# Valuation Actuary Symposium*

Boston, MA
September 20–21, 2004

## Session 39TS
## Approaches to Determining Unpaid Claim Liabilities: Old and New

**Instructors:** DOUG FEARRINGTON
ROBERT LYNCH

*Summary: This session discusses the different techniques for determining health unpaid claim liabilities such as the traditional loss development method, casualty reserving techniques and new approaches that have been discussed in recent actuarial literature. Case studies explore and illustrate the issues and the pros and cons associated with each technique.*

**MR. ROBERT LYNCH:** I'm with Blue Cross/Blue Shield of Michigan, and my co-presenter is Doug Fearrington from Anthem Blue Cross/Blue Shield in Richmond, Va. This session title refers to old and new methods, but both Doug and I are mostly going to be focusing on new methods that we've been working on. I will be giving a brief review of old methods, mostly for comparison purposes. Probably the primary difference between Doug's approach and mine is that mine is spreadsheet-oriented. I use Excel a lot, whereas Doug is using high-order programming, especially SAS for his approach, which gives access to a wider variety of statistical tools than in Excel.

Why should we worry about the accuracy of incurred-but-not-reported (IBNR) and incurred-but-not-paid (IBNP) claims? We worry about them because our goal in life as actuaries is to make people happy. There are a fair number of people to make happy. An important one is keeping regulators happy. Regulators like to know that our reserves are adequate to cover our liabilities. That's their big purpose in life, and as long as they know that, they're happy. We also have to keep the nice people at the IRS happy because if they think that we're inflating reserves to hide profits and avoid taxes, they get unhappy. They want to know we're not hiding profits. We need to keep auditors happy just because they can make our lives miserable if they're unhappy.

One group that is often not thought about by the valuation actuary is the underwriters. If you're pricing large group experience, you have to complete claims to get estimates of incurred claims to project. That's what the premium's based on. Once a premium is delivered, you can't restate the premium because your IBNR estimate was off. You're stuck with it. Anybody who has done pricing in that area knows there's a constant trade-off between trying to get the most recent experience and trying to have something that's credible.

Another group that we want to keep happy is the investors because they don't like surprises, such as large restatements of reserves. Finally, an important person to keep happy is the CFO because if the CFO gets unhappy, we can be out of a job. That's our goal in life.

I'll start out with reviewing the three most common traditional IBNR calculation methods, and there are a lot of variations of these. The completion factor method includes the chain ladder and link ratio methods, which are mathematically equivalent. It's probably the most popular, and I say that from my own bias. It may not be, but I think it is. That is probably because for any given combination of incurred claims and payment run-out, you have a single factor that, once calculated, is easy to apply to any set of incurred and paid claims. Especially for pricing purposes and for the underwriters, this is easy to apply over and over again, so that makes it popular. Back in the days when we didn't have computers handy, once you had this factor, you could keep it for a year or so and didn't have to do a lot more calculations again.

The incurred claims method, including the loss ratio method, goes about by backing into the IBNR by first making an *a priori* estimate of total incurred claims and then subtracting what's already been paid. The difference is the IBNR. Variants such as the loss ratio method usually involve different ways of getting that *a priori* estimate of total incurred claims.

Finally, the Bornhuetter-Ferguson method essentially combines the completion factor and incurred claims methods by multiplying an *a priori* estimate of total incurred claims by the complement of the reciprocal of the completion factor. In English that means that if X is the completion factor, you multiply the incurred claims estimate by 1-1/X. The completion factor and incurred claims methods are mentioned explicitly in the *NAIC Health Reserves Guidance Manual* as the preferred methods for calculation of IBNR.

All three of these methods present problems of varying severity when used. The completion factor method suffers from a high variance and a high error, and that is basically because it depends exclusively on the use of an estimated parameter, which is then multiplied by the incurred and paid claims. This is itself a statistic, and when you multiply two variables together, the nature of statistics gives you a high variance because you multiply the variances together. It's also based on the implicit assumption that there is a high degree of correlation between the claims that have already been paid and claims that will be paid in the future. While in my opinion this assumption is questionable, it doesn't seem to have attracted much attention. Despite the fact that a lot of actuaries, including myself, have recognized that the completion factor method gives poor results, it's probably still the most popular of the three methods mentioned here for calculating reserves.

The incurred claims method suffers from the obvious flaw—that the actuary must somehow come up with an *a priori* estimate of incurred claims. As I said, the variants of this method mostly deal with different ways of trying to estimate the total claims. The two most common approaches appear to be using the completion factor method to complete claims up to some time point prior to the valuation date and then projecting those claims forward or opting out and shuffling off the responsibility for the projection to the pricing actuaries by multiplying the projected loss ratio by the expected premiums and using that as the incurred claims. Because the incurred claims method ignores how many claims have already been paid, it gives the rather disconcerting result that the IBNR estimate is negatively correlated with the claims already paid. This characteristic directly contradicts the basic assumption of the completion factor method, which leaves you wondering how they can coexist in the same universe because they're relying on opposite assumptions.

Finally if the incurred claims method is applied to data with more than a couple of months run-out, it frequently encounters the problem that the claims already paid may be more than the incurred claims estimate for any given month. Since negative IBNR estimates are not allowed, this results in the potential for a biased estimator of overestimating the amount of IBNR.

The Bornhuetter-Ferguson method avoids these last two problems of the incurred claims method. It still requires an *a priori* estimate of incurred claims; however, in my opinion, it gives the best results of these three. Depending on the accuracy of the total incurred claims estimate, it gives accuracy in its results that approaches the simple paid claims method. Unfortunately, the Bornhuetter-Ferguson method is complicated and kind of clumsy to use compared to the previous method, and I think that's restricted its acceptance in the general usage.

If we start over to invent some new method for calculating IBNR, what would the ideal IBNR calculation method look like? Here are some objectives that I put forward as a standard for comparison. The ideal method would be mathematically sound and practical to use. By mathematically sound, I mean that the method would be based on a defined and robust model of how claims develop and are paid, with all material assumptions clearly identified. The method would yield estimators that are unbiased, and the estimator would have a low variance as measured by standard error. By practical, I mean that the method would be relatively simple to use and understand, and it should be spreadsheet–friendly, since most actuaries like myself are acutely addicted to their spreadsheets, whether it's Excel or Lotus.

Let's look at how to go at this. The first step in the calculation of IBNR is usually to organize the incurred and paid claims by calendar month of incurral and payment in the familiar lower triangular claims payment matrix, with the columns representing the calendar year of claims incurral and the rows representing the calendar months of claims payment. In Chart 1, I'm showing 24 months of claims incurral and payment from January 2002 through December 2003.

Next the claims are rearranged into the upper triangular matrix format where the row represents the paid lag month rather than the calendar year of payment, which is defined as the difference between the calendar month of payment and the calendar month of claim incurral (Chart 2). By convention, claims paid in the same month in which they are incurred are assigned zero months of claim run-out. The upper left portion of the upper triangular matrix now represents claims payment data—claims that have been incurred and paid—while the lower right represents future claim payments, and it's empty. The stair-step line between the two parts represents the valuation date, in this case December 31, 2003. The task at hand is to obviously fill in the cells at the lower right half of the matrix.

I started developing what I call the simple paid claims method. The simple paid claims method is based on the assumption that for a given claims incurral month, the amount of per member per month (PMPM) claims paid in a given lag month is related only to the durational lag and is independent of the amount of claims paid in prior months. Based on this approach, the intuitive estimator is simply the average PMPM claims paid in the respective durational lag months and prior incurral months, after being adjusted for trend and seasonality.

If we look at the upper triangular claims payment matrix, the projected estimates for the cells in the lower right are simply the mean of the values of X and the corresponding row or lag months in the upper left paid portion of the matrix. In this exampleChart 3, for paid lag month three you just take the average of these across and fill in the cells with that. It's simple, but you still have to adjust for trend and seasonality. To get the total IBNR, you do the same process for every row and then add all those cells together. That's it. It's simple. That's why I call it the simple method, and since the total IBNR is based on the summation of independently derived cell estimates and variances additive, this approach in theory should develop an estimate with a much smaller variance than the completion factor method.

The next question is whether we can do better. In an article in the January issue of *Contingencies*, I gave a short outline of it—one paragraph and that represents the simple paid claims method. Since then I've worked at improving it. If we want to look at improving it, we go back to the completion factor. The correlation of past and future paid claims is a basic premise of the completion factor method. This represents a conventional wisdom, and since we would like to think that 15,000 actuaries can't be wrong, it should have some predictive value.

How can we bring this assumption back into the simple model? Let's put it into the mix. I'll call this new iteration of model development the multiple linear model. Like the simple paid model, it incorporates the assumption that future paid claims are linearly related to the durational lag time. That will be the starting point, and then we add the assumption that future claim payments are linearly related to prior paid claim amounts by lag period. Since this model requires some sort of parametric estimation, we'll assume that these parameters are determinable from historic claims data; otherwise we'd be stuck. We'd have an idea but wouldn't be able to do anything with it.

The natural development of this model would suggest a multiple linear regression where the PMPM claim amount in a given durational lag period is the dependent variable, and the PMPM paid claim amounts in past periods prior to the valuation date represent the independent variables. If we go back to the upper triangular claims matrix to see what this approach would look like, the estimate of the future paid claims for a given cell is based on the array of claims paid in the equivalent past lag periods and the claims for the given incurral month paid prior to the valuation date.

In this example (Chart 4), I used claims incurred in October 2003 to be paid in January 2004, with a December 31, 2003 valuation date. The estimate will be a function of all previous paid claim amounts for the third lag duration month, plus the previously paid claims for the incurral month multiplied by some array of parameters.

Putting this relationship into a formal multiple linear formula gives the expression as shown where the expected value of X equals that stuff down in the corner. Alpha is the zero intercept parameter primarily determined by the past lag duration three paid claims, which is the row going across, and the respective values of beta are the slope parameters for each of the lag durations, in this case durations zero, one and two, which are primarily determined by their respective arrays of corresponding durational and paid claim amounts. Since regression is a two-step process, the first step is determine the parameters of the regression followed by applying those parameters to the durational paid claim amounts for the incurral month in question. X is a combination of all those parameters there.

This concept of the multiple linear model is not new. It was put forward by M. Weiss in a 1985 paper in the *Journal of Risk and Insurance*, but it has not gained widespread use since it violates one of the objectives that I put forward earlier—it's complex. Each prior lag duration requires a separate parameter to be determined, and each projected lag duration requires a different formula with a different number of parameters. Each cell in the lower right of the claims matrix requires a different formula with a different number of parameters. Some of them may have a lot of parameters, so I would say this approach is spreadsheet-unfriendly. If you're an Excel user, you won't like it. You'll have a hard time making it work at all.

What do we do? We have something that looks good, but it's too complex to use. I like to take the approach that when a problem is too complex to tackle directly, you simplify. You go about simplification of the multiple linear model to something I call the cumulative linear model. Bivariate linear regression is a lot easier than multiple linear regression, especially on a spreadsheet, so we insert a simplifying assumption that will convert the multiple linear model to a bivariate linear model. We introduce a simplifying assumption. The expected future paid claims for each incurral month are independent of when the claims were previously paid for the given incurral month (when they were actually paid), as long as they were paid by the valuation date. It doesn't matter when they were paid, as long as they were in by the valuation date. That's the only thing we look at. We can apply this simplifying assumption in the multiple linear model by replacing paid claims by lag duration as the multiple independent variables with cumulative paid claims as of the valuation date as a single independent variable, so now we've reduced it to a bivariate linear regression model.

As we develop this into a methodology, for purposes of the independent regression variables, we substitute the cumulative paid claims matrix for the paid claims matrix as the source of the independent variables in the multiple linear model, perform bivariate linear regression where the cumulative PMPM paid claims replace the multiple prior paid claim amounts as the single independent variable, and the paid claims matrix is retained as the source of the dependent regression variables.

Chart 5 is an example of the cumulative paid claims matrix to be used for the independent regression variables. It looks almost identical to the paid claims matrix, except that instead of each cell containing the amount of claims paid in a durational lag month, the cells contain the cumulative amount of claims paid through the end of the durational lag month.

I have a little side note concerning the completion factor method because I think this will support what I'm talking about. I also want to come back to it later on a theoretical basis. Chart 6 represents how a completion factor method can be represented on the paid claims matrix. The fundamental assumption of the completion factor method implies that the ratio of the expected claims paid in any given lag month for a given incurral month to the cumulative claims paid through the start of that month is equal to the ratio of the total historic claims paid in the corresponding lag month for all prior incurral months, divided by the total historic cumulative claims paid in all prior incurral months up to the start of the corresponding lag month. If this latter ratio is viewed as a single parameter, the projected estimate of the claims to be paid in a given lag/incurral month is simply the cumulative paid claims multiplied by that parameter, which we show as a beta down there in the corner.

Chart 7 shows the same thing, but I've replaced the paid claims matrix in the upper half of the matrix, above that dashed line, with corresponding entries from the cumulative paid claims matrix. When the completion ratio is applied, instead of sending all the values above the dashed line, you take the cumulative values row directly above the lag month of interest. These items represent the sum of all the cells in or above that from the paid claims matrix. It's a hybrid matrix. It's showing one thing in the bottom half and a different thing in the top half, but it's useful for illustration. I'm going to be using this hybrid matrix representation of the combined paid claims and cumulative paid claims for the next couple of examples to help illustrate what's going on.

I'll continue on with the development of the regressed paid method. We're operating under the assumption that the expected paid claims in a given lag duration associated with a given incurral month are linearly related to the cumulative paid claims as of the valuation date for the given incurral month. Using the hybrid matrix for illustration purposes and focusing on claims to be paid in lag month three, the claims incurred in October 2003 and expected to be paid in January 2004 will be based on the cumulative total claims paid after two months of run-out. This amount is basically going to be our alpha parameter times beta times that number right there.

The February 2004 paid claims incurred in November 2003 will be based on cumulative total after one month of run-out. That's the last one before the valuation date. That will be used for that estimate. March 2004 paid claims incurred in December 2003 will be based on the amount of claims incurred and paid in December with no run-out. This estimate will be based on that item as the independent variable.

Let's look at the entire process for a single cell. For a valuation date of December 31, 2003, what's the process for determining the amount of claims incurred in November 2003, for which there were two months of known paid claims and expected to be paid in February 2004, which is lag month three? The independent variable data to be used in the final regression calculation is the total claims paid through the valuation date. The first step is to calculate the regression intercept, which I call alpha and slope—beta parameters. The dependent variables for the regression are all lag month three paid claims by incurral month as of the valuation date. Those are the dependent variables for regression. These include incurral months through September 2003. Since the data variable for the regression is the cumulative claims paid through the first lag month, the independent variables for the regression are the cumulative paid claims through the first lag month for the incurral months corresponding to the dependent variables, that is through September 2003.
The standard linear regression is then performed on these two sets of variables to generate the intercepted slope regression parameters. In Excel this is easily accomplished using the LINEST function. The regression parameters are then applied to the cumulative paid claims amount to generate the estimate of claims to be paid in February for claims incurred in November, so that E of X is now alpha plus beta times the sum of Y1 right there, and that's the estimator.

Let's take a look at the whole process from the top as it would look on a spreadsheet application. If this seems repetitive, I am repeating because one of the rules of teaching is to repeat. Say it three times and people will remember it, so please bear with me. The starting point is the paid claims matrix in the upper triangular format. This is our raw data converted to the PMPM amounts. The cumulative paid claims matrix is derived directly from the paid claims. For lag month zero, the paid claims and the cumulative paid claims are necessarily the same amount because there's only one month there. For lag month other than zero, the cell entries are the sum of the paid claim amounts in or above the corresponding cells in the paid claims matrix.

We now performed the linear regression using the paid claims matrix for the dependent variables and the cumulative paid claims matrix for the independent variables. The regression parameters are put into two matrices of their own: one for the intercept parameters, the alphas, and one for the slope parameters, the betas. If we're looking specifically at the projected paid claims of February 2004, for claims incurred in November 2003, the paid claim values to be used as the dependent variables are simply those in the same lag row as the cell we are trying to depend on (Chart 8).

The values to be used as the independent variables of regression are not quite as obvious. Since for this particular cell the parameters will be applied to the cumulative paid claims at the end of the first lag month, it is necessary for the first lag month values in cumulative beta matrix to be used.

You can visualize the correct array of cells to use by taking the last value of the incurred month columns of interest and following that across the lag month row to the cells that match up with the appropriate array of cells in the paid claims matrix. We're looking at this month here incurred in November. We come down, and for the last cell we hit, we just use the figures from that row, and they have to match up with the dependent claims. Again, as I said, the LINEST function in Excel or its equivalent function in Lotus can be used to easily calculate the correct values for the intercept and slope parameters to put into the arrays over here.

The second step of the regression is to apply the intercept and slope parameters to the values in the cumulative paid claims matrix to calculate the correct estimate of future paid claims liabilities for each cell in the lower right of the paid claims matrix. The total IBNR liability amount is then calculated by multiplying the PMPM values for the lower right by the exposures to get a total dollar amount and then adding them all together. Do the same thing for all the cells in the lower right, multiply them by their exposures for each month and add them all together. There's your total IBNR.

I'll recap just in case you missed the first two times around. Using the hybrid matrix illustration, the regression parameters are calculated using the paid claim amounts as the dependent variables and the cumulative paid claim totals as the independent variables. These parameters are applied to the cumulative paid claim amounts by incurral month to determine the estimates of future PMPM paid claims by incurral month and lag duration. After multiplying the PMPM estimates of the paid claim liability by the exposures for each incurral month, they are all summed together.

That's it for the hands-on, how-to part of my presentation, but before presenting some actual results from the calculations, I want to take a moment to show how this regressed paid method ties together the simple paid method and the completion factor method. The basic model for the regressed paid method is the linear relation by any given incurral month of the expected paid claims with the cumulative paid claims as of the valuation date. Since the regressed paid method was developed from the simple paid method by adding on the assumption of some sort of linear relationship of the expected paid claims for the cumulative paid claims, it should come as no surprise that the simple paid method represents a special case of the regressed method, whereby the slope parameter is forced to zero, leaving the future paid claims estimate as determined by a single parameter, namely the average of the past PMPM paid claim amounts for the given lag month. There's the simple method.

You'll recall from my side note on the completion factor method that that method could be represented as a single parameter model also. Specifically, the cumulative paid claims as of the valuation date are multiplied by the ratio of the average paid claims for the lag month divided by the average cumulative paid claims through the end of the prior lag month. As it turns out, this is exactly the result that the intercept parameter in the regressed paid method is forced to zero, thus the completion factor method is in reality also a special case of the regressed paid method. We have all three of these methods tied together with the regressed paid method as the general model, and the simple paid and completion factor methods representing two special cases of that model.

It's easy enough for me to stand here and spin off what seems like a nice method to significantly improve how you calculate your reserves, but it doesn't mean much if it doesn't work. I've applied the various methods to three different sets of real claims data to see how and why they stack up against each other in practice. The three data sets represent claims from a traditional open-panel fee-for-service health plan, a closed-panel HMO managed care organization (MCO) and a mixed-panel plan from a PPO/POS plan. The data represent 36 months of claims from which 28 monthly IBNR values were calculated. The data were essentially complete with 15 months' run-out on the most recent month. Each sample represents approximately 200,000 lives, although the final results have been normalized to represent 100,000 members for comparison purposes. Before performing the calculations, the data were leveled for trend and calendar seasonality effects, so that the result would not be skewed by those factors. If you're wondering what I used for trend, since I had complete data, I had the real trend, so it was fairly easy to get rid of it. We don't usually encounter that in the real world.

The data were also transformed to preserve confidentiality since they were not mine, but with the claims payment patterns left intact. No time waiting was applied. You can use time waiting in these methods if you want to make them even better. First off, the 28 actual monthly IBNR amounts for each data set were calculated directly from the data, and then I ran the data through spreadsheet algorithms to calculate monthly IBNR estimates by the completion factor method, two variants of the incurred claims method, the simple paid claims method and the regressed paid claims method. The incurred claims method involved completing claims using the completion factor method up to either three months or six months before the valuation date and then projecting the average incurred claims forward to the valuation date. I refer to these two estimates here as the three-month incurred claims and the six-month incurred claims methods, respectively.

I present the results in two different formats. One of them is scattergrams as shown here (Chart 9), how the individual monthly data points plot against actual IBNR on the horizontal axis. This axis represents the actual IBNP. The vertical axis represents the estimates on month-by-month points. The diamond points represent estimates calculated using the completion factor method, and the circles are IBNR estimates calculated using the three-month incurred claims method. This first data set is that representing an open-panel fee-for-service type of health plan. The relative standard deviation of the actual IBNR values is 7.9 percent of the average total IBNR amount. As you can see, there is a lot of scatter around the perfect fit line for both of these methods. It's this diagonal line. If the estimate equaled the actual, the dot would fall right on that line. The standard error for the completion factor method is 26.4 percent of the average actual IBNR, which is pretty bad in my opinion. The incurred claims estimate does a little better with the standard error of 15.2 percent of average actual IBNR, which is still nearly twice the standard deviation of the actual IBNR.

Chart 10 shows the results of the IBNR estimates using the simple and regressed paid claims methods respectively, and the picture's a lot different. The IBNR estimates are all within a fairly tight range, and the standard error for the simple method is 8.1 percent, which is close to the actual standard deviation. The regressed method is the only method that delivers a standard error smaller than the standard deviation at 7.5 percent.

Chart 11 illustrates the second format for my results, with the standard errors for the different methods plotted against each other and the standard deviation of the actual IBNR values. This is a summary of the previous two slides, but it highlights the differences in accuracy of these methods. This is the actual relative standard deviation and then completion factor method, three-month and six-month incurred methods, simple method and regressed method.

I also looked at IBNR estimates with some claims run–out, since there are situations, such as pricing experience-rated groups, when such estimates are demanded. Not surprisingly, the standard error of IBNR estimates decreases from those with no claim payment run-out. The completion factor estimates improved the most, but they also had by far the longest way to go. There are two items of note here, however. First, the relative standard errors for the IBNR estimates from both the completion factor and incurred claim methods with one month of run-out are still worse than those for the IBNR estimates from either the simple or regressed paid claims methods with zero run-out. The relative standard error is 10.1 percent for the completion factor method, whereas with no run-out, the simple and regressed had 8.1 percent and 7.5 percent. There is a similar story with the incurred claim methods. The second item is the regressed paid claims method again yields a standard error less than the standard deviation of the actual IBNR values, which was 5.9 percent versus 5.7 percent for standard error.

If we look at the results for IBNR amounts, estimated with two months of claims run-out, the completion factor method is finally getting close to the accuracy of the simple and regressed paid claims methods (if you count a standard error half again as large as close), and the incurred claims method estimates are pretty dismal. Part of their error is due to the increasing tendency of the method to produce underestimation bias with the increasing length of claim run-out time. With three months of claim run-out, the advantage of the simple and regressed paid claims methods over the completion factor method is down to less than 1 percent of the standard error. But the six-month incurred claims method gives results that are inaccurate and biased. With three months' run-out, the incurred method is not good.

The next example is the same analysis applied to data from a closed-panel MCO health plan. It should be noted that since claims come in faster in a closed-panel system, the total IBNR tends to be smaller. The average IBNR for this sample data is about half the total IBNR for the fee-for-service plan we just looked at. The relative standard deviation of the actual monthly IBNR values for this data set was 11.9 percent of the average total IBNR. The standard error for the incurred claims method was pretty close to that at 12.5 percent, but the completion factor method is again performing poorly with a relative standard error of 22.6 percent.

If you look closely at this data set, you can see that the IBNR estimates from the completion factor method are negatively correlated with the actual IBNR values, suggesting that there's a problem with the claims processing. Rather than using this opportunity to let fly with some comment about claims departments and the woes that they cause actuaries, I'll just put forward that this observation with this data set is probably why the incurred claims method performs reasonably well on this data set compared to the other methods. It's less sensitive to problems in the claims processing department.

Chart 12 represents the IBNR estimates from simple and regressed paid claims methods respectively for the same data set. The estimates are not as tightly bunched around to mean as they were for the fee-for-service data, probably because of the apparent problems in the data themselves. However, the regressed paid claims method yields easily the lowest standard error coming in at 10.7

percent compared to an actual IBNR variance of 11.9 percent of average IBNR amount.

Summarizing the results of the previous two scattergrams, the completion factor method is significantly worse than the others, while the incurred claims method approaches those of the simple and regressed paid claims methods. After one month of claims run-out, the IBNR estimates under the various methods get much closer together, although the paid claims method still outperformed the completion factor in incurred claim methods, and the regressed paid method is still the only method to have a standard error smaller than the variance of the actual IBNR values. After two months of run-out, the effects of the underestimation biased with the incurred claims method is giving it a larger relative error than the completion factor method. After three months of run-out, all the methods except the incurred claims method come out pretty close to the actual IBNR estimates.

The next data set is data from a PPO/POS type of plan with a mixed panel. Again the estimates derived using the completion factor incurred claim methods present a scattered pattern with a standard error for the completion factor method of 18.7 percent, being about three times as large as the variance of the actual monthly IBNR amounts. Again the simple and regressed paid claims method gives a much tighter pattern of results. The standard error from the regressed method is 4.9 percent, significantly smaller than the variance of the actual IBNR.

Here is the summary of the results for the zero run-out IBNR for the mixed-panel claims data. The pattern from the fee-for-service data is pretty much repeated, and once again the regressed paid claims data is the only method that gives a standard error better than the actual variance. With one month of run-out, the completion factor method again loses a large portion of its error but still performs significantly worse than the three-month incurred claims or either of the paid claims methods. With two months run-out, the pattern looks similar to the other two sets with the underestimation bias of the incurred claims method giving a significant impact on those results. With three months run-out, and claims largely completed, the only standout is the bias error from the incurred claims estimate of IBNR.

Finally, to give another perspective on these results, because pricing actuaries and underwriters are usually more interested in total incurred claims than in IBNR, I've plotted the estimates of rolling 12-month incurred claims totals against the actual 12-month incurred claims. The relative performance measures of the different methods reflect what was observed for the IBNR estimates by themselves. The results from the completion factor and incurred claims methods are somewhat scattered, while the simple and regressed paid claims methods give a much tighter pattern. If we put that in a bar diagram (Chart 13), that pattern reflects the IBNR results as you would expect. The closed-panel MCO data again provide a similar pattern of results, estimates from the completion factor and incurred claims methods are relatively scattered, and the estimates from the simple and regressed paid claims methods lie closer to the perfect fit line, although they're a little more scattered, again because of apparent data problems in that set of data.

In summarization of those, the completion factor stands out as having poor performance relative to the others. In looking at the PPO/POS data, again we get a lot of scatter with the completion factor and incurred claims methods compared with the tighter clustering of the estimates from the simple and regressed paid claims methods. Summarizing the results for that data set, simple and regressed, the paid claims method yields a standard error of incurred claims less than 1 percent.

I have some conclusions. The completion factor method gives a poor estimate of IBNR liabilities with no run–out, in my opinion. The simple paid claims method gives a significantly better estimate of IBNR amounts than either the completion factor or incurred claims method, with the advantage of being easy to use. The regressed paid claims method gives better results than the simple method, but at the cost of some complexity. Finally, both the simple paid claims and the completion factor methods can be shown to be special cases of the regressed paid claims method.

**MR. DOUG FEARRINGTON:** I wanted to talk today not so much about a specific methodology or model, but instead about a class of models that I think have relatively good applicability to the type of data we deal with when we're doing incurred claim estimates or IBNR calculations. I'm not doing this because I thought it was a neat thing to talk about. A lot of what I'm about to discuss grew out of a goal of mine, which was not necessarily to produce a better mousetrap or a more accurate estimate of IBNR, but to produce some type of valid distribution around an estimate of IBNR. With a lot of the work that we did, it felt like we had debates over confidence levels with IBNR estimates that got down to: Do you think you thought of everything or not? That was my motivation in getting into this and why I'm here today. The approach and the ideas that we have here today are all around trying to come up with a way to put a valid distributional form around an IBNR estimate.

I thought a bit about how to do this. One straightforward way would be if you had a given IBNR methodology that was replicable, an algorithm or something like that, you could go back in time and run it again and again and again as you progressed at monthly increments of valuation dates and see how well it performed. You could use that series of residuals of past performance to construct some type of distribution around a given estimate. The question becomes what kind of distribution? Is it normal? Would you use a student's T distribution? The answer is that you don't know; you'd have to look at the pattern of residuals to see if they fit over time a sample from one of those distributions. I think what you'll find is when you get up to this level where you're looking at IBNR amounts, you're going to have residuals that are correlated over time, primarily because each IBNR amount contains a lot of information that the one just before it does.

Just looking at past performance on total IBNR to develop some sense of the distribution ended up being a little bit of a blind alley for me. Instead the idea was to look at each month of incurred claims and put a distribution around that, some type of technique that puts distributions around monthly incurred claims, and then we could sample from those distributions of each month's incurred claims to calculate other things of interest, be it IBNR or trends or whatever could pop out that's a function of incurred claims. That's the basic approach that's going to be outlined here.

When I think of estimating IBNR, I look at it as almost a forecasting process, but it's a little different from what we typically think of as a forecasting process because you do have some foreknowledge of what's going to happen. That's normally what we refer to as the given payments made on a month that's not complete. If you buy into looking at IBNRs as forecasting with a decaying amount of foreknowledge, that leads to a general set of models that might be worthwhile looking at. Specifically, I'm talking about time series models. If you were to do a forecast without any knowledge of the future, if you had just a series of data, I think this would probably be the first place that most practitioners would turn. They're straightforward, and it's generally thought a forecasting model.

Another nice thing about them is that they are stochastic in their specification, so that gets back to the concept we were talking about of putting distributions around things. That could be a good starting point. You can combine them in ways to maximize the information or content that's in each one of them. I'll get to a little bit more of that later.

The other reason that I think time series models as a general class are good for dealing with our type of data is that our type of data (be they on an incurred basis as you think about claims being incurred through time or with the way claims are paid if you think about just payment processes in general), you come across trend, seasonality and, to a large extent, autocorrelation. Let's say that we saw a trend only in seasonality in any given set of health data, as in a series of incurred PMPMs for medical claims over time. If that series had only trend and seasonality in it that were relatively straightforward to model, you probably wouldn't need to use a time series risk model; you could handle that with regression, a standard, ordinary squares regression model where you have a variable for time and maybe some indicator variables for what month it is. It's the autocorrelation piece that, at least in my experience, shows up in a lot of our data that makes time series models a good starting point to look at.

Chart 14 is a basic representation of an ARIMA model, which is a particular type of time series model. I use them interchangeably, which is a mistake. We have some type of variable here that we want to project or predict. That's our target variable or response series. Basically that target variable is a mixture of some constant mean plus this weird function of random error terms over time. If we were to eliminate this completely from this model, we're back in standard linear regression world, and maybe this would effectively be some type of slope parameter times some other independent variable or something like that. The key difference is this polynomial of factors that are multiplied in effect by this series of error terms, and this is where you incorporate autocorrelation of values across time.

The numerator and denominator of this ratio are explained a bit. The letter B stands for a back-shift operator, which means that if you have one B, take the value immediately preceding in time (instead of $a_t$ it's $a_{t-1}$). If you have B squared, take the value two periods of time back. It makes it easier to write out so that you can say, "Wow, what a simple formula," instead of, "Oh my goodness, what a mess."

The only reason I put this up here is not to drill into the theory behind ARIMA models, but to say that this is not a complicated model conceptually. You're just saying that you have some type of series that is a mix of a constant mean and a polynomial combination of back-shifted error terms. One thing I did want to mention is this response variable can be an actual value or a difference of the values, like a first difference. That's a handy way that time series models normally detrend the series.

Chart 15 is another time series model that's a little bit more complicated, and it must be more complicated because it has all these additional letters in it. All this is saying is that we have the same terms from before. You have your constant mean, you have these back-shifted error terms, and then you can throw in fun variables to explain behavior in the response series. That's what these $X_{i,t}$s are. These are independent variables that you know the values for. Guess what: you can back-shift those and use values at different times than the T that's indicated here. That's typically referred to as a transfer function model. You are transferring information that you have about this series through some function to predict this one, but if you were to eliminate a lot of this, you're back to regression world. Again, these two slides are to give you a flavor of the types of models that I'm talking about.

I'll get back to the task at hand—trying to come up with these distributions around incurred claims. The general approach that I've taken is I don't have any particular preference or bias over what I'm trying to predict. I'm going to try to build models for every single thing. Getting back to that $W_t$ in the formula before, your target variable can be anything. Separate time series models can be developed for completion ratios, monthly incurred claims (just meaning like PMPMs over time) or the incurred and paid amounts in a triangle, on a PMPM basis as opposed to the ratios. The whole point of doing that is then to combine all those different models that you build for these different target variables in a way that's relatively rigorous and to figure out how to maximize the contribution of each.

Once you've done that, one of the nice by-products of this is that you have these prediction error distributions around each estimate that you've made that form the basis for sampling.

When you're building these models, your goal is to specify a model whose residuals have expected value of zero, have constant variance, are uncorrelated over time and are normally distributed. This is not any different than if you were doing linear regression, but it's something that I think a lot of times gets overlooked. You can build any model you want, but if you've violated the basic assumptions of the model, then you're going to draw erroneous conclusions particularly around confidence intervals. How do you get there? How do you make sure that any given model that you specify meets those constraints?

There are all sorts of fun tests to do using the residuals of the model. I don't know the level of familiarity with these. These are pretty standard. There's the Ljung-Box White Noise Test and the Dickey-Fuller Unit Root Test for evaluating whether or not you've explained away trend and seasonality.

Beyond violating model parameters, you want to know how good they are. In building these models, you're going through this constant balance of goodness of fit versus robustness. Have I included too many terms such that the goodness of fit looks awesome, but it's never going to generalize to something in the future?

Let's go through this process, this general approach. It basically breaks down into: build your models, combine them and then sample from them to calculate your IBNR amount. Let's get a little bit more into building these models. The first thing that we've already discussed is you want to select some type of target variable. It could be a completion ratio for the second month where you were looking at the ratio of two payments to one payment and are trying to push that one payment up just one time period. It could be anything. You also want to include as many explanatory variables as you can for whatever it is you're trying to model. Here you're limited only by your own imagination and your ability to get the data. Chart 16 is an example of what a data set would look like. When I do this on my own, this is what my basic data set looks like. I have time values over here, and then I have all these different series. Some of these series are values that I'm trying to build a model on and forecast into the future, and others I have in there to help me explain process variance in the things that I'm trying to forecast. You could have completion ratios on back as far as you want, and those could be series that you're trying to forecast. These are a few explanatory variables that I typically include, like the number of payment cycles in a given processing month; we're normally either four or five, so it's a 01, depending on whether it's four or five. There's also the number of holidays. Obviously that tends to affect payment speed and sometimes incurral patterns. Then you can have all your actual PMPM values from a given triangle. Here we're talking about an incurred date, and then you can sum those up to everything that's been incurred and paid for that month as of your valuation date. This is a typical data set that I feed through and try to build models off of.

You're trying to predict that lag one completion ratio. I have one payment that's been made for this month—what's the second payment going to be? I have a series of those completion ratios over time, a long series. The first thing I want to do is examine that data set for outliers. Are extreme values occurring in there? What about trends, seasonality correlation and things like that? How do you do that? You could eyeball it. Particularly with trend, that's easy to do. Fortunately there are some relatively rigorous tests that you can do as well that involve looking at correlation across time, autocorrelation and partial autocorrelation going back.

The next thing that you get to do after you've cleaned up your data set and tried to think about what kind of patterns are in it is to specify a time period for fitting and a time period for evaluation. I don't care if you never use a time series model. You can walk out this door and completely purge anything I've said about time series models. There's one thing I hope you would retain—it's this framework for building models. In my experience, I haven't seen this a lot for any given forecast or for any given prediction that's made. There's a nice way of going about doing it, making sure that you're choosing a model that's appropriate and the best of the ones that you could have developed.

The basic approach is you have some historical series that you're trying to fit a model to. It needs to be broken up into a period of time that you're going to use to estimate the parameters of the model and then a period of time that you're going to use to evaluate the model. In the period of time that you're going to use to evaluate the model, you don't want the model to have ever been seen before. You don't want its parameter estimates to reflect what was going on in that time period. It's a hold-out sample, and that hold-out sample is going to be your basis for comparing different models that you've built. Say you're extremely imaginative and come up with 20 different models that you could use to forecast or predict a given series. How are you going to pick one? You're going to pick it based off of its performance on your hold-out sample.

There are a couple of measures that I tend to like to use to balance goodness of fit with robustness. One is root mean squared error. That's a goodness of fit measure. Akaike Information Criterion (AIC) is basically not that much different from goodness of fit, but it includes a penalty for the number of terms that are in your model. It's kind of a balance. Obviously you can increase goodness of fit by putting more terms in, but there should be some type of penalty for robustness that may not translate into a different data set that your model has never seen before.

As a crude illustrative example (Chart 17), which model would you pick out of these four if you had these measures? I'll tell you this—AIC is better the lower it is. Anyone? I think it's a hard choice. I would pick maybe A or D. For B, there's a substantial improvement in root mean squared error, but I don't know. I don't know about that. I might be worried that I have a model that's not going to translate well. I would definitely rule out C. For C, I would say, "I'm sitting pretty well, but I've penalized myself by including a lot of terms." The whole point is that it's tough to know. There's some judgment involved. You could maybe argue over two or three of those. The point is to do the exercise, and then you can defend which one you pick.

Once you've gone through that exercise, another part that you have to think about is evaluating the residuals that each model produces, and those are its past errors. That gets back to making sure that you haven't violated any of the assumptions built into the model, that your residuals are expected value zero, normally distributed, have constant variance and are uncorrelated over time. That doesn't have much of anything to do with goodness of fit or robustness; it's just a well-specified model.

Even though that looks relatively cut and dried, and you have measures to help you out to make choices along the way, just like in the example we saw, there is judgment that comes in. The first thing is how much data you have to use. The answer to this question a lot of times raises eyebrows. Let's say I'm trying to forecast PMPM claims. I have some data set. Let's say I have a good sense of actuals, maybe 18 months prior to evaluation day. I would go back in time maybe seven years to build a model to forecast PMPMs. That's the one for which I don't have that much data. The reason I'd use that much data is I'd take maybe five years to do my fit and two years to do my evaluation. Five years to do fit is necessary because of seasonal effects. If you only have two years of data, I would argue that's not enough to estimate any type of seasonality parameter and trend changes, as we all know. That's my two cents on how much data you have to have.

All these other things too are difficult sometimes to pick. When you're kicking out outliers, that's obviously an extreme value. How do you know? What if there are a couple that seem relatively close? I think the more important part on that is that at least have some reason for doing it and make sure it's something that's relatively defensible. Specifying fit and evaluation periods can be tricky. Sometimes there could be stuff historically you think there's no way that's going to happen again, and you might want to kick it out of your fit period so that it's not reflected going forward. That one can be tough.
Knowing when you're done is difficult, at least if you're like me and want to analyze stuff to death. Again in this world where we have 20 models to choose from, picking which one is the best and not thinking about it anymore can be tricky.

Let's say that we were going to try to use time series models with just the traditional completion ratio approach. What would we do? We still have this conceptual framework that we're going to use ratios. The question is: How are we going to estimate future ratios that haven't happened yet? It's relatively straightforward to simply build a model for each past series of ratios and forecast it forward. Again these are forecasting models; that's what they're meant to do. In practical effect, I've picked the time period maybe 12 or 13 months back where I don't look at ratios. I don't build time series models for the ratios anymore; it's not worth it. You also run into the problem that these ratios are constrained normally to be greater than or equal to one if you ignore negative adjustments. As you get further and further back in time, are these residuals really normally distributed if the target variable can never be less than one? You're going to have to pick some point where you say, "I'm not going to do this." Then it's straightforward. Once you've built your models to that forecast, the product of the forecasted ratios is applied to whatever has been incurred and paid for a given incurred month just like always.

Getting back to those distributions around incurred claims, I then go back in time and say, "How well would this have done historically?" If I used the product of all of my completion ratio models and only had assumed that one payment had been made for all past incurred months, how well would this product of models have done? Look at those residuals. Those residuals form the basis for estimating a standard deviation of the prediction error, and that prediction error forms the basis for putting a distribution around things. You still have to make sure now that you've taken all these products that these residuals over time are still uncorrelated. Do they have any unexplained bias or trend? Are they growing over time?

For the most part in my experience, if you do that on the ones for each specific target model, things tend to work out even when you multiply them all together. I have to say though that an additive model, like Rob says, is better. I'm not trying to endorse or slam any given approach, but it is true. Multiplicative error terms introduce a lot of headaches when you're trying to make sure that you have uncorrelated residuals across time.

Chart 18 is a quick example of the fit. This is a series of lag one completion ratios. You're taking two payments to one payment over time. The little triangles are the results of fitting a model and estimating all the parameters. This particular model is a second-order autoregressive model, which means that when it tries to predict the completion ratio, for lag one it looks at the two values previous to it. The "I" means it's integrated, which means the whole series had first differences taken to eliminate trend. It has three different regressive variables in it, which you could argue one way or another whether or not they're relatively complicated. There's a simple one in there for the number of payments that were made in a given processing month and then the lag two completion ratio, which he had the month before, and then a little one that I like to put in there that looks at the changes and skewness in the payment factor over time.

The actual terms in here aren't important. This is not some endorsement that this is a great model for lag one completion ratios. It's just an example, but as a result you can easily produce a forecast going out to plug in. For a lag one completion ratio, we need only the one-step-ahead forecast; we only need one value.

Now I'm going to look at historical PMPMs and try to do a projection to give me a basis for judging whether or not I believe my completion ratio models. That's about the easiest series there is to develop a time series model for, again because it speaks to all the issues that usually go on with those things—trends, seasonality and autocorrelation. Typically what I do is just pick a point in time back like normal and deem it to be either complete or apply traditional method to it. If it's 12 or 13 months ago, it's going to be close enough. I use that historical set to build a model and project forward. Chart 19 is an example of doing that. Again these are PMPM values now and not completion ratios, but the dots in this example are an ARIMA model, first order across all the terms. For that model that we just looked at, it would be easy to put in some seasonal dummy variables or other regressors. There are different variations of the basic ARIMA model approach that you can use.

Let's say that I had stopped there, that I had built time series models for completion ratios and projected those and had nice little standard error values for each of those. Then independently I developed a model to project PMPM values going forward from some past actual date. How would I combine them? I think a lot of times what goes on in our heads as we're doing this valuation work is that the combination of these models is essentially just judgment for each month going back. I think there are some different ways to do it that are at least a little bit more rigorous. We'll go back in the past and try to figure out what the optimal combination of the models would have been to perform on whatever incurred month you're trying to estimate. Typically what I've done using my historical data set is go back and develop coefficients for a linear combination of two sets of answers—the PMPM answer and the completion ratio answer—for a given incurred month and try and minimize some error measure. I usually do sum of squared error.

The coefficients for combining the PMPM model and a lag one completion ratio model are different. All I mean by that is if you go back and look historically at how well your completion ratio models would have done if they had only known one payment, it's probably going to be pretty bad as Rob had pointed out to us. How they do if they had known two payments is going to be a little bit better, three payments even better and progressive improvement thereon, so the amount of weight that you would want to apply from your PMPM model would decrease over time as you knew more about the month. When talking about these completion ratio models, you have a series of different models that you go back independently and combine with the PMPM model to figure out what set of weights you should use. The whole point of that is to go back historically and try to minimize some error measure to figure out your combination. That would be my only guidance.

Let's say that we've done that. We've figured out how to balance these two competing set of models for each given incurral month of our valuation date. We have a predicted error distribution around each one of those estimates. Now we sample from each of those distributions and take the sum of the incurred estimates, subtract off the paid claims and have our IBNR. Each time we sample, we recalculate that value. It's a little bit harder than that. In sampling from these predictive error distributions, there still is the possibility of correlation of the specific prediction errors. That could be because even for a given model, if you've made sure that that won't happen, if you have two different models even for two different values, they can be wrong in the same kind of way even if they were developed independently.

That's why I suggest specifying a covariance matrix to handle that sampling. It can be difficult, and the truth is it is difficult. We could talk about it for maybe six hours.

One simple way to address it is to set your correlation of your sampling at one. It's the most conservative thing you can do because everything is positively correlated. You get the biggest tails that way. It's probably a little bit overconservative, but if you don't have any other guidance or can't develop that, it's a simple way to solve the problem.

There are some implications for doing this. Going through this model-building framework, taking a large historical data set, testing lots of different models and lots of different approaches and getting a sense of how wrong they can be in estimating the parameters, it divides up what I tend to think of as a typical month-end process. Normally you're waiting until you get your most recent data, after your last payment cut-off, and then you have five or six days to go crank. The models that you're using need the most recent information that you have because you know that they're all going to be wrong, so you have to go through and figure out how wrong it is. You have to have that most recent information. If you've gone through this process of developing these models, the answers for which models are best and even to some extent what the parameter values are associated with them, these are not going to change with an additional month of information. You can divide up your work time. On the downtime, on off-cycle times, you can spend that on model development, and then once you get your most recent data set as of a given evaluation date, it's a little more plug and chug. You reestimate parameters, but you don't change model forms and just recalculate.

I've tried to anticipate at least some questions by providing answers to questions that I've gotten over time. Is this approach any more accurate than a traditional method? The truthful answer is for a given month, not necessarily, given that these things are random variables. If you were to put up a set of models that were developed over a past set of data versus a seasoned actuary, I would say in a given month a seasoned actuary could do better, but over time I think that the probability of being right is much higher. The other thing too is that the point isn't so much to produce an exact right estimate, but to quantify the amount of error around that estimate. The thing that you end up worrying about is whether these confidence intervals that you're establishing are right or accurate.

The next question is: Do you have to think at all? Can you just push a button and be done? No. I think we've briefly touched on a few points going back where there's a good amount of choice that's still involved, particularly in that model-building phase. The nice thing is you can do that up front, and it's a one-time price for pain. Is it a lot harder, and does it create more work? Again, most of the work is up front.

Can a time series model predict something that's never happened before? What about large scale process changes? I think all of us have lived through a new payment system being implemented or some type of trend over time of shifting to more electronic claim adjudication versus paper, things like that, and the answer is it depends. If it's something that truly has never been seen before, which can be difficult to establish in a lot of cases, you're going to have to most likely handle that as an explicit adjustment. There's not going to be anything in your historical data set that's going to reflect that occurrence. Once it starts to manifest itself in your data, whatever the change was, after a period of time, there are ways to model this intervention or change in your series with time series models.

To me the harder question to answer is: How do you know it's never happened before? We get hit with that all the time. Is it really some type of new thing, or is it already reflected in the process variance that we have in our series? I don't have a good answer for determining that. I think there's judgment involved there.

What software do you use to do all of this? I don't use anything all that fancy. It's a combination of Excel and PC SAS. Particularly there's a module in SAS for time series—ETS. I used @Risk just for the sampling. It's an add-on to Excel. Most of the model development work is done in SAS itself. On a production basis, that's used to reestimate parameters for models that have already been through some testing. On a month-to-month basis, using SAS doesn't occur that much.

Can you use this approach of time series models for anything, not just claim payments? What about days, admissions, visits or any kind of unit cost figures or costs for admission? The answer is yes. Again you're going to go through that whole bill process to make sure what you're doing is appropriate, but there's nothing specific about this that just would be applied to dollars. From the sampling that you do, in fact, your forecast can reflect a lot of the potential starting error that we're normally trying to explain away as we're four or five months into the forecast, so you have a random starting point that you can sample from to model to and go forward.

I have a few pros and cons. The time series approach creates an actual modeling framework for estimating IBNR. You can replicate it. You can go back and say this is specifically why I chose what I chose and end up with some guidance that typically isn't available from deterministic methods. Some of the minuses include the fact that you need a lot of data, and it gets difficult to figure out if you have enough. Once more, not all of the potential error that could be occurring is captured in the confidence intervals produced. There is parameter uncertainty. Even though you're going with a mean estimate of a given parameter, there's variance around that. You could have picked the wrong model—you didn't do a good job in your model specification period, or the data that you're using, even though they're part of a large set historically is not at all reflective of what's going forward. That's a risk that most models run.

Another thing that I've run into is that you can give an inexperienced audience a false sense of confidence. When you start saying you looked at the 97$^{th}$ percentile this month or something like that, for some reason that conveys some type of wizardry that isn't there. It's been ironic. I think that we were pursuing these approaches to limit the amount of discussion we had about uncertainty at a given month. If anything, it has only increased the number of conversations that we have.

Anyway, that's a fair warning.

Streamlining and automating this process of going through this and building things up can take a lot of time. You're looking at probably at least a few months of developmental time and then a lot of parallel testing.

I have additional materials I can provide to people. I can be contacted. I'm in the directory. I can provide prototype spreadsheets. I've submitted a paper to the *North American Actuarial Journal*. If anyone wants to see a draft of that, I can provide it.