went up and down and turned back up again which worked very nicely on this data, but it suggests some caution on extrapolating with the curve. Because the curve has that extra shape parameter, you do have to be careful about what's happening in the tail and look at some plots. I guess I'm wary of extrapolation as a general rule, but in this particular case I think we received generally very nice results on the average claim sizes. Of course at the tail of the data is an average claim size that you're observing on very few claim counts, sometimes just one or two or three claim counts in a cell. So that's where you're going to see more and more erratic actuals as you move to the right, and that's why we have the model. Chart 27 shows average claim size for a more recent accident year.

Using that same model, we can also plot the actual and projected payments (rather than average claim sizes) as shown in Chart 28. What I was trying to emphasize with these plots is that these fitted data will not always look so smooth because we've actually fitted an average claim size so smoothly. We multiply those by actual claims closed in-period. Sometimes the fitted data will look a little erratic, usually in a shape that matches the actual data. Chart 29 is another example of that.

Chart 30 shows one of the distinctions between my model and Tom's. He uses a variance structure that comes directly out of the theory he built, but my approach is much more empirical. I use curves that look fine, and then I check to see if the variant structure is working or if there's any residual heteroscedasticity. If I see it, I try to model it and correct for it. I check for weighted square errors against delay. I only model it in the delay direction which is the same direction that Tom's model picks up. You can see that I have fit a curve with some significance to the fit on the way the squared error relates to the delay.

Chart 31 is the operational time model. It's a little harder to see the curve on it. On these models the reading of the (heteroscedasticity) goes the other way; it's heavy in the first columns. It's hard to see but the curve is trending down. We corrected for it and it works much better. Again, there was a lot of (heteroscedasticity) before we applied the model, but after correcting for it, the heteroscedasticity came down very nicely in the heart of the distribution. We're less than the 50% level, so we're doing well on that.

CHART 22
PLOT OF FITTED VS. ACTUAL DATA POINTS
BODILY INJURY LIABILITY IN-PERIOD LOSS PAYMENTS
ACCIDENT YEAR 80

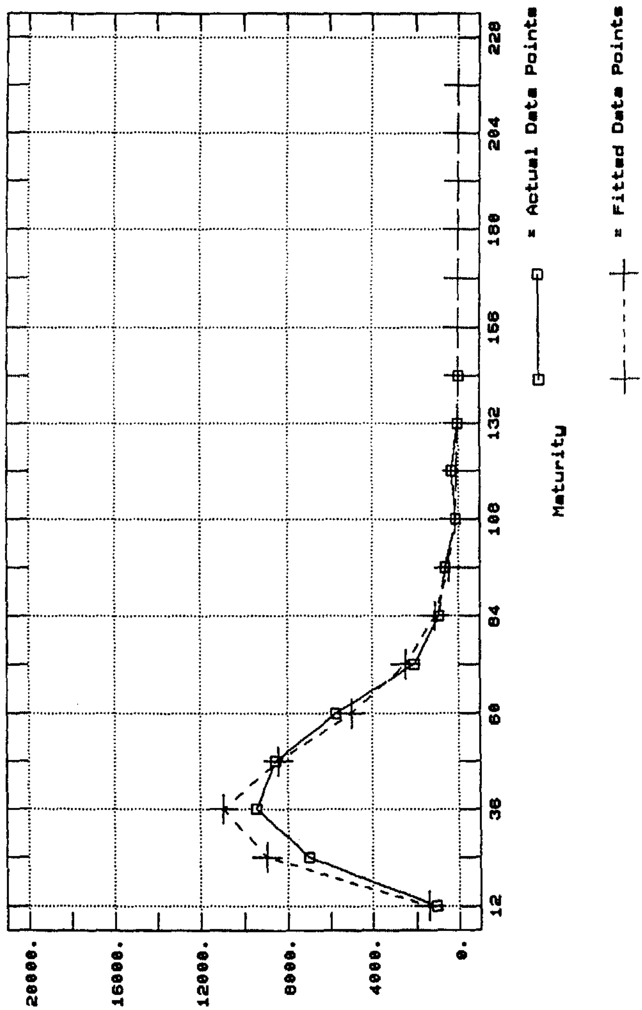


CHART 23
(M1)
PLOT OF FITTED VERSUS ACTUAL DATA POINTS
BODILY INJURY LIABILITY IN PERIOD LOSS PAYMENTS
ACCIDENT YEAR 89

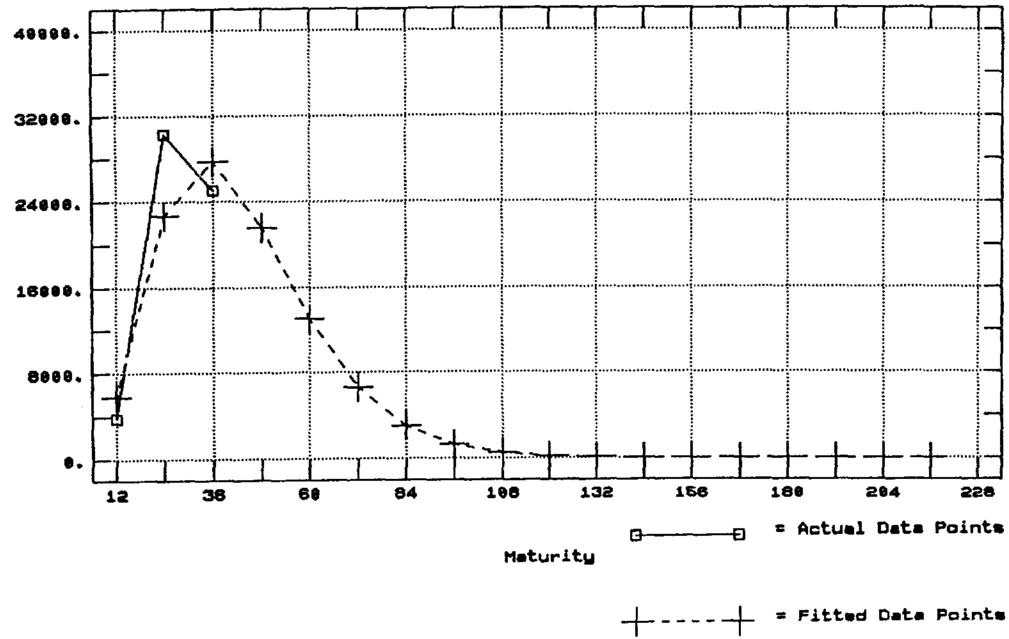


CHART 24
(M4)
PLOT OF FITTED VERSUS ACTUAL DATA POINTS
BODILY INJURY LIABILITY IN PERIOD INCURRED LOSSES
ACCIDENT YEAR 81

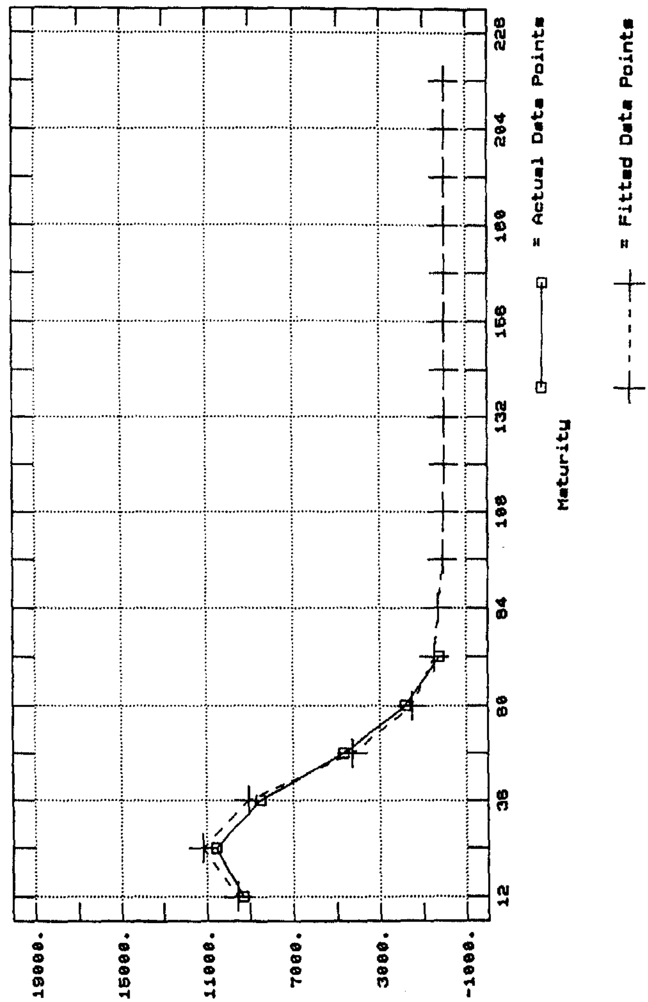


CHART 25
(M4)
PLOT OF FITTED VERSUS ACTUAL DATA POINTS
BODILY INJURY LIABILITY IN PERIOD INCURRED LOSSES
ACCIDENT YEAR 84

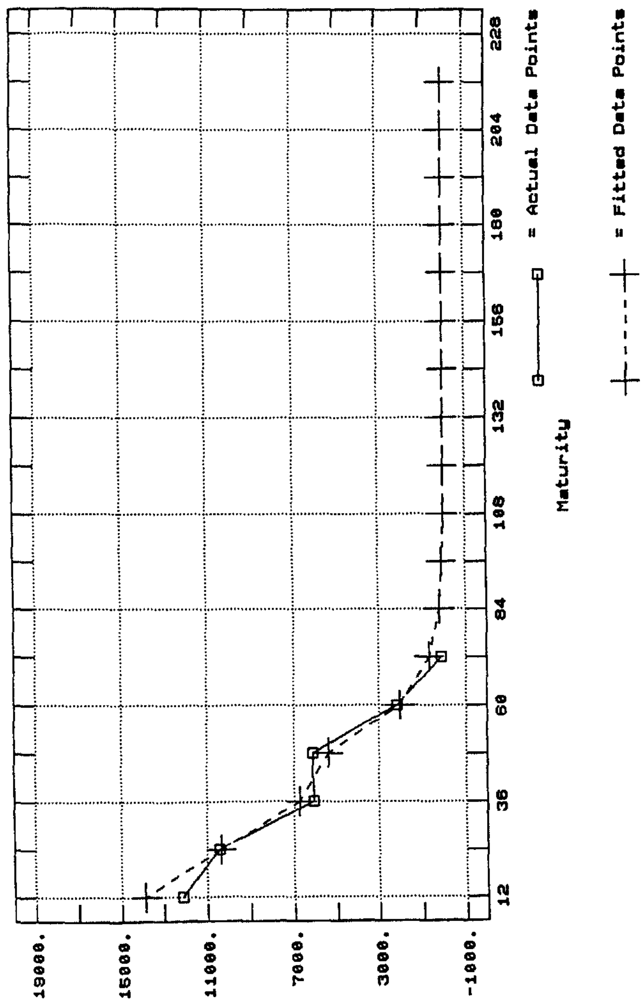


CHART 26
(M5)
PLOT OF FITTED VERSUS ACTUAL DATA POINTS
BODILY INJURY LIABILITY IN PERIOD AVERAGE LOSS PAYMENTS
ACCIDENT YEAR 78

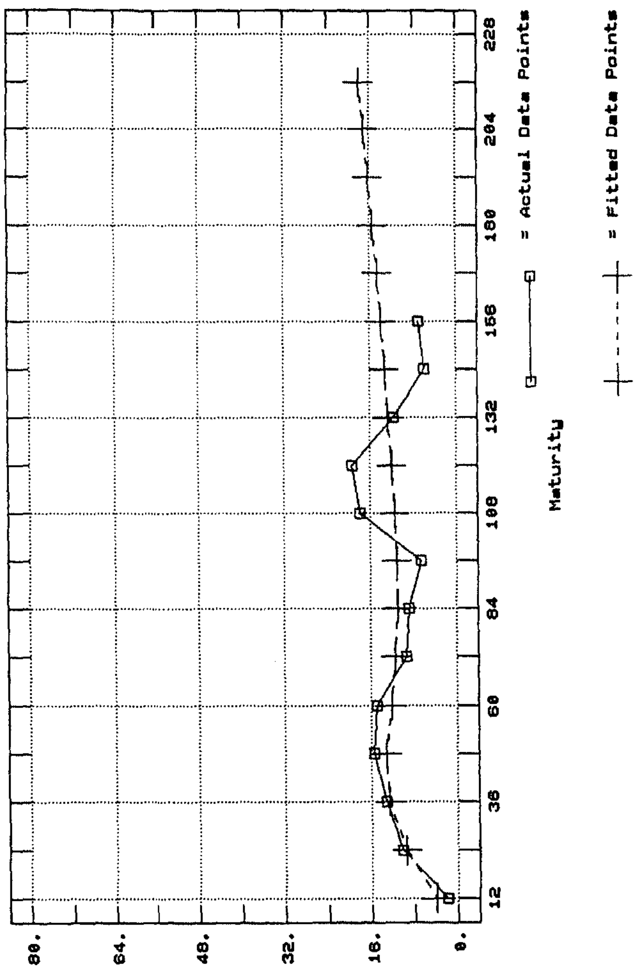


CHART 27
(M5)
PLOT OF FITTED VERSUS ACTUAL DATA POINTS
BODILY INJURY LIABILITY IN PERIOD AVERAGE LOSS PAYMENTS
ACCIDENT YEAR 85

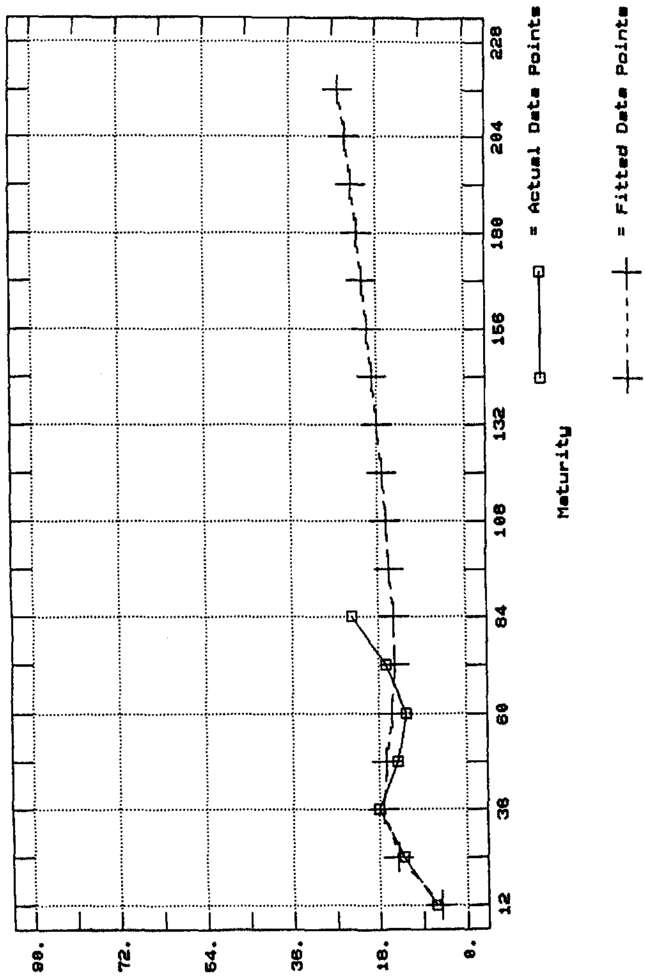


CHART 28
PLOT OF FITTED VERSUS ACTUAL DATA POINTS
BODILY INJURY LIABILITY IN PERIOD LOSS PAYMENTS
ACCIDENT YEAR 78

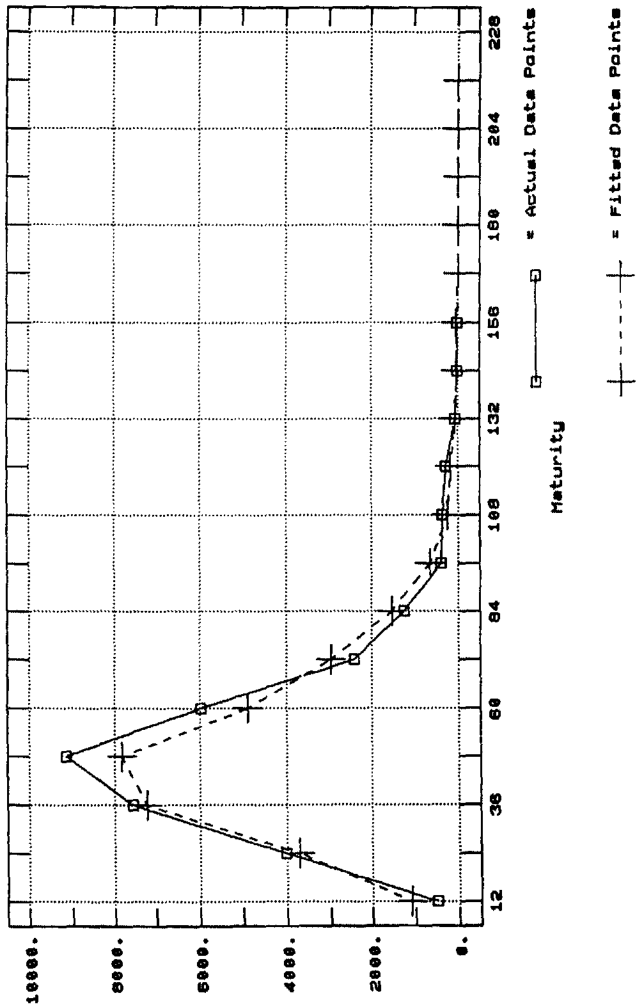


CHART 29
PLOT OF FITTED VERSUS ACTUAL DATA POINTS
BODILY INJURY LIABILITY IN PERIOD LOSS PAYMENTS
ACCIDENT YEAR 84

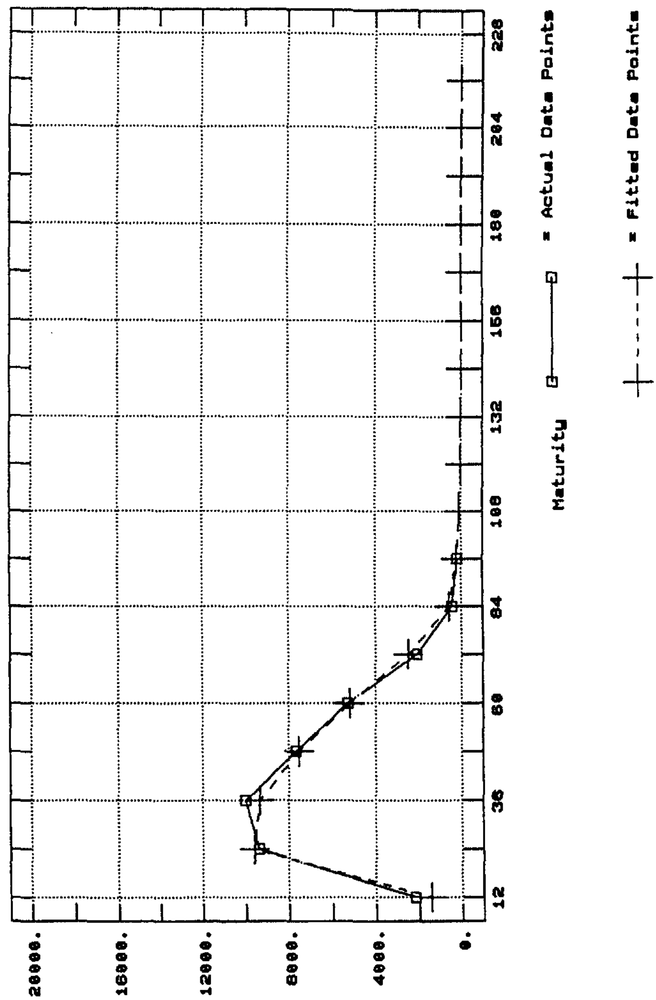
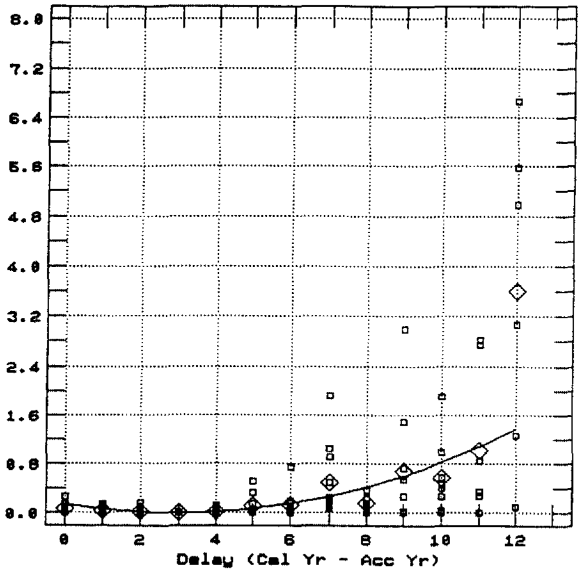


CHART 30
HETEROSCEDASTICITY
WEIGHTED SQUARED ERRORS VERSUS DELAY
BODILY INJURY LIABILITY—IN PERIOD LOSS PAYMENTS
BEFORE APPLICATION OF MODEL



AFTER APPLICATION OF MODEL

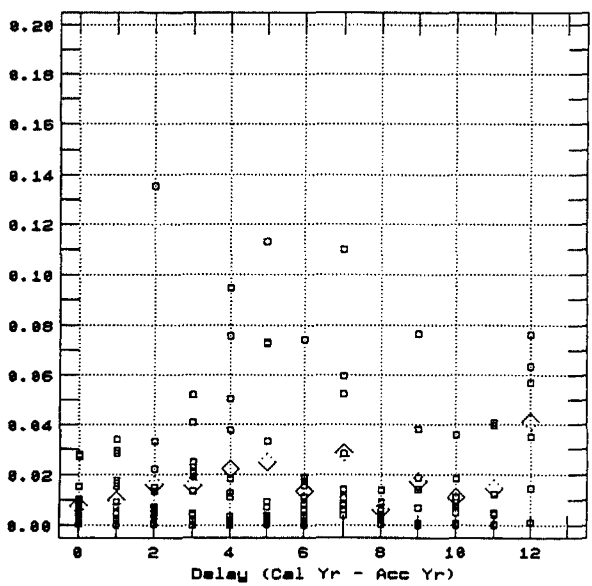
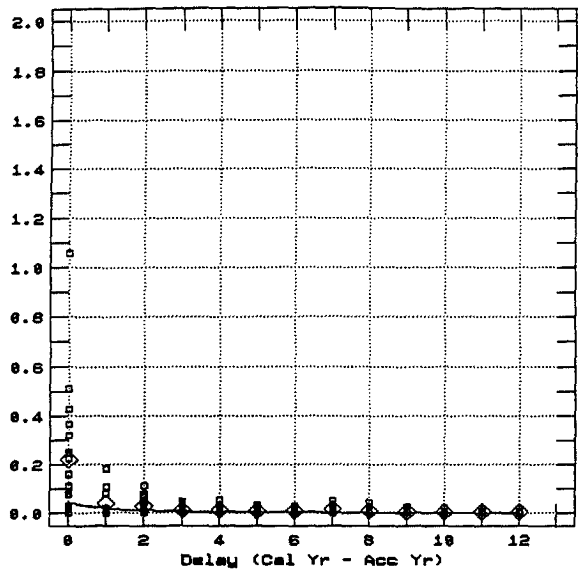
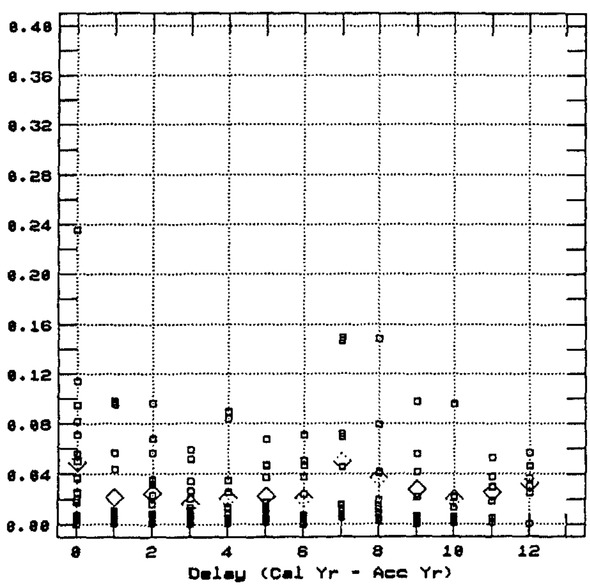


CHART 31
HETEROSCEDASTICITY
WEIGHTED SQUARED ERRORS VERSUS DELAY
BODILY INJURY LIABILITY—IN PERIOD LOSS PAYMENTS
BEFORE APPLICATION OF MODEL



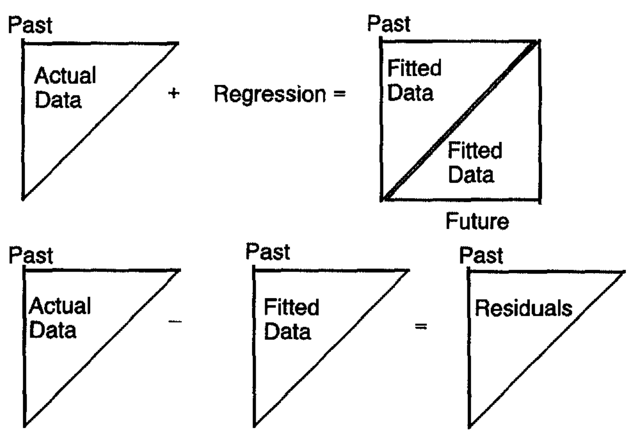
AFTER APPLICATION OF MODEL



So those are my diagnostics. Now let's discuss the standard errors. I use a bootstrapping approach rather than just the algebraic calculation. The math is less complicated, but the computer power is more complicated. The programming is also more complicated. One of the advantages of it is that when I'm using my hybrid models, which are not a pure regression model, I can still run all of that through the bootstrap.

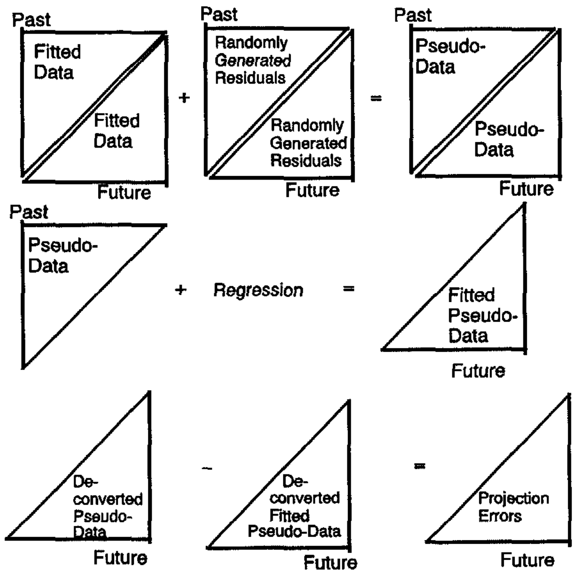
Let me give some basic concepts of bootstrapping. Chart 32 is not really an equation but shows how you apply bootstrapping. When you've got actual data, you apply regression and that gives you fitted data in the past and in the future which I've represented with triangles. Subtract the actual data minus the fitted data to get your residuals. In regression, we require that the residuals be random; that's what we look for in all the scatter plots; they must also be independent, identically distributed, and normally distributed. The normal distribution assumption is what makes the regression estimate optimal. In bootstrapping we drop normally distributed as a property. We don't presume how the errors are distributed, but it is still important that they are random, independent, and identically distributed; by random I mean they are not systematically related to any of the independent variables.

CHART 32
HOW TO APPLY BOOTSTRAPPING



Now we have this triangle of residuals, and rather than presuming any particular form for the error distribution, we assume that, if I have 120 residuals, I will use the actual residuals as the error distribution (Chart 33). So now the error distribution is a discrete distribution with 120 likely and equal results. I randomly generate residuals by selecting from that distribution with replacement. So now I take my fitted data in the past and future, and I randomly generate a whole square or a whole rectangle of residuals. I add that to the fitted data, and that gives me what I call past and future pseudodata. So for this pseudodata I know the future. I take the past pseudodata, I run it right through the whole regression model, and that gives me a future triangle of fitted pseudodata. So now I have a triangle of future pseudodata and a triangle of fitted future pseudo data. I run them through the whole deconversion process. I subtract them, and I get a triangle of projection errors on the pseudodata. I then run that repeatedly 500 or 1,000 times. It uses a lot of computer time, but the computers keep getting faster and faster, so it's not a problem. It takes a lot of space to store all those results.

CHART 33



Now I have a projection of the whole distribution at every point. I have a great deal of power to look at the data subcombined in any way which is important to the extent that the projection errors are not independent of each other. Don't forget that the process risk will be independent at the different points, but the parameter risk will be highly dependent. If I want to look at subcombinations, (e.g., the total reserve or any combination of three accident years) I can combine it any way I want. In Table 11, I have summarized it in total by individual accident year and by individual calendar year of payment. We are most interested in the total line in this case.

Notice that there's a bias correction and that's because we've taken a log transformation of the data, so that means my fitted model is not the mean. If observations arise from a log normal distribution, then the fitted model corresponds to the mean of the underlying normal. But if you take the antilog of that value, that does not give you the mean of the log normal. So it's important, when you're using log transformed models, which a lot of people do, to remember that the fitted model is not the mean. You have to correct for bias. Because we've run off this bootstrap, it gives us the correction for bias directly. You can also do it algebraically if you're willing to presume the normal distribution. This is distribution free, but I have compared my results with people who do it algebraically and they usually come out about the same anyway. So I'm not sure I've added anything.

TABLE 11
BODILY INJURY LIABILITY IN PERIOD LOSS PAYMENTS
PROJECTED RESERVES AND ANALYSIS OF ERRORS
Inflation (+) or Discount (-) Rate: 0

				Total Projection Error			Sources of Variance	
Year	Original Fit	Bias	Corrected Fit	Std. Dev.	Variance	Skew	Statistical (Process) Error	Parameter Estimation Error
Accident Year								
74	0	0	0	0	0	1.000	0	0
75	0	0	0	0	0	1.000	0	0
76	32	-23	55	52	3	-3.678	3	0
77	0	0	0	0	0	1.000	0	0
78	0	0	0	0	0	1.000	0	0
79	0	0	0	0	0	1.000	0	0
80	15	-14	29	40	2	-2.897	2	0
81	49	-62	111	136	18	-3.989	18	0
82	64	-68	132	146	21	-3.909	21	0
83	145	-93	238	154	24	-2.564	24	0
84	127	-83	210	155	24	-2.445	24	0
85	455	-120	575	209	44	-1.452	43	1
86	975	-158	1,133	306	94	-0.925	89	4
87	3,079	-203	3,282	533	284	-0.503	249	34
88	11,128	-354	11,482	1,225	1,502	-0.172	1,216	354
89	28,051	-538	28,589	2,282	5,207	0.006	3,458	2,044
90	59,343	-752	60,095	4,361	19,015	0.246	10,837	8,377
91	83,700	-763	84,463	5,939	35,271	0.006	19,876	17,171
Calendar Year								
92	78,165	-778	78,943	5,183	26,862	0.068	17,970	11,208
93	56,328	-503	56,831	4,188	17,537	-0.043	10,272	7,464
94	30,143	-532	30,675	2,681	7,188	0.044	4,253	2,840
95	14,136	-402	14,538	1,534	2,354	-0.080	1,548	820
96	5,195	-195	5,390	802	643	-0.393	537	139
97	1,813	-227	2,040	431	186	-0.430	170	21
98	721	-166	887	295	87	-1.161	84	4
99	356	-141	497	259	67	-2.029	67	1
100	174	-120	294	230	53	-2.853	53	0
101	79	-83	162	216	47	-4.335	46	0
102	55	-83	138	192	37	-4.794	37	0
103	0	0	0	0	0	1.000	0	0
104	0	0	0	0	0	1.000	0	0
105	0	0	0	0	0	1.000	0	0
106	0	0	0	0	0	1.000	0	0
107	0	0	0	0	0	1.000	0	0
108	0	0	0	0	0	1.000	0	0
109	0	0	0	0	0	1.000	0	0
Total	187,165	-3,230	190,395	10,462	109,449	0.176	34,227	82,124

Note: Projection Bias: E(Projected Reserves) – E(Actual Reserves)

This is the projection for what I thought was my best model, the operational time model. I came out with a mean of 190 which I believe is about 15 lower than the mean that Tom got and a standard error of about 10.5. My standard error doesn't account enough for the potential claim count projection error, so I think my standard errors are too low. I know how to fix that within my bootstrap. I just haven't gotten around to programming it yet.

As an example, I have the same model with the actual distribution in Table 12.

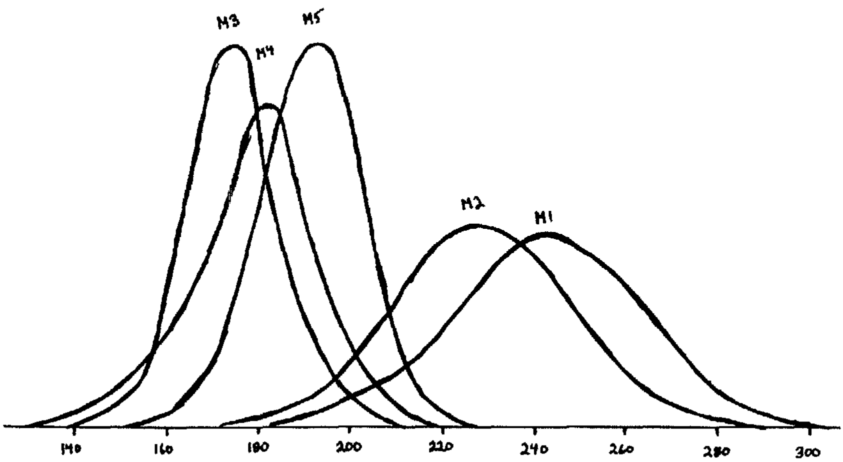
TABLE 12
BODILY INJURY LIABILITY IN PERIOD LOSS PAYMENTS
CONFIDENCE INTERVAL ANALYSIS
Probability Distribution of Reserves

	Percentiles						
Year	**5**	**10**	**25**	**50**	**75**	**90**	**95**
Accident Year							
74	0	0	0	0	0	0	0
75	0	0	0	0	0	0	0
76	10	15	23	42	72	99	131
77	0	0	0	0	0	0	0
78	0	0	0	0	0	0	0
79	0	0	0	0	0	0	0
80	1	2	5	13	33	82	121
81	18	25	43	76	130	193	267
82	24	33	59	92	156	241	354
83	79	96	146	203	297	395	477
84	59	73	112	169	260	355	494
85	305	351	430	541	680	817	906
86	715	783	922	1,095	1,301	1,529	1,698
87	2,479	2,630	2,911	3,247	3,626	3,937	4,150
88	9,468	9,948	10,625	11,467	12,238	12,990	13,580
89	24,831	25,628	27,111	28,665	30,064	31,461	32,406
90	52,415	54,138	57,054	60,287	63,351	65,481	66,788
91	74,759	76,954	80,255	84,481	88,823	91,846	93,352
Calendar Year							
92	70,679	72,235	75,500	79,031	82,462	85,452	87,684
93	50,271	51,546	54,074	56,559	59,770	62,275	63,930
94	26,310	27,409	28,819	30,720	32,414	34,007	34,908
95	11,971	12,564	13,397	14,608	15,658	16,494	16,975
96	4,149	4,438	4,837	5,363	5,913	6,420	6,786
97	1,392	1,533	1,746	2,012	2,312	2,618	2,760
98	493	561	679	849	1,040	1,257	1,380
99	216	252	324	444	588	781	942
100	83	103	154	238	343	511	696
101	23	31	53	102	198	309	444
102	11	20	36	79	168	308	389
103	0	0	0	0	0	0	0
104	0	0	0	0	0	0	0
105	0	0	0	0	0	0	0
106	0	0	0	0	0	0	0
107	0	0	0	0	0	0	0
108	0	0	0	0	0	0	0
109	0	0	0	0	0	0	0
Total	173,281	176,806	183,486	190,856	197,647	203,681	207,157

I ran 500 iterations of the bootstrap through each model. I ran 500 iterations more than once on a couple of the models and they came out about the same.

I plotted the results on Chart 34. They show how the various models came out. This will give you an idea of the whole concept of specification error. You can have reasonably well-specified looking models, but they don't project the same result. If you have a few competing models that you think, based on your diagnostics, look reasonably specified, you can actually put those distributions together and get a larger error measure. The two poor performing models were the paid models (M1 and M2). I would reject those because the diagnostics were not that good. You can see that they had standard errors at least twice that of the other models. I felt the validity of these models was questionable. I thought the best performing model overall is M5. It was the operational time model, but the standard errors were similar on the M3 and M4 models and projections were somewhat lower.

CHART 34
TOTAL ALL YEARS



In the projection intervals in Table 13, you see M3 and M5 were well-performing models. They're overlapping at the 90% projection interval. When I say 90% projection, I do not mean 5-95; I mean 10-90, so it's 80% wide. They are overlapping so I guess it's reasonably plausible.

TABLE 13

Model	Mean	90% Projection Interval
M1	233	206–259
M2	228	206–250
M3	174	161–186
M4	178	160–195
M5	190	177–204

Roger, who knows the real data behind this data set, has told me that he has accumulated two years more data since he did this and things keep going down. So I like the fact that I have the lowest answer.

MR. HAYNE: The micro approach that I applied was based on a collective-risk model. I looked at the collective risk model in terms of reserves for individual accident years. I looked at the number of outstanding claims, at the distribution for the outstanding claims, and I tried to build in some parameter uncertainty. Basically the collective-risk model simply says you take many samples out of distributions and add them up to get your aggregate losses. You do this several times. A paper by Glen Meyers and Phil Heckman published in 1985 or 1987 described an algorithm for calculating the aggregate distributions. They've also built in certain methods that incorporate parameter uncertainty. You can actually use those methods to calculate the expected value for the total and the variance for the total in terms of the parameter uncertainty and the underlying distributions. It turns out that the parameter uncertainty is handled by nonzero parameters B and C. In this case, if B and C are not zero, the variance in the mean will not go to zero as the number of claims increases.

My approach starts off with standard actuarial-type methods. The examples that I'm bringing up are written up in a paper that I've submitted in response to a call for a prize paper from the Risk Theory Committee of the Casualty Actuarial Society (CAS). That paper, along with quite a few other papers that were submitted for that competition, will appear in an upcoming issue of the *CAS Forum*. My first approach is to take the more standard actuarial methods and come up with the estimate of the expected reserves or the expected outstanding claims. My best estimate is about $203 million. This compares with Spencer's $190 million and Tom's $210 million. So we're close.

I attempted to reflect parameter uncertainty by the variation in the projection methods. In my study, there really is not a lot of process uncertainty. But when you build in some reflection of the parameter uncertainty, diffusion becomes much greater. In my case, as I said, I had a mean of $203 million. I have a standard deviation of about 13 million, which comes in very close to Tom's results which are about $210 million for the mean and a standard error of 13. Spencer's mean is $190 million and standard error is about 10.

What's encouraging is that we have three methods that are coming at the same problem from three different angles and the answers are at least close. There is uncertainty in the loss reserves, but I think we're in agreement that the standard deviation of 10-13 is not bad. I think reserves in the neighborhood of $200 million aren't bad either. I wish we could have all come in and given an exact reserve answer and standard deviation answer, but the real world isn't that way. As Spencer pointed out, the real data continues to behave very pathologically. We don't know why, but the payments have just dried up almost in an unbelievable fashion, and the estimates continue to drop. So Spencer's closer to where I would be if given a little bit more information.

FROM THE FLOOR: If you were the appointed actuary for a casualty company, what would you have felt comfortable with as an estimated mean?

MR. HAYNE: I don't know. I'd have to evaluate an individual carrier. From my point of view, I'd have to know an individual company's underlying data, or know what's going on in the company, before I could feel comfortable signing at a certain level.

MR. GLUCK: Let's say that the best you can do, whether you're using typical methods or regression methods, is conclude that there's a large error, or that you have a badly behaved data set or an uncertain situation. As the appointed actuary I'm not going to provide an answer, just another question. Where does that bring you? If my standard error turns out to be 25% of the mean, does that mean that I'm happy with reserves 25% below my best estimate? They're within one standard error after all. I think that your reaction could very well go the other way. If there's a large amount of uncertainty, I don't think that that should be used personally as a justification for going well below your mean estimate. If you are unsure about your mean estimate, wouldn't you be better off being conservative than using that uncertainty as an excuse to sign off on a result that's too low because you can't be sure that that low result is wrong? That's my personal reaction to a large standard error. It doesn't necessarily give you a lower reserve that you can sign off on as an appointed actuary. That's my opinion anyway.

MR. WRIGHT: I have nothing to add. I'm not an actuary, and these are difficult questions I prefer to leave to you actuaries who are paid to answer them.