RECORD Volume 30, No. 1*

Spring Meeting, Anaheim, CA
May 19-21, 2004

Session 18PD
Stochastic Modeling for Health Actuaries

Track: Health

Moderator: Darrell D. Knapp

Panelists: Doug Fearrington
Charles S. Fuhrer

Summary: This session presents a summary of actuarial and potential uses of stochastic modeling for health actuaries. Attendees learn the types of issues health actuaries may want to consider using stochastic tools to address.

MR. DARRELL KNAPP: This is a session on stochastic modeling for health actuaries. We have three different presentations that are each going to present a problem and describe the specific modeling used to produce the solution. So, we'll do a little bit of learning by example and case studies. Hopefully that will be an effective learning style for everybody.

Our presenters this afternoon are the following: Chuck Fuhrer with The Segal Company will present a method to rate health plan reimbursement provisions and some modeling analysis. Doug Fearrington, who is with Anthem Blue Cross and Blue Shield, out of Richmond, is going to do a presentation on time series analysis and developing claim liabilities. Then the third presentation will involve looking at reinsurance models and how to use the stochastic modeling approach to help define what the appropriate reinsurance level is in a health plan. I'll let Chuck introduce himself and give some of his background.

MR. CHARLES FUHRER: Thank you, Darrell. I've been working on a problem that has bugged me ever since I started in health insurance back in 1973. There was

this manual, and I was told that it was a sacred book that allowed the underwriters the opportunity to figure out what a plan change and the health coverage was worth. I thought that was great. The manual weighed about 273 pounds and had lots of tables in it, which you used to multiply and divide and sometimes raise things to powers. It was great, and the answers it gave were perfect. After I was with the company for about six months, they took me aside and said I would have to update that manual. I asked how do I do that? And they said, "We don't know. The actuary that did it has left, but it couldn't be that hard. Just do some modeling. We've got lots of data here."

This was during an era when we had those big mainframe computers, and you had to go to the computer department and get on your knees and beg them for data from the data warehouse. So things have improved a little bit. But, forget my presentation title, it's a method to do manual rates. Fortunately we have PCs today, so we don't need those big, heavy manuals (Figure 1).

Figure 1

The Reimbursement Function

- *r(x)*, the amount paid to the insured for allowable charges *x*.
- Basic Properties
  1. continuous
  2. non-decreasing
  3. $0 \le r(x) \le x$
  4. $0 \le r(x_2) - r(x_1) \le x_2 - x_1$
- Usually
  1. piecewise linear
  2. $r(x) = c(x-d)^+ + \left(1-c\right)\left(x - d - \frac{l}{1-c}\right)^+$ for constants: $0 < c < 1$, $d > 0$, and $l > 0$. $x^+ = \max\{x, 0\}$.

We have a lot of complicated benefit formulas with deductibles and out-of-pocket limits. I figured that if I could write an equation for how much the plan reimbursed, given so much in cover charges, that would be the first step. I said that if we have cover charges of *x*, then the plan will reimburse some function of *(r)x*, and it had some basic properties.

It's continuous. For example, if somebody spends an extra dollar, we don't suddenly reimburse him or her five more dollars. Generally, it's also non-decreasing. If you have cover charges that go up, then you don't end up getting

paid less. You don't get two dollars extra reimbursement for a dollar extra cover charge. In fact, it's usually linear, at least piecewise. We don't have quadratic equations, like a square of the amount of the cover charges.

Then I got this nasty little formula where a typical plan says that you subtract the deductible, and if that's positive, then you pay a coinsurance of "*c*." Then if you have an out-of-pocket limit of *l*, you divide by 1 minus *c* and subtract that from the cover charges and deductible. Then you pay 1 minus *c*. So, I have this nice, easy formula, a deductible, a coinsurance, and out-of-pocket. I've made a really complicated formula out of this nice explanation. But we want to come up with a manual rate.

Typically we have a huge amount of data—at least if we work for Blue Cross or someplace like that. I don't have quite as much data as I used to, but in any case, it's not quite as big as we thought it was. First, we had to subdivide it by area and then maybe by age and sex, but the main thing is that it is also different by plan. Then there are some other problems with the data in making manual rates. Generally we don't have negative claims, although there are sometimes reimbursements. It's highly skewed; therefore a lot of claims are very small. Then as we get higher it drops off quickly. But there are still thick tails in the sense that even though we have mean cover charges of a couple thousand dollars, there is a pretty high probability of spending $100,000, which is 50 times the mean, but a surgery or a long hospital stay can easily get you there. So, it's not that uncommon. That's called a thick tail distribution.

How do we do our manual rating now? The first thing we do is estimate the value of the expected claims under some reimbursement arrangement. To do that we usually put together all of our claims. . We then come up with a size of claim distribution, or basically we just repay the claims with the reimbursement "*r*." We take the average of all those claims, and that's our answer. We use a smooth version of it. In fact it's on the syllabus: Stuart Klugman's book on loss distribution. We have this huge amount of data, and we fit a curve to it.

So now we have a nice distribution where we know that the deductible is worth so much, and the out-of-pocket limit is worth so much. But the problem is that we know that the reimbursement actually affects the claims. I just saw something on the news about how someone said that higher co-pays are causing people to not fill as many necessary prescriptions, and I thought, "We've known that for years." But the doctors were upset because they thought these were necessary prescriptions, but the point is that we know very well that if you reimburse more, you're going to spend more.

In fact, one of the problems we had with our plans way back when was that we were paying 100 percent of everything. With that blank check people found a way to use it, and it wasn't clear that it was of any value to people. It was sort of "overcare." But I'll leave the value judgments aside here. We know very well that

we put deductibles in to keep the cost down, so people will think twice before going to the physician. So they'll only go when it's necessary, or maybe they won't go when it is necessary. The point is that this is a very significant effect, and we cannot ignore it. This method, unfortunately, does ignore it. We look at the mean from each of our plans, and then we guess and we fit a smooth curve to it and say, "Well, if the coinsurance goes from 80 percent to 90 percent, we're going to make it a little more. It's all kind of ad hoc, and yet it seems to me that there is something wrong with it.

We should be able to let the data speak for itself, at least as a first try (Figure 2). Then, if it doesn't look good, we can always adjust it later, but let's do something. So, one problem is that when we do number one we have different plans in the data. So we could get different answers depending on what plans we had in there. Then the problem is that there was a big disconnect between the two steps—we do one thing, we push all the data together, and then we do this other thing where we're looking at the plans separately.

Figure 2

Current Methods

Two Steps Method

1. Estimate Values of $E[r(X)]$ using either empirical (i.e. just repaying the claims) distribution or a smoothed version based on all claims.
2. Trying to estimate effect of plans on $E[r(X)]$ by looking at me from each plan

Problems with Method

1. Different mixes of plans in data
2. Disconnect between two steps

Finding an independent variable for step 2

So when you're fitting that curve, do we use coinsurance? Do you use deductible? I know some people have done number two. They come up with this average coinsurance. I guess it's the average amount paid. But that implies that a deductible has the same effect; if it's eating up 10 percent of the claims, it's going from 100 percent to 90 percent, but we know that's not true. I'm not sure which way it goes, but the deductible *i*s going to stop people from that initial visit to the physician when they don't need to, whereas the 10 percent coinsurance is going to cause people to think twice about more expensive procedures. But it doesn't really stop them from going to the physician initially. So, it's going to have a different

effect, yet people will put it all together. We don't really have a very good independent variable to measure where these plans stack up.

We're basically doing curve fitting on a whole set of curves, which, I don't doubt is a possible method, but I wanted to introduce another method of doing this today, which I think is pretty straightforward to understand. The formulas here get a little nasty, but let me just explain what's happening (Figure 3). My suggestion is that we go ahead and repay the claims, just like we did under step one of the prior method. But instead of just using a simple average of all those claims, let's weight those plans—the ones that are closest to the plan—based on how close they are to the plan we're trying to rate. So, if we're trying to rate a plan with an 80 percent coinsurance, and we have some information that's 90 percent, 80 percent and 70 percent, then we should weight the 80 percent coinsurance plan the highest and give a lot less weight to the 90 and to the 70.

Figure 3

A Solution

Repay claims at each data point then weight data points based on the distance between the plan they were incurred under and the plan whose mean is being estimated.

Weights: $w_{i,j} = \frac{1}{mn_j|\mathbf{K}|} K_r\left(\frac{\mathbf{r}'(\mathbf{x}_{i,j}) - \mathbf{r}'_j(\mathbf{x}_{i,j})}{\mathbf{K}}\right)$. $\mathbf{r}'(\mathbf{x}_{i,j})$ and $\mathbf{r}'_j(\mathbf{x}_{i,j})$ are the vector gradients (coinsurance rates)

Then for $\hat{\mu}_r = \frac{1}{w}\sum_{j=1}^{m}\sum_{i=1}^{n_j} w_{i,j}\mathbf{r}(\mathbf{x}_{i,j})$ where $w = \sum_{j=1}^{m}\sum_{i=1}^{n_j} w_{i,j}$

For suitable Kernel function: K_r and constant (matrix) $\mathbf{K}$

The thought here is that we've based that weighting on the dollar amount of each claim. So we would actually use the current coinsurance percentage of the plan in which the claim was encountered. We have this *j*'s claim that had a reimbursement function of *r* sub *j*. So this is actually the coinsurance that was for the particular data point, and this is the one we're trying to value. We're going to base a weight, and we're using this Kernel function, which is just a sort of peaked function, a bell-shaped curve. So we're going to give a greater weight, the closer these two are together.

This method is basically like nonparametric regression, but there is a difference in that we're actually doing the differences between the plan. But don't worry about the formulas, the whole point is that we're giving greater weight to the plans that are closer to the one we're trying to value, so the points that have coinsurance; that *i*s closest. And I'm assuming that coinsurance, or the rate of change to the reimbursement, is actually the prime variable. So, for example, if this point here was at 90 percent coinsurance at the place where it was encountered, and we're valuing a plan that's 80 percent, then the difference here is going to be 10 percent, and this *K* function is going to give that less weight than if these two matched.

We can vary how spread out the weighting function is. For example, if we make it very spread out so that it's very even, that would be equivalent to giving equal weight to everything and then we're back to number one of the traditional method. On the other hand, if we make it very peaked, very thin, then we're only giving weight to those values that are very close to the plan we're trying to vary, and the problem there is if we don't have one that's very close. We don't want to throw away too much of the data. But the point is that we're going to vary the weights depending on how close we have values so that we can actually get a better answer here. Then we'll divide by the sum of the weights, and if we find that the data supports this, that is, that there are greater means when there's greater reimbursement. Then we will get a bit greater means for the plans that have larger reimbursements than the other way around. So, this will automatically do that change in the reimbursement, and we'll actually see that our answer will be moved over.

I've got some notes on what these Kernel functions are, but they're not important (Figure 4). But we're no better off if we have bad data. In other words, if we have outliers that show very low means, even though they have very high reimbursements, or vice versa (very high means) even though they have low reimbursements, then that's obviously going to throw this off, but we're no worse off than when we started. Those would have created the same problem in step two in the traditional case.

Figure 4

Notes on Using Kernel Functions

These functions give more weight for smaller values

Desirable Properties

1. Symmetric so $\int \mathbf{x} K_r(\mathbf{x}) d\mathbf{x} = 0$
2. Positive $K_r(\mathbf{x}) \geq 0$
3. Normalized $\int_{-\infty}^{\infty} K(x) dx = 1$.

$\boldsymbol{\kappa}$ controls the amount of smoothing in the weights. Very high values of $\boldsymbol{\kappa}$ and the method reduces to the old method of a simple average. Very low values just uses the closes plan.

Let me just talk a little more, and for those of you who like the math, there it is (Figure 5). For those that don't, let me just explain what we're doing here. The problem with the method of just repaying the claims is that we can have some very sparse data, and we want to do some sort of smoothing. That's why we do the stuff in loss distributions. The same thing applies here, except that we don't want to lose what our contribution was from each point. My recommendation here is to use a different method of smoothing, and this is a very well-known method. It's called density estimation. Instead of fitting a curve to all of your points, you take each point and fan it out into another bell-shaped curve or something like it, another Kernel function. So instead of having a single point where it was actually encountered, you assume that there were more points around that point that could have occurred. Just the sample size wasn't large enough, and you fan it out that way. Then the process is the same, except now you're integrating over that curve, but it's really just a very straightforward extension of the method.

Figure 5

Better Solution - Smoothed Version

Let: $\tilde{f}(\mathbf{x},r)=\frac{1}{m}\sum_{j=1}^{m}\frac{1}{n_j}\sum_{i=1}^{n_j}\frac{1}{|\mathbf{H}_{i,j}|}K_p\left(\frac{\mathbf{x}-\mathbf{x}_{i,j}}{\mathbf{H}_{i,j}}\right)\frac{1}{|\mathbf{K}|}K_r\left(\frac{\mathbf{r}'(\mathbf{x}_{i,j})-\mathbf{r}_j'(\mathbf{x}_{i,j})}{\mathbf{K}}\right)$ for suitable kernel functions, $K_p:\Re^d\to\Re_+$ $K_r:\Re^d\to\Re_+$, smoothing matrices $\mathbf{H}_{i,j}$, and $\mathbf{K}$. Choose the smoothing matrices $\mathbf{H}_{i,j}$ using an adaptive method (more smoothing in the tail).

Here $\mathbf{r}'(\mathbf{x}_{i,j})$ and $\mathbf{r}_j'(\mathbf{x}_{i,j})$ are the vector gradients.

Set $\hat{f}(\mathbf{x},r)=\frac{\tilde{f}(\mathbf{x},r)}{\int\tilde{f}(\mathbf{x},r)d\mathbf{x}}$

Then estimate $\hat{\mu}_r=E[r(\mathbf{X}_r)]=\int\hat{f}(\mathbf{x},r)r(\mathbf{x})d\mathbf{x}$

Let $w_{i,j}=\frac{1}{mn_j|\mathbf{K}|}K_r\left(\frac{\mathbf{r}'(\mathbf{x}_{i,j})-\mathbf{r}_j'(\mathbf{x}_{i,j})}{\mathbf{K}}\right)$ and $w=\sum_{j=1}^{m}\sum_{i=1}^{n_j}w_{i,j}$.

Define: $I_{i,j}(r)=\frac{1}{|\mathbf{H}_{i,j}|}\int K_p\left(\frac{\mathbf{x}-\mathbf{x}_{i,j}}{\mathbf{H}_{i,j}}\right)r(\mathbf{x})d\mathbf{x}$

Since $\frac{1}{|\mathbf{H}_{i,j}|}\int K_p\left(\frac{\mathbf{x}-\mathbf{x}_{i,j}}{\mathbf{H}_{i,j}}\right)d\mathbf{x}=1$

$\int\tilde{f}(\mathbf{x},r)d\mathbf{x}=w$,

$\hat{f}(\mathbf{x},r)=\frac{1}{w}\sum_{j=1}^{m}\sum_{i=1}^{n_j}w_{i,j}\frac{1}{|\mathbf{H}_{i,j}|}K_d\left(\frac{\mathbf{x}-\mathbf{x}_{i,j}}{\mathbf{H}_{i,j}}\right)$,

and $\hat{\mu}_r=\frac{1}{w}\sum_{j=1}^{m}\sum_{i=1}^{n_j}w_{i,j}I_{i,j}(r)$.

So the expected value is a weighted average of the integrals.

I just wanted to present that as something to think about. I was going to try and get some data together, and I just didn't manage it. I'm going to try and do this with dental. It's a lot easier because there are four different types of cover charges instead of medical where there are somewhere around 100. So, my plan is to get some data together and probably make this into a paper for the *North American Actuarial Journal,* which will give some more practical examples.

MR. FUHRER: It's time for questions.

FROM THE FLOOR: You're talking about using this kind of a technique to come up with new factors for a rate manual or a new way to come up with a final answer. How do you get to the what-the-underwriter-can-understand version of this? I see how you can come up with nice ways of modeling the actual claims appropriately, but how do you get this into something applicable that the underwriter can look at and make a good judgment about because when you get into Kernels, they're just going to shut the door.

MR. FUHRER: Right. The first thing you do is you make a manual or, in this case, a computer program that does these calculations for them. Second, you do the calculation two ways, one using the simple average, and the other way doing the weighted average, and then you present that to them. "Well, here's what it would have been if there had been no utilization changes, and here's what it would be due to the utilization change." The manual itself is a black book anyway. So there's not that much difference between doing it with a program that does this calculation as to the old manual method, whether it was done on a PC or by human calculation, but, in any case, if you give them those two facts, then they have the outcome, and they're really no worse off than they are under the current system.

FROM THE FLOOR: And then a second question, I'm sure I'll offend somebody, but when it comes to filing rate changes, when you get to more complex methods like this, which are actually technically more defensible, do you see problems with them being more difficult to explain to a state rating agency?

MR. FUHRER: Well, I don't think so. Right now, depending on the agent and on the policy, you are allowed to make assumptions for utilization changes based on plan changes, and right now I don't think you necessarily have to give much supporting evidence—just send them off to the state government and say, "This is what we assume the effect would be" and if they look reasonable. So, I don't know if you're any worse off. You did it this way, and this is what you came up with, and I don't think you'll have a problem with them asking what the calculations are under there. Getting these formulas and saying forget it; I just don't see it going to that point. My experience with the regulators is that they expect you to be guessing at these things anyway. So now we have a method of a little better guessing, which is really what stochastic modeling is all about anyway.

FROM THE FLOOR: When you're saying data points are you talking about policy-year claims for a member or individual claims?

MR. FUHRER: I'm talking about policy-year claims for a member, and I'm using *x* for the cover charges.

FROM THE FLOOR: The second comment I have is when you said if you have anomalies in the data, it doesn't really matter because you're no worse off than today where everything gets an equal weight, isn't it true, though, that if your anomaly happens to have the same benefit design that you're trying to price, that you could give it proportionately higher weight—so that anomaly could completely drive your results if it was in a much smaller set? In other words, if you eliminate or weight down, say, 80 percent of your data because of the disparity between what you're pricing and what that data had, but that one anomaly had a very high weight, it could completely distort your results, right?

MR. FUHRER: Yes, and presumably in the situation where your data was sparse, you would mostly correct for that by using a wider averaging so that things got closer to even, which, of course, puts you right back where you started.

FROM THE FLOOR: Right.

MR. FUHRER: One cure, obviously, is to look at the data and decide where your outliers are and somehow reduce the weights on them. You can either do that on an ad hoc basis, or there's probably some statistical technique for that, that could apply here. There are, actually, but I'm not sure how they would apply in this particular circumstance. But you're right. That could be a problem.

FROM THE FLOOR: How do you handle, in the Kernel function, that it seems like you have some of your data points, which due to a richer benefit plan would have more claims than you'd expect under your benefit plan, your pricing, and some of your data points that would have less? Is that necessarily a symmetrical shape that you want your Kernel function to be in, or could you end up wanting an asymmetrical shape to that function?

MR. FUHRER: Are you talking about the one where you're doing a smoothing in the second part or the one where you're actually re-weighting by difference in the plans?

FROM THE FLOOR: I guess the re-weighting.

MR. FUHRER: I don't really see any need to use an asymmetric Kernel. The problem does come up if you're not actually fitting a curve to it. So, if you're on the borders of your data, then you may not be getting the right answer, and that's a problem. It's important to try and have as many plans as possible that bracket each of the plans you're valuing, so you'll get some sort of average that makes sense.

And sometimes having enough groups (i.e., enough plans) is hard because a lot of times you think, "Well I've got this huge group, and I could use lots of data, but it won't work in this case.

MR. RAY MARTIN: Would there be any advantage to looking at the large claims where there's no influence on out-of-pocket expenses or payments by the individuals as taking those out, and that part of your continuance, making that separate, and then just looking at the part of the table where out-of-pocket is going to influence the utilization more?

MR. FUHRER: Actually, this method prevents that very thing from happening. In a typical analysis, when you finish with step one, if the very large claims just happen to appear on your plans that had high coinsurance, then those would have a high mean, and you would think, incorrectly maybe, that there was a big influence of that coinsurance on your answer. In this case, all of those claims, which are paid at 100 percent coinsurance because they're all past out-of-pocket limit, and the plan you're valuing, all have 100 percent coinsurance at that point, and, therefore, you're giving them equal weight in your analysis. So in that sense, I think this actually corrects that problem almost automatically because we're no longer subject to the randomness that has nothing to do with what we're looking at. We don't care how many really big claims a particular group had. It tells us nothing about what the deductible has an effect on, but it has a huge influence on those means that we're comparing. This takes care of it. It's not going to have any influence at all because they're all paid out at 100 percent because they're all past your out-of-pocket limit. So, I actually think that this method does exactly what you're looking for.

MR. DOUG FEARRINGTON: I'm an actuary with Anthem Blue Cross and Blue Shield of Richmond, Va. Today I'd like to share some of my experiences in attempting to develop stochastic approaches to reserve estimation. This might be a little bit more of a story than a case study. It is something that I've been actively working on for a little bit more than a year, but I've gotten to a point now where I think there's probably some validity to pursuing this. This is probably not the ideal forum to discuss models, which is ironic because that's why we're here, particularly with technical details. So I think you'll find that the thrust of my presentation is a bit more conceptual.

Another thing that I'd like to emphasize before we get into the meat of this is that I don't intend to endorse or proscribe these approaches as "the" ways to go about doing this. I think that these show some promise and that there's some validity to it. There are also several challenges and problems with it as well, which probably many of you will have comments on.

This is my attempt to come up with a way to put some valid distributions around incurred but not reported claims (IBNR) estimates or unpaid claims estimates. Our goal was to be able to say with some degree of confidence, "Where are we? If we

book this number, what percentile does it represent out of a possible distribution?" In pursuing that goal of trying to come up with these models, I think that we've stumbled on some other realizations about our general process that have value in and of themselves and I think are interesting regardless of whether you even want to pursue trying to come up with distributions around IBNR. So I'll try and highlight those as I go through.

Our basic approach in trying to come up with these distributions around reserves is to develop models that make distributional statements around incurred per month per member (PMPM) estimates, going back for some period of time, at least to the point where you feel like it's not worth doing this anymore. If we could do that and then somehow sample from the distribution in a way that we could add everything and calculate other measures of interest, specifically reserves, but also trends, that might be a way to do it. I can tell you that our initial thought was to just figure out a way to put a distribution around reserves, and who cares about incurred estimates from month to month? We ended that quickly, primarily because normally you want to know what's going on in a given month. Where do you have budget variances? Where are things more uncertain than others? So, most of what I'm about to describe essentially has the goal of making estimates about any function of incurred claims, the actual incurred claims in a given month themselves, the reserve estimates, trends and things like that.

The modeling approach presented here is essentially just a time series model approach. The whole concept, when you're formulating a time series model, and by a time series model I mean something more formal—we're talking autoregressive, moving average model. The whole concept is that your error term is essentially the thing that you're interested in. That is the stochastic variable of interest.

And the goal is to specify a model that has residuals with certain properties. You want a white noise process, where the errors at any given time are uncorrelated with each other with a constant variance mean of zero. The way you go about figuring out whether you have a model that is essentially fit to a process with these properties is a series of diagnostic tests that are associated with time series models. That's really what you're going to use to check whether you're doing a good job of building the model. It's not so much whether it feels right, it's that you have some tests that allow you to know how well you're doing.

The other key things when you're building time series models are the error measures involved for evaluating goodness of fit and robustness. It is a balancing act you get into. Obviously, the more variables or terms you include in a model to a given historical dataset, the better job you're going to do of fitting it, with the tradeoff being that it's not going to generalize well, and there are a variety of error measures that attempt to either talk about one or the other or balance both.

In going about this, I think that any time that I've talked about models for IBNR or unpaid claim reserves, people get fixated on the target variable. Is it a completion

ratio, or is it an incurred amount for a given month, an incurred PMPM amount? What is essentially the thing that you're trying to produce? Almost every method I've seen takes the results of forecasted statements for those output variables and then combines them in some way to calculate a reserve. I don't have any particular affinity for one method or another. I think that the point is to go in and say, "Let's look at a variety of methods and see how we can combine them."

So, some of the examples that I have are a bit more specific, and we'll talk about how to combine them as well. Most of what I've done has been centered around three basic target variables, actual incurred amounts, completion ratios and then if you imagine a claims triangle or a lag triangle, the individual cells within it, an incurred and paid amount for a given time. Normally we're looking at building models for all three of those possibilities and then looking at ways that they could be combined to maximize contribution of each. Then once you get done with that, and you have your ranges around all of your forecasts and incurred estimates, go sample and then keep track of the values each time to calculate whatever it is you're interested in.

Let's talk a little bit about the basic framework for how you go about doing this. Most of the comments are going to fall into these three categories: build, combine and sample. Let's look at the first one, building. Obviously, the first step is to determine what are you trying to build a model for? What's your target variable? And then you need to go through some type of evaluation process to identify what type of model would be valid at all, and that means looking at your data, looking at trends in it, examining it for outliers, looking for seasonal patterns, cyclical patterns, things like that. It's something that seems kind of obvious. When I compare the way that you go about, for example, a time series process of identifying, fitting and evaluating models, in my experience it wasn't something that I have really seen done.

Normally for a given valuation method it's almost like a default model that has some different assumptions you can put into it, that you're going to run on a data set and essentially evaluate real time. Once you get an answer, you say, "I don't think that answer makes a lot of sense because I know the weaknesses of this approach, and I know that there's a bit of information that it doesn't incorporate that's true this time." It's kind of weird, and, as you do that enough, if you accumulate enough experience, you almost have this set of residuals in your head of how well things have done in the past, including how you've manipulated answers after you didn't agree with the first one that came out. So, it's like a real-time or prospective way of building and evaluating models. Then to a historical period evaluate them on a holdout sample that they haven't seen, select one based on some set criteria, which does involve some judgment, and then use that to go forward. That's certainly not a novel or radical concept, but the more I've thought about it, the more I thought, at least in my experience, we don't do that with our IBNR methods very frequently. I haven't seen it done.

So that's really what this building summary is talking about. Once you've identified a target variable, go through the process of specifying and fitting these models, and that should be based on some historical data set. Then once you have a candidate list of possibilities, you're going to go through an evaluation period on a data set that they haven't seen before. In trying to figure out which model you like the best, you have to, again, go through that balance of goodness of fit and robustness. How well is it going to generalize? You can certainly look through, for a given model, the predicted contribution of any of its parameters to weed out some. There are many tests you can perform to actually help you do that. It can be overwhelming at times, but the key thing is to evaluate the residuals themselves—make sure that you're not doing something with the model that totally violates the assumptions that it was built on.

If you've specified a model and fit it to data, do the residuals that are produced meet all the criteria for the model itself? Is there any bit of unexplained trend or seasonality that I haven't taken into account that would suddenly violate the model assumptions? A couple of useful error measures, the routes mean squared error measure is probably my favorite for goodness of fit. Then there's Akaike Information Criteria (AIC) and Schwartz's Bayesian Criteria (SBC). It's a measure that's basically introduces a penalty for the number of terms in your model. It's a good one for figuring out, whether I have really gone overboard.

When you're building these models, even though it might seem like this is a statistical process, I have tests in place to help me make decisions, but there's still a lot of judgment involved. That was another thing that was a bit surprising for me in going through this. A key fit is: what set of data am I going to use for fit and evaluation? How big should it be? What's going to be included in it, things like that. Then any type of filtering or transformations you do to your data upfront, kicking out values that you deem to be outliers, things like that—there can be some real judgment involved there as well. Certainly once you get down to things, you're trying to pick a given model out of a candidate list, and you get a couple of models that have very similar goodness of fit, they're pretty close, they don't have too many parameters in them, how do you know which one to go with? It can be tough. Probably the biggest thing that I've run into that's difficult when you're building models is knowing when to quit. It can be very tempting to continue to build for a long time.

Method as Applied to Completion Ratios

- A separate model is built for each of lags 1 through n
- Each model forecasts future ratio values. The lag 1 model forecasts one value, the lag 2 model two values, etc.
- Lags >n are arithmetic averages (no stochastic treatment)
- The product of forecasted ratios at each lag is applied to I&P for each month (like normal)

Just as a quick example if you were going to try to build time series models for completion ratios: It's step-by-step. The last step of going through and taking a forecasted value from each of your time series models for each of the lags and going ahead and multiplying them to produce estimates is shown above. When I had initially done this, I had just projected completion ratio values forward with associated distributions around them and then sampled from those and specified correlation for the sampling to produce estimates. That got to be quite tedious, and I'm not sure that I was doing a good job of maintaining appropriate relationships between the sampling of the different lag values. So, instead, I decided to go ahead and multiply all these up and go back historically and make 60 estimates of what this model would have done if it were estimating incurred claims with only one payment. Then I'll do the same thing as if it had always had two payments and then three, and use the goodness-of-fit or sample variance measures from doing that exercise as a way to go forward and sample the forecasted values.

Here's a quick example (Chart 1) of some fits for trying to look at lag 1 completion ratios. This is the ratio of two payments to one payment. The blue line is historical. The red dots are fit. I have no particular fondness for this model. It's just an example of what you can do. Here you have a basic autoregressive model, a two-term autoregressive model. There has been transformation for integration. It's taking first differences of all of the series values to explain away some trend, then incorporating a couple of other different regressors that are uncorrelated with—or at least not functions of the value that we're interested in—the lag 1 completion factors, that give some indicators of changes in speed of completion, as well as an indicator variable for the number of payments that were made in any given month. I don't know if this is universally true, but that's a big deal for us—our payment system arbitrarily, depending on dates, can have more or fewer payments, which affects completion speed because of cut-off times.

Let's do the same thing with incurred PMPMs instead of actual completion ratios. Here there's a choice that you have to make, and this is where you're more of a practitioner than an academic. At some point you're going to have to determine where your series ends. What set of values for incurred PMPMs are you going to use to project forward? And, of course, the closer you get to more recent times, the less sure you are that you actually know what the actual is. That's the whole point. So, normally I've gone back 12 or 15 months, and I include values that have some estimate to them based on a traditional method, thinking that they can't be so far off. I use that as a baseline series and then project forward. Then it's just a matter of fitting a time series model to what we're deeming to be complete incurred PMPM values.

Here's an example of a fit (Chart 2): With an integrated autoregressive model, you could easily include variables, such as actual indicator variables for season terms or any other series that you want to include. If you are doing only medical claims as your incurred series that you're trying to estimate, you think that you have a good idea of drug claims and want to use that as something you think would be

correlated with medical claims for a given month; it's very straightforward to go ahead and introduce it as another regressor.

Some models I've had a lot of success with are good because they do a good job on goodness of fit, but you don't include a ton of terms, are either subset models or factored models where you only include terms for lags 1, 3 and 12, as opposed to 1-12, which can be an extremely large model, or a factored model where you take the products of separate models to get some interaction terms as well. Those seem to work pretty well, at least on the health insurance data that I have available to me.

So let's say that we had gone through a modeling process doing both ways. In other words, you're going to have a completion ratio view of the world. We're going to do an incurred PMPM view of the world. We've built time series models for each. I think for everybody here who has done valuation work, this is essentially the balancing act that you go through. You have some models that try to reflect payment speed and some models that try to reflect just the pure incurred process, and you need to figure out how you're going to blend these two answers in your head. Normally I think that the assumption is that the more recent it is, the less I'll believe models built off of payment speed—as I go further back, I'll give them more weight. A straightforward way to do that would be to go back historically, perform that weighting and see what would have been the best set of weights to use to maximize the contribution of each model.

In our example of using completion ratios and incurred PMPMs, let's say we're trying to estimate the most recent month of incurred claims. That month would only have one payment made against it so far, so we would combine our completion ratio models that only assume one payment being made. It's the product of a lot of factors. Let's say that when we built our incurred PMPM model, we went back 12 months and said, anything 12 months or prior, that's good enough. Then when we combined the models for evaluating, how do we combine models for the most recent month? That's a 12-step-ahead forecast on the incurred PMPM side in association with the estimate or completion ratio that a model makes when it only has one payment.

So we would need to go back historically into our data set and say, okay, now produce a series of 12-step-ahead forecasts using this model. What are the estimates from this model when you only have one payment? Then combine those estimates to minimize some error measure. It could be squared error or something like that. The key there is that you don't want to use the one-step-ahead forecast from the incurred PMPM models. You're making a 12-month projection and then an 11-month and a 10-month. So you need to do that going back all along.

Once you've done that, then the whole idea behind this is that by using time series models to project all these values, you can make distributional statements around these estimates, and you want a sample from these distributions that you made for

incurred claims. The big challenge, or at least the big challenge for me has been, let's say that I have distributions around the last 12 months of incurred claims. I'm going to sample from this 12 distributions. How do I correlate the sample? For a lot of models that's relatively straightforward. For more simplistic time series models, there is an analytical solution that's easy to derive. Once you start getting into combinations of models, and particularly even with taking the product of all these completion ratio models, I think it gets a little bit trickier, and I don't know that I have a great single way to figure it out for more complicated model combinations.

If you get really bogged down in how correlated any given month is to itself in this projection, you could just go ahead and say, "Well, I can at least develop some maximum. Maybe I can't figure it out exactly, but I can develop some maximum." I've done that a few times. The one problem with that is your covariant structure for sampling—it makes a really big difference, at least with all the datasets that I've looked at. The differences between no correlation and perfect correlation is a lot of money for a reserve. So, it's not something that you can just brush aside lightly. If you're going to specify a maximum, you might want to be relatively conservative.

Another idea that I've had on this, particularly, is with a completion ratio model, is there some type of filter you could apply upfront, like develop a simpler model that doesn't exactly give the same answers that the completion ratio models would, but would be pretty close, and if it's simpler, it's a nice, two-term autoregressive model. Then it's a lot easier to talk about the covariant structure of your forecasted estimates. It's definitely something to think about, though.

Briefly, I'll present some pros and cons. If we suddenly develop relatively reliable ranges around our reserve estimates, then does that give folks even more room to debate what is actually going to get booked? I don't know the answer to that. It's something I worry about, though. Who's to say 50th percentile? 70th percentile? Can it just be in the range?

Then, even for the current month it's an issue, but say that you decide in a given month to book the 80th percentile of a distribution that you feel good about, and then the next month when you go to do re-estimates, as part of your process you'll get a distribution around your re-estimated IBNR for the month that you booked the 80th percentile. Suddenly the number that you booked is now looking more like a 50th percentile. Do you make any adjustments? Then you go another month, and now it looks like that has a 75 percent chance of being low, but it's still within a range. When do you know how to make adjustments if you miss something—favorable or unfavorable? That's not something I have the answer to. I think it's more just a policy that's going to be established by a given company, but it also means that from company to company there could be a lot of variation. So that might be a bad thing.

One other quick thing, too: streamlining and automating this process can take substantial time. That's an understatement. It's taken quite a few months to figure

out how to make this something that's a little bit more of an industrial strength process that we could reasonably use on a monthly basis. Certainly the practical constraints are probably paramount, even though I haven't mentioned them, in developing any of these models. It can't take much longer than it does now. There's a very short amount of time available, so, speed and efficiency are important. One other thing that I'll mention related to that is there is a particular software burden as well. I don't think that it's feasible—I could be wrong, but I don't think it's feasible to attempt to develop these models in Microsoft Excel on your own where you're essentially recreating the wheel. You're going to have to buy, or at least have access to, some software that does it for you. There's plenty of stuff out there, but as soon as you introduce more software, then you have to figure out how to fit it into the rest of your existing software processes. So that's a challenge. I think it's possible, but it definitely takes a lot of work, and it's almost like all fixed costs as well. You're doing it all upfront. Once you get it in place things can go along reasonably smoothly, and there's not a lot of effort involved, but it's a substantial burden upfront. Are there any questions?

MR. FUHRER: You mentioned wanting to have normally distributed residuals, but you can't get that with completion ratios because by definition they can't go over 1. So I am wondering if you gave any thought to using transformed data for that.

MR. FEARRINGTON: Actually, I was talking about the residuals associated with the models. So just error terms, if you fit a model to a series of completion ratios and you look at the residuals associated with it.

MR. FUHRER: Yes, but they can't be normally distributed because it has to be skewed because they can't go over 1.

MR. FEARRINGTON: If we go back to the example, we're talking about the difference between the blue dots and the red dots.

MR. FUHRER: As long as you're there, you're at lag 1, which I assume is early. But if you're looking at completion ratios that are back four or five months, then they'll be very close to 1, and they can't get over it.

MR. FEARRINGTON: Yes, I see what you're saying. I guess the only answer that I have for that is in my own experience, going back even to lags 11 or 12, where 12 or 13 payments had been made for the residuals themselves, I have not had a problem developing models that produce residuals that are normally distributed.

FROM THE FLOOR: What do you have to do to your model or how does your model handle the kind of nonstatistical variations in data, for example, you shut down claim payments for two months because you've misinstalled a new system or something like that, that we all have to deal with?

MR. FEARRINGTON: So, basically, I think the question is how does this approach deal with reality? And the answer is in some cases it doesn't deal with it at all. It is definitely reliant upon a large enough historical dataset to be able to evaluate things, and it's something that you run into every month. It can be either a huge change in a payment process, or it could be a huge change in an incurral process, for example, in December it happened to us with flu. So, if it's something that has occurred and has absolutely not really manifested itself at all in the dataset, then this is not going to make any statements about it. You could use this approach perhaps to develop a baseline and then try and estimate separately the impact of whatever significant intervention is of interest. The problems with that are when you get down the road and you learn the right answer, were you wrong because of your baseline model or because of your estimate of this kind of one-time thing? You'll never really know.

If it's something that's been around for a few months—a good example with us was about a year ago we took all of our primary care physicians (PCPs) off of capitation and went to fee-for-service. So actually there's this huge step-up in our fee-for-service data beginning at a certain point, and after about six months the models that we had I felt were doing a pretty good job of taking into account—we just specified an intervention variable. Beginning at this time put in a little gap and estimate a parameter for it, but it took about six months before the forecast that came off of it made any sense at all. It's definitely a weakness. If your historical data is not at all representative of what's going on, then you're going to have problems. That's a good question.

MS. BECKI HALL: I have a question about the time series over the deterministic, and if you can expand upon that, as well as—without giving out any proprietary information for your company—have you looked at doing the time series analysis in comparison to some deterministic methods, looking with hindsight, and what have your results been?

MR. FEARRINGTON: Actually, all I meant is that the output of this is not only just a mean estimate but also a distribution around it. So you can make percentile statements or talk about adequacy of a given reserve, although I don't think you can make it with the same amount of confidence using deterministic methods. Normally the ranges that I see from deterministic methods are essentially generated by looking at a combination of different sets of scenarios and different sets of assumptions. What if this is going on? What would the answer look like? What if we used a different set of historical completions or percentages to average them together somehow? That's all I meant by that statement. As far as our more deterministic tried-and-true method, we've been doing this in parallel for about nine or 10 months now. So far, nothing particularly strange has come out of it. I will say that, for the most part, the estimates that come out of the time-series-based approach are not radically different—at least the mean estimates are not radically different from the more deterministic approaches. This could just be a function of the data of our processes. It doesn't necessarily mean it's a universal statement.

There have been some big differences on particularly small blocks of business, which I don't think is all that surprising. But for the most part, we haven't had a complete divergence every month, wondering, for example, where is this going? A good point related to that, as well, is that ultimately the test is going to be how well it comes up with IBNR estimates.

So, you could go back historically. That could be another test to do. And forget about goodness of fit and everything like that. Ask how well would this have done in the past? Since it is formulaic, it's easy to go back and replicate what the answers would have been or, easy is probably the wrong word, it could be done, as opposed to if you have method that relies a lot on actuarial judgment and interpretation, that's difficult to go back and replicate in all past time. So, I think that's definitely the last sanity check that you would want to do before implementing anything.

MR. ROWEN BELL: Doug, you mentioned at the end of your talk a percentile distribution for your IBNR estimate and the issues that may lead to in terms of which number do you book and certainly there's been some talk that the future of reserving and capital requirements would be to have a situation in which everyone would be doing models like this. There would be an indication that you would book a reserve number that might be in the 65th percentile, the 70th percentile, and that you would use the extreme right tail of your model to calculate what your capital requirement would be using conditional tail expectation-type of method. The lifelines have already done a lot of modeling down this path, gotten some people interested in this. From a medical standpoint I always shrug my shoulders and say, "Well, no one's really done this yet." I'm really glad to see you and people like you starting to do this.

My question is, with that as background, how confident do you feel about your tails? I mean when you do this, and you come up with a probability distribution around your IBNR estimate, do you think your 95th percentile is a 95th percentile? Do you think your 99th percentile is a 99th percentile? Or the fact that, as you said a minute ago, your model's only as good as the data that goes into it. Does that mean that we're really not at a point where we could rely on the far-right tail of this sort of approach yet?

MR. FEARRINGTON: I think that's a great question, and I would have to admit that I would be a little reluctant to say that a 95th percentile that comes out of this approach is definitely a 95th percentile. There's uncertainty around every single parameter estimate that's in your model. There's the danger that you've blown it—at least for me there is. And there's certainly the danger that the data you're using is not indicative of what's going to happen. The real question behind those is all these potential sources of error that are not inherent in the distribution. How do you quantify those? I don't have a good answer for that.

At a minimum, at least it gives you some progress toward identifying the elements of variability that you can't quantify. It's like, let's get the stuff off the table associated with process variance that we think we can capture and then identify the things that we don't know. If you're thinking in terms of a booked estimated and then perhaps in the explicit margin that's held on top of that, then maybe you've got some guidance on what that explicit margin ought to be and how it could vary from period to period. But as far as getting down the road of what should a policy be or is this going to be an end-all-be-all, I don't know and would be reluctant to say it is. Any others?

PANELIST: From the reinsurance models, the third approach to the modeling that we're going to do today—the question we really tried to address was what's the appropriate stop-loss level for a 25,000-member plan? This has been a question that I've struggled with for a really long time. Early in my career as an actuary we had someone come in from a stop-loss vendor, and since I was the low person on the totem pole with free time, I got assigned to talk to them. They asked, "What is the right stop-loss level for your organization?" I thought, "I really don't have a clue." So I went back and talked to my boss about it, and he said, "What do you think we buy the stop loss for?" I said, "Probably to cover the variability in earnings." And he said, "Okay, so how would you define what the right stop-loss level is that you'd want?" I thought for a minute and said, "How much variability of monthly earnings do we get, and I get, to keep my job?" That seemed to be a reasonable definition at that time.

Key considerations in trying to define the level include a balance somewhere between expected cost and variability—that decreasing the variability has a cost that you pay to the stop-loss carrier. An expected value approach would essentially conclude that purchasing stop loss doesn't make any sense because you're paying out to the stop-loss carrier, so with whatever stop-loss level you get, you reduce your expected value. But you need to somehow determine the level of variability. And then there's the question of how you define the variability. Is it variability from the mean? Is it variability from what you would have if you didn't have stop loss? There are a lot of different ways you could look at defining variability. But however you look at trying to get your arms around that question, you really come to the conclusion that stochastic modeling is a really good tool for this because, in essence, what we're trying to do here is understand *the distribution* of results, not just a fixed point on the results.

In constructing a model, we started with a number of assumptions. We came up with a claim distribution. We found that since we were really looking at variability, the claim distribution itself didn't matter as much as the fact that we had a representative claim distribution and tried to make sure we had something that really dealt with the tail pretty well and has a tail out there. We also looked at assumed stop-loss administration costs and profitability, so that we recognized the cost for the stop loss by assuming a loss ratio on the stop loss. In order to make the whole exercise work, we assumed that the stop loss is fairly priced—in other

words, that the expected costs in the stop loss are actually what the expected claim costs that go into the stop-loss premium are truly what the expected claim costs are. There are a number of people out there—some of which I've had as clients, that really look at stop loss—they will go out and take it to market every year. They'll say, well, this is what I think my costs are going to be, and they'll move their stop-loss level up and down based on where they think the stop-loss market is, and they view that as another source of a profit center for them. But in terms of this model, we really started with the assumption that the stop-loss coverage was fairly priced.

And then we wanted to look at several stop-loss alternatives—different levels of deductible in the program. From the model standpoint, just for convenience, we used an at-risk model in this exercise. At-risk is an Excel-based add-in that essentially helps you with some of the statistical packages. It's got advantages. It's all on PC. You do the basic spreadsheet in Excel, so it's easy to farm out and for everybody to understand. There are a number of models that are similar to at-risk. That's the one that I picked up and seemed to grow comfortable with and the one we used here. It also does a lot to help you capture the data outside of your model.

For the at-risk, on our 25,000-member HMO, we did 5,000 iterations. It took approximately 20 hours to run on my PC. That seems like a long time. It was over the weekend, so it wasn't much of an inconvenience. The 5,000 iterations on the 25,000 models, that's, I think, 125,000,000 samples. It's a pretty substantial number of calculations to go through in terms of plugging through to try to get to the model, and in the results we compared both the ultimate cost and the variability among our various stop-loss alternatives. So we looked at what the ultimate cost would be on a deterministic basis and on an expected value basis, and then we tried to look at different measures of variability for what I'm going to present here as far as results. We really looked at measures of variability that focused on what would be the cost at different percentiles in our distribution. Candidly, when we looked at the results of the model we looked at it graphically, but as Doug said, it's difficult to talk about modeling in a lecture format and we couldn't figure out a good way to present that in a mechanism that would convey the results effectively.

In terms of the results, the mean claim costs, including stop-loss premium for this whole packet, if we had no stop loss, we had 45.7 million. Then we looked at three basic stop-loss alternatives, 100 percent coverage over the stop-loss level. The three alternatives were 25,000; 50,000; and 100,000; and then the associated costs going up from there. If I recall correctly, I used a 70 percent assumed loss ratio for the stop-loss product, so that we said for every dollar of premium that we're paying to the stop-loss carriers, we're only expecting 70 cents in benefits, as you can see. So the 25,000—really that $3.5 million additional claims cost expectation or claims cost plus the reinsurance cost expectation is really the cost of the profit or administration that goes to the stop-loss carrier.

And then, as you can see, with the higher levels of stop loss you have a lower stop-loss premium and, thus, a smaller premium and profit that's going to the stop-loss carrier with the assumption of a level stop-loss, policy-loss ratio. Then if we look at the 70th percentile, that stop loss increases by about 800,000 in going to the 70th percentile of the claims cost plus stop loss. At 25,000 you only have half of that increase; the increased costs for the $25,000 stop loss only increases by 400,000 because a lot of that extra 400,000 gets in, in a change profitability for the stop-loss carriers, then the 50,000 and 100,000 drop down. We started at the 70th percentile because when we originally had some conversations with the client their initial thought was that was really where they wanted to look at variability, but, as we found at the 70th percentile, any of the coverages, with even up to a $100,000 stop loss, you're still paying more than you would pay under an expected value, even if you had what they defined as moderately bad experience, and even at the 100,000 you note that the no stop-loss option goes up by 800,000. The 100,000 only goes up by 600,000. So the 100,000 stop loss really only takes out 200,000 in additional to being well above the mean. It only takes out a fairly small amount of the variability at the 70th percentile.

One other caution I ought to throw in here is that one of the features in at-risk has some mechanisms—and a couple of the other statistical packages have some mechanisms—that will allow you to get to convergence faster, that at-risk includes a Latin hypercube sampling approach which essentially carves up your buckets, and when it draws the samples, it makes sure it gets them from certain buckets and it will allow you to get to overall convergence faster. The is that sampling mechanism will distort some of your tails, which, when you're doing this type of an exercise, what you're really looking for is the tail. So you need to watch what internal processing you do in your model.

At the 9th percentile on results we had a little bit more, the 100,000 and the no stop loss get to where they're very close in terms of overall cost, but even at the 95th percentile the 100,000 still became a little bit more expensive than the no stop loss, and clearly the 25,000 and 50,000 were up above that. Then we dropped it clear up to the 99th percentile, and finally we got to a level where the 100,000 stop loss program actually had a lower claims cost at the 99th percentile than the no stop loss, and the 50,000 began to get very close to the no stop loss there. A couple of other ways we wanted to look at variability, we looked at what the 95th percentile was to the mean, so that in the comparisons we were talking about before if you had stop loss or if you didn't have stop loss, what would be the difference in the expected cost? In this measure of variability we're really looking at how much does variability occur?

In other words, how much does the stop loss give you in terms of protection between the 95th percentile and the mean in terms of overall variability? For the no stop loss, the variability of the mean would be 2.5 million, and then if you use the 25,000, you really compress that variability from the mean a lot so that the mean claim cost at the $25,000 stop loss, the 95th percentile is only 1.1 million greater

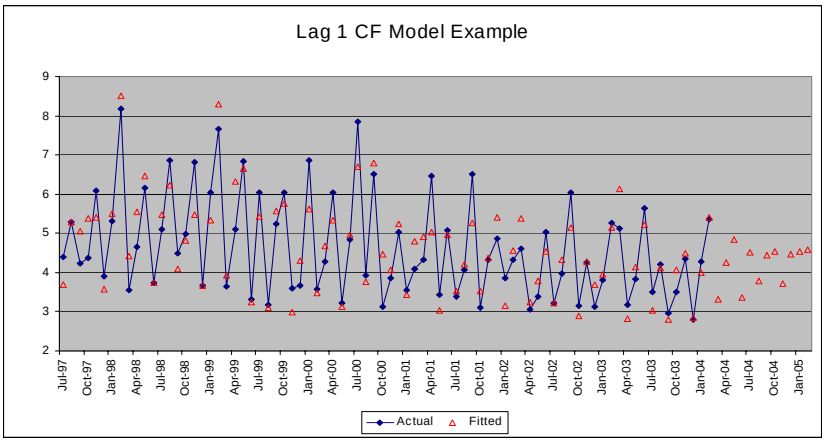
than the mean there, so even though it costs more in premium, it significantly reduces that variability, which may be attractive to some organizations in a specific situation. They may say it's easy for us to get the premiums in, but what we really can't deal with is any variability at all, in which case you may have a situation where that actually becomes attractive. That's why we wanted to look at that as another measure of variability here.

The other measure we put in here was the crossover percentile, and that was essentially at what percentile did the stop loss go, where the claims' costs were lower than it was under a no-stop-loss scenario, and the crossover percentiles for the three plans we looked at. For the 25,000 it was actually at the 99.9th percentile, then anything under that, having no stop loss, actually you have a lower expected claims cost. For the 50,000 stop loss it was 99.2, and for the 100,000 it was 96.2.

The conclusions that we got out of this model, aside from the obvious kind of answers on the stop loss and what it would look like, it seems like the model replaced a lot of facts with impressions, that the variability reduction was demonstrated to be very expensive in terms of profitability—you really gave up a lot in terms of profitability to reduce your variability. The other thing that we saw as a significant conclusion was that at a relatively small stop loss level it almost never crosses over, and, candidly, the 25,000 stop loss is probably too small, but we do see organizations that have those type of stop losses in there. In terms of some of the additional things we're looking at doing in the next step on expanding the model, we're trying to take a look at corridor stop loss where you have a deductible of stop-loss coverages, where you say that the company absorbs the first million dollars or two million dollars of claims over the stop-loss level in a way to try to eliminate some of the profit and administration going back up. Looking at what happens, if instead of the 100 percent you have a 90 percent that grades down, and then looking at modeling out with some experience refund contracts.

Chart 1

Example for a Lag 1 Completion Factor

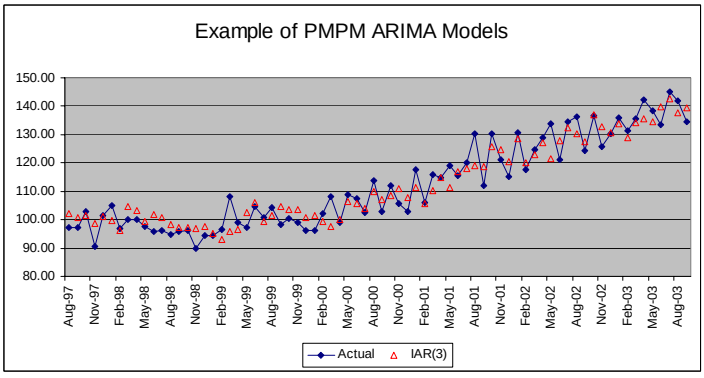


Fitted Model is : IAR(2) + α_1 * **B** Δ Paid Lag 1 + α_2 * **B** Lag 2 CF + # Payments Indicator

11

Chart 2

Example of PMPM Model



13