RECORD, Volume 29, No. 3*

Orlando Annual Meeting
October 26–29, 2003

Session 91OF
Risk-Based Capital Requirements on Variable Annuities with Guarantees

Track: Product Development and Financial Reporting

Moderator: DOMINIQUE LEBEL
Panelists: GEOFFREY HENRY HANCOCK
JEFFREY A. LEITZ

Summary: New risk-based capital (RBC) requirements for variable annuities (VAs) with guaranteed benefits such as guaranteed minimum death benefits (GMDBs) and guaranteed minimum income benefits (GMIBs) have been developed by the American Academy of Actuaries and are being considered by the National Association of Insurance Commissioners (NAIC). The proposed requirements may significantly increase the level of capital required on such products from today's typical levels. Moreover, the approach to determining the required capital involves stochastic modeling of guaranteed benefits. Industry experts discuss: overview of the new capital requirements; implementation issues such as model, validation, number of scenarios, grouping, interest rates and policyholder behavior anticipated impact and methods to reduce the necessary number of scenarios. Attendees learn the new requirements and how they affect the bottom line.

MR. DOMINIQUE LEBEL: I'm with the Tillinghast business of Towers Perrin in the Hartford office. I will be your moderator for this session. RBC for VA guarantees is quite a hot topic in the marketplace, which is evidenced by the large number of people in the audience today. We have a distinguished panel of experts to talk about this topic. We'll start off with Jeff Leitz from Tillinghast, who will provide an overview of the Academy's current proposal. He will also discuss the impact of the proposal and talk about some scenario-reduction techniques. Next will be Geoffrey Hancock from Mercer Oliver Wyman, who will look at practical issues through case studies.

Note: The chart(s) referred to in the text can be found at the end of the manuscript.

This is an open forum, so audience participation is welcome. Let me introduce Jeff Leitz. Jeff is a consultant in Tillinghast's financial services practice located in the Hartford office. His areas of focus include asset-liability management (ALM), equity risk management (ERM), mergers & acquisitions (M&A), embedded value (EV) and investment strategies. Jeff holds an honors Bachelor of Science degree in mathematics in actuarial science and is also a graduate of the SOA Wharton ALM and Risk Management program. He's a Fellow of the SOA, a member of the Academy and is currently a member of the RBC working group of the Academy's Life Capital Adequacy Subcommittee (LCAS).

MR. JEFF LEITZ: We have several topics to cover today. Your questions and comments are welcome.

I'll start with a brief background on VA RBC. Traditionally, the RBC formula was formula-based and designed as an early warning system. The RBC factors were essentially set at the 95th percentile of the risk distribution. They reflected industry-wide experience, rather than individual company experience. And they ignored extreme tail events.

The C3 component in particular was the most subjective, so near the turn of the last century, the NAIC decided that they needed to look at more specific company requirements. They first implemented C3 Phase I, which covered the interest rate risk for fixed annuities and single premium whole life. Fixed annuities included just about everything: structured settlements, guaranteed investment contracts (GICs), synthetic GICs, funding instruments, etc. It was the first standard to recommend scenario testing; the interest rate generator was provided as part of the requirement. And it did indeed, if you actually tested it, reflect your company's assets and liabilities. However, there was an exemption on implementation; two tests allowed you to not actually do the testing.

Last December, the Academy made its first draft recommendation on the Phase II portion, which was submitted for comment, and in September the revised version was up for comment and voted on at the December meeting this year. The equity risk and interest rate risk on VAs with GMDBs and guaranteed minimum living benefits (GMLBs), and in fact, all VAs is covered. The equity risk on group annuities with GMDBs and GMLBs is covered. Any insurance contracts with GMDBs are also covered. Excluded are equity index products, equity-indexed life insurance and equity-indexed annuities. Those are covered under a recently adopted proposal in June 2003, providing two alternative approaches: a look-through method and a tracking error method.

The proposed effective date of C3 Phase II is the end of 2004. With just a quick look at today's current formula, the components generally are C1, C3 and C4. The C1 covers default on general account assets backing any guaranteed-type reserves, just like they would for any default factors on assets. If you have surplus in your non-guaranteed separate account, that also receives a default factor. The

Commissioner's Annuity Reserve Valuation Method (CARVM) expense allowance, which is transferred from a separate account to the general account, is also subject to a default charge—effectively, the risk of not being able to amortize that expense allowance. That piece of C1 would no longer be needed if this proposal is passed, since it would be implicitly covered under the C3 Phase II modeling.

A C3 component is currently in place for GMLBs only, and the percentage range was from 1 percent to 3 percent of the reserves, depending on risk. This is the part that would be replaced under the new proposal, and would not only cover GMLBs, but also GMDBs. The C4 piece is five basis points on separate account liabilities and would remain unchanged by this proposal.

C3 Phase II is a stochastic methodology, whereby you would be required to create a model of your business and essentially project balance sheets at the end of each year, and look at year-by-year accumulated surplus or deficit positions. There are some nuances to this approach, so I want to provide some definitions.

The first definition is of model assets. These are the assets that you start the model with at the end of the year. These should be equal to your statutory reserves, and that means what you're holding in the general account for this product as well as the cash value in the separate account. So if the assets equal the statutory reserves, then you would think the statutory surplus is zero at the start date. That would be true, but this proposal redefines what we call model surplus. Model surplus is defined as model assets less cash surrender value (CSV) because CSV is used as a proxy for the reserves at every point in time along the projection. That decision was made to help speed things up because otherwise you may have to do stochastic-within-stochastic projections in order to compute the statutory reserve that in the near future may be based on a stochastic standard. With this definition, the starting surplus is generally positive.

The first step in getting toward the capital requirement is to compute the additional asset requirement (AAR). The AAR is the negative of the lowest present value of surplus at every year-end, including the starting surplus. Now, if you were to project balance sheets forward yearly, find the lowest surplus/deficit position on a present value basis and change the sign, that's the AAR for a given scenario of investment performance.

Next, the AAR is added to your model's initial assets; that constitutes the total asset requirement (TAR) for that scenario. This simulation is repeated. It's resimulated for a total of N times and N can be very large. We have some numbers in mind, maybe 10,000, but we have techniques to reduce the number to a more manageable level, perhaps 1,000.

After repeating this process, you end up with a distribution of the TAR. Based on that empirical risk distribution, you calculate the conditional tail expectation (CTE) at the 90th percentile level, CTE90. This is the arithmetic average of the 10 percent worst cases (the largest TAR values) from the risk distribution. Once you have that number, you subtract the statutory reserves that you started with—the actual statutory reserves held by the company. The incremental difference is the C3 Phase II RBC component. Note that this piece would be included with the C1-CS component for RBC covariance purposes only.

I have just a few comments on proposed scenario requirements. Unlike the Phase I, we're allowed to use just about any equity scenario generator and interest rate generator available. However, there are calibration restrictions imposed on them because of the desire to ensure that high impact, low probability events are captured. If you do have an integrated—and what I mean by integrated is a combined and correlated equity and interest rate generator—you can use that, or if you don't want to use your own generators, you can use a prepackaged set of scenarios that will be provided by the Academy.

The calibration percentiles are based on the S&P 500 equity returns over the last 50 years. That was quite a point of discussion, but it was decided that that made the most sense in our economic conditions today. Once we looked at the historical data, we used it to parameterize the regime-switching-lognormal (RSLN) model with two regimes. Many of you are familiar with that. Other multi-state-dependent models are permitted, as long as they are justified by the historical data and meet the equity calibration criteria. Also, the interest rate scenarios used, if any, should be as robust as they are in the Phase I. Chart 1 shows the RSLN.

You can see that the gold-colored distribution here is the one that we actually use. It has a negative skewness and a positive kurtosis, ensuring fat tails for the tail-dependency risks. This is where risk parameters collide, and tend to go south all at once. I believe those parameters are also provided.

The proposal asks that you calibrate your equity model, not only for the downside tail, but also for the upside, and the calibration requirements are represented by wealth accumulation factors for a $1 invested at time zero. For example, one calibration point would expect 10 percent of the time after 10 years to have $1.41 or less.

Some other features—VAs with GMLBs—must be stochastically tested under this proposal. Stochastic modeling is also strongly recommended for VAs that only have GMDBs; however a factor-based alternative method would also be available. The modeling should take into account credit for hedges and reinsurance and also their effectiveness. The discounting of the accumulated surplus at each point in time could use the integrated interest rates. They could use the implied forward from the year-end swap curve and they should be after tax. Both income tax adjusted as well as tax reserve deduction adjusted policyholder behavior needs to be reflected vis-à-vis the performance of the funds so that you get dynamic lapses, withdrawals, transfers, etc. You want to try to capture as much as you can of embedded options in the product.

We had some analysts at Tillinghast look at the capital impact of the new requirements, and on a generic product basis, here are the results. The impact can be fairly large. The largest impact appears under the GMIB, in some cases, 16-fold. But how do you evaluate the implications for your own company? It really depends on when you wrote the business. How much of the business is in the money? One of the best ways to evaluate it is to do the stochastic modeling early and find out the impact on your company.

In the process of doing the modeling, you'll also have a better understanding of the risk exposure and how to mitigate those risks. For example, asset hedging can reduce capital, but at this time there's no allowance for product hedging. I'd also like to mention a method for reducing the number of scenarios. It's not the only method, but it's a nice one.

As you know, the cash flows are the outcomes of the ALM model, and creating those path-dependent cash flows is time consuming, so it would help to employ some sort of scenario-reduction process. We have a method, which is basically a stratification-sampling approach, whereby a subset of the about 10,000 scenarios, for example, is picked. How you pick them is based on a relative present-value-distance formula. Essentially, you're looking for the extreme scenarios, those that are farthest apart or most unlike the others. Once you've selected an extreme subset, you then use the same distance formula to assign relative probabilistic weights to each by counting the number of scenarios most similar to each as an approximation to how much probability they would represent.

This tends to over-sample in the tails, which is a good thing, because that's where we're calculating CT90. Here's an example of the distance formula:

- *D* is a function of the two scenarios' risk parameters

  e.g.,

$$D = \sqrt{\sum_{t=1}^{30}\left(\prod_{k=1}^{t}\frac{1}{1+i_k} - \prod_{k=1}^{t}\frac{1}{1+i'_k}\right)^2}$$

  where *i*t, t = 1, 2, ... , 30, is an equity rate path consisting of one-year returns for 30 years

- *D* attempts to capture the relationship between scenario risk parameters and model outcomes by selecting extreme scenarios

The rates for i and i-prime are rates of return from two different scenarios. They're attempting to capture the relationship between scenario risk parameters and model outcomes. This is not usually one-to-one, but for GMDBs it's fairly close and it usually gives you satisfactory results. In an example where we have 1,000 scenarios which have been selected employing this metric, the tail is replicated quite nicely and produces a fairly accurate calculation of the CTE90.

So scenario reduction can be done successfully using the distance formula and an iterative recursive algorithm. I'd like to also acknowledge and thank Yvonne Chueh and Alastair Longley-Cook for their work in this area of research. The reduction algorithm is robust for guarantees that depend on volatility as well as return, and we hope that it helps you reduce the number of scenarios that you need to run.

MR. LEBEL: Geoffrey Hancock is a consultant and director of Mercer Oliver Wyman. He's an FSA and a Fellow of the Canadian Institute of Actuaries (FCIA). He holds an honors co-op bachelor of math degree with a double major in actuarial science and statistics from the University of Waterloo. Geoff has over 13 years of experience in the insurance industry. Currently, most of his time is devoted to evaluating the market and product risks on equity-linked, individual variable insurance and annuity contracts on a stochastic basis. For the past 15 months, he's been closely involved in the development of recommendations put forth by the LCAS regarding C3 Phase II RBC for VAs with guarantees.

MR. GEOFFREY HANCOCK: Now that Jeff has given you a good overview of the Academy's LCAS proposal, I'm going to assume that with that background, we can jump right into some of the practical aspects of implementing it and examine some of the challenges that you might face. So I'm going to demonstrate some of the techniques and some of the hurdles that you'll need to overcome, some of the issues you'll have to consider, and I'll do this primarily through case study.

As Jeff mentioned, the methodology is based on stochastic testing, looking at aggregate results. That's important. These are aggregate calculations, not calculations by policy. The assumptions are "prudent best estimate." Effectively this means that these are best-estimate-prospective-based assumptions with some margins in them, where those margins would relate to the underlying uncertainty in setting those future assumptions. So you would have to consider the quantity, relevance and credibility of experience data and unfolding trends. In areas where you are less certain about future behavior, you would add more conservatism to the assumption.

Also as Jeff mentioned, the metric being used here is the lowest present value of accumulated surplus. The risk measure is the CTE at the 90 percent confidence level, or CTE90. We're all familiar with that statistic by now. It's just a conditional expected value and the confidence level is $\propto = 0.9$. The method defines the AAR, and that's what I'm going to focus on here. All my results show the AAR above the modeled or working reserve, which is just the cash value. I'm focusing solely on that additional piece that comes out of the modeling.

The equity model that you use, the economic model, is not prescribed. However, there are some calibration criteria. I want to look at the impact of that and walk you through how you would take a model that would not meet the calibration and change the parameters so that it would.

Certification is required. The actuary must certify that the work meets the standards set forth in the proposal. It's not an opinion on the adequacy or the sufficiency of the company's current surplus or future financial condition. It's an important distinction. It's just a certification that the work has been done in material compliance with the proposal, which of course, when it comes to pass, will not be a proposal anymore!

Finally, there is strong recognition in the LCAS proposal that RBC is only part of the solvency solution here. So as complicated as all this modeling might seem, and as challenging as it may be, it's only part of the total solution. Sensitivity testing, hedging, and proper risk management are all elements that bear upon the company's financial position. So this is a means to an end, not an end unto itself.

Let's look at the calibration of the scenario models. This is a pretty significant element of the proposal and, of course, the scenarios drive everything else. The calibration criteria are specified as a table of wealth accumulation factors that you have to meet. It's important to realize though that different models are acceptable. There is no specific guidance on interest rate models. The focus of the calibration is on diversified U.S. equity, and we've used the S&P 500 total return index to represent that class of assets.

Table 1 shows the calibration points. The table is designed to narrow the range of practice and ensure that the return distribution has sufficiently fat tails. It is not meant to exclude models, and it's not meant to hand you the model on a silver platter either. It's really about ensuring that there's a sufficient sampling with respect to frequency and severity of those tail events. That's the sole purpose of it. The table looks at wealth accumulation factors. It was developed not from a direct analysis of the historic data, but by taking the historic data and fitting a model for stock returns. The data period includes monthly total returns for the S&P 500 from 1952 to 2002. The process of taking those monthly returns and fitting a model to them—in this case, the RSLN model with two regimes (RSLN2)—is well documented.

Table 1

S&P500 Total Return Wealth Factors at the Calibration Points

Calibration Point (α)	One Year	Five Year	Ten Year
0.5%	0.65	0.54	0.60
1.0%	0.69	0.62	0.72
2.5%	0.76	0.75	0.93
5.0%	0.83	0.87	1.13
10.0%	0.90	1.03	1.41
90.0%	1.34	2.67	5.55
95.0%	1.41	3.01	6.57
97.5%	1.47	3.31	7.55
99.0%	1.54	3.71	8.91
99.5%	1.59	4.00	10.00

The RSLN2 model fits the monthly data well on a statistical basis, and the calibration table is derived directly from that fitted model. Remember, it's not based on the data directly. For example, it's not based on a boot-strapping approach to the data or anything like that. It's simply based on the results of a fitted model. This model does produce fatter tails, and calibrating your model to the table will produce fatter tails than you may otherwise get. Now for the regime-switching model, it's—I don't like this word—a safe harbor, but it is only a safe harbor in the sense that the table was constructed using it. So if you choose that as your model, it should be a slam dunk—meaning that calibration is straightforward. Other models are a little bit more challenging. There is something in the paper that talks about the strictness about meeting the calibration. I think it's important to remember there is some judgment involved here, and I'll show the impact of this in a moment. The certifying actuary should be certain that he or she is materially satisfying the calibration table with the company's scenario model. It does not necessarily mean that you have to satisfy each and every point, which could be tricky with some models.

Burn that table (see above) into your memory because it's pretty important to keep in mind. The interpretation should be clear now. The significance level, also called percentile, quartile or alpha, is shown in the first column. So we have the left tail for alpha less than 50 percent and the right tail for alpha greater than 0.50. So this is a two-tailed calibration. We have three holding periods. The table shows the future values of a current dollar over various holding periods: one, five and 10 years hence. For example, with 10 percent likelihood (probability) over a one-year horizon, the model must show the future value of a $1 at 90 cents or less. Hereon, I will focus on the top line in the calibration (alpha = 0.5 percent), because it tends to be one of the stricter requirements—in particular, the one-year holding period constraint. It tends to be the hardest one to meet and as such drives the calibration of many common models, including the standard lognormal with constant mean and variance.

We've already mentioned that a monthly regime-switching model will meet the calibration by definition. The independent lognormal (ILN) model is a very simple model and widely used—loved by some and hated by others—and the foundation for a lot of option pricing and financial economics work. The assumption is that the log returns in any period are normally distributed with constant mean and constant variance. So, when you look at the monthly S&P 500 data over the calibration period, you have 600 values. Calculate the mean of standard deviation of those log returns and you have the maximum likelihood estimates. On an annualized basis, these are approximately μ = 10.5 percent and σ = 14.69 percent. Note: this is the mean and variance of the associated normal distribution, not the lognormal distribution.

The expected total return here is 12.27 percent annualized. That's an effective return. Here's the kicker though: these scenarios would not satisfy the calibration. So we'd have to adjust the model and we'll see how to do that. It doesn't mean you have to abandon the lognormal model, because it can be salvaged. It means that you have to adjust the parameters. Typically, you're going to have to increase the volatility quite substantially. A standard deviation of 14.69 percent isn't going to cut it. Remember, we usually call it volatility when we move out of the statistical world and start talking about returns.

What should we do? For the equations (below), you need to solve for mu (μ) and sigma (σ) such that the calibration table is materially satisfied for all points. Remember, each calibration point has two elements: a time horizon T and the quartile level alpha (α), also known as the significance level.

$$\Pr\{S_T \le q_T\} \ge \alpha \qquad \textit{for } \alpha \le 0.10$$
$$\Pr\{S_T \ge q_T\} \ge 1-\alpha \qquad \textit{for } \alpha \ge 0.90$$

"S sub T" (S_T) is the value of a dollar at time T. "Q sub T" (q_T) is the calibration point, and then alpha is the significance level. You can also re-express the accumulation factor as returns; so you can do it either way.

Now you have to work through the math. The natural log of S_T is normally distributed with annualized mean and variance of μ and σ^2 respectively; q_T is just a number, and alpha is just a number. Those are plucked from the table. The table has 30 points, so you go through the exercise of solving multiple times. Here's the equation you need to solve:

$$\frac{\ln q_T - \mu \cdot T}{\sigma \cdot \sqrt{T}} \geq \Phi^{-1}(\alpha)$$

That's the inverse of the standard normal CDF on the right-hand side. Now's there's a problem—can you see it? In solving this for a given alpha and time period T, we have two unknown parameters—mu and sigma—but only one equation. Hence, there are actually an infinite number of solutions. How do you narrow those down? It's hard to cope with an infinite number of solutions! We want one reasonable solution. There are a number of ways you can proceed, but one reasonable way is to fix the total expected return at some appropriate or reasonable level. Recall,

$$\ln\left(1+E[R]\right)=\mu+\frac{1}{2}\cdot\sigma^2$$

Here, E[R] is the total expected return as an annual effective rate. This doesn't have to be the historic value of 12.27 for the given data. It could be just some sort of reasonable expected total return, inclusive of dividend reinvestment, for a well diversified U.S. equity index. So maybe it's 12 percent; maybe it's 10 percent—whatever you think is appropriate for long term modeling.

Now that we have this assumption, we can re-express one of the parameters, either mu or sigma, in terms of the other variable. Now you're dealing with an equation in one unknown variable. That's easy to solve!

In my examples I've fixed the total expected return (*M*) at the historic value of 12.27 percent effective. I then solved for mu and sigma for each of the 30 calibration points using the previous relationships. I've plotted them in Chart 2. Mu is on the X-axis and volatility on the Y-axis. All values are annualized.

The uppermost point, the one that I'm calling full calibration, shows the ILN parameters to meet every single point in the calibration table. In fact, using these parameters you'd be quite a bit more conservative than the calibration table at the longer holding periods, and we'll see that in a moment.

The volatility is about 20.5 percent. That's a far cry from the 14.69 percent that the historic or maximum likelihood parameters give you. That's a 6 percent annualized increase! This result is being driven by the upper leftmost point in the calibration table. Looking back to Table 1, you'll see 0.65—meaning that there's a 0.5 percent chance that a dollar in equities will be worth less than 0.65 in one year's time. Note that this is the left tail, not the right tail.

The maximum likelihood parameters are shown by the caption "MLE." We saw those before: sigma is about 14.7 percent and mu is about 10.5 percent. These parameters would satisfy a few of the calibration points, but not the vast majority. That's why the maximum likelihood model would not materially satisfy the calibration table. So what does "materially satisfy" mean? Well, that's a personal decision of course, but in my modeling I used the parameters labeled "satisfies calibration on average." Volatility is about 18 percent and mu is pretty close to 9.5 percent annualized. These parameters don't meet every single calibration point, but I believe that they would materially satisfy the calibration requirements. So I'm going ahead with them and claiming that would be a reasonable assumption. Now you're looking at about a 3.5 percent increase in the annualized volatility to take the lognormal model that we know and understand and make it meet calibration.

I've graphed a few of the calibration points in Chart 3. I'm focusing on the left tail here, because the left tail really drives the calibration when you're working with the lognormal model. In fact, once you've calibrated the left tail, the right tail is already calibrated. Chart 3 deals with the one-year horizon. The Y-axis shows the left tail calibration point. The X-axis shows the alphas, 0.5 percent, 1 percent, and so on up to the 10 percent probability level.

The solid navy line connects the calibration points. To meet the left-tail requirements, we want the model statistics to be below that line. That's the challenge.

The line labeled "ILN (MLE)" shows results for the lognormal model using the maximum likelihood parameters. None of those points are below the line. This would not be an acceptable model. We already saw that.

The line for the regime-switching model is difficult to see, and the only reason it doesn't lie exactly on the calibration table line is because of sampling error. I used sampling here to demonstrate the impact. But of course, we know the regime-switching model satisfies the calibration table. The other two lines here are for the lognormal model with different parameters. The full calibration, where the model meets every single calibration point, is marked by the open squares. The model points for "ILN (CAL-AVG)" are solid squares. This is the model which I've assumed materially satisfies the calibration standard. I guess that's open for discussion, because as you can see, not all the points lie below the calibration line.

Now we're going to move to the five- and 10-year holding periods and see what they look like using the same models and parameters. See Charts 4 and 5.

How do the models now compare? The solid navy line still connects the calibration points and the regime-switching model sits on top of that. The ILN model with maximum likelihood parameters still doesn't meet the calibration. Then I have my two other lines.

What is the takeaway from all this? The full calibration lies well below the line. It's more conservative than you need to be, and the reason it's more conservative is that the 0.5 percent one-year calibration points are driving everything. But look at the brown line with the solid squares (the ILN (CAL-AVG) model). Remember, this is the one I'm saying would materially meet the calibration. It lies entirely below the calibration line for the longer holding periods.

Now you may look at all these lines and say, "Okay, great. Those look nice, but how do they actually affect results?" We'll see that in a moment, but before we continue let's reiterate a few points. When you're working with the ILN model with constant mean and variance, it's the low alpha, short horizon tail points that drive the process. In fact, it's one point, with alpha = 0.5 percent.

I would go out on a limb though and say that if you're going to work with the lognormal model for U.S. equity returns, you're going to have to use a volatility of at least 18 percent. For the given data series, you'd be hard pressed to justify anything lower.

Volatility affects mostly how things work in the short term. Of course, it doesn't affect long-term projections too. Using the calibration method described earlier, you actually have to adjust the mu as well. Mu drives how well the market performs over the longer term. I often refer to it as the "trend parameter." Since the calibration table includes short and long holding periods, both parameters need adjustment. That's why it's a little bit challenging to solve for the confluence of those two parameters that materially meet the calibration standard.

In the end, a calibrated lognormal model will have extremely fat tails, much fatter than you really need to meet the calibration. It incorporates more "conservatism" than is strictly necessary. What all this is really telling you is that you need more than a two parameter model to match the calibration table.

So what's the impact on the AAR? I'm going to look at a model for GMIB business here. It's a pretty well diversified block of business: a mix of attained ages, policy durations and so forth. The first example assumes that all variable account money is invested in diversified U.S. equity funds. I've used a fully integrated stochastic model for interest rates and equity returns. Remember, interest rates strongly impact the value of the GMIB at election. I've also assumed a dynamic model for annuitization (election of the GMIB) consistent with the value of the underlying option. That is, the GMIB election rate goes up when the guarantees become more "deeply in the money."

So let's have a look at the results shown in Chart 6. Rather than just give you the one point, the AAR at CT90—the proposed standard, Chart 6 shows different CTE levels so you can get a picture of the tail of the distribution.

The open circles are for the RSLN model. This model doesn't have any embedded conservatism relative to the calibration standard. The lognormal model with MLE parameters is way down at the bottom (open squares), producing the lowest numbers of all. We expected that, since the ILN (MLE) model isn't calibrated. On the other hand, the full calibration lognormal model, labeled ILN (CAL-FULL) and denoted by solid squares, gives the highest values. Even the ILN model with "average calibration parameters" produces a CTE90 result that is about 15 percent higher than the RSLN2 model.

The difference between any of these lines and the regime-switching line is possibly a measure of relative conservatism (or lack thereof). We know that the bottom line—the ILN (MLE) model—doesn't meet the calibration criteria, so actually we can't use it. The other two ILN models do meet the calibration, but there's a substantial difference between them. Since the Y-axis is the CTE for the AAR expressed as a percentage of account value, the differences between these lines are actually quite material.

I've highlighted the differences with the purple solid line, which is the ratio of a calibrated lognormal model result, using those average calibration parameters, to the results from the RSLN2 model. This ratio is expressed on the second Y-axis. At CTE90, the ratio is about 1.15, meaning that the AAR is 15 percent higher than strictly required by the calibration criteria. That concludes my comments on calibration.

Let's look at policy grouping, another issue that you're going to have to come to terms with. Seriatim valuation is perhaps preferred in the sense that you don't have to make any decisions about grouping, but for very practical reasons, many companies are going to be forced into grouping due to the potentially long run times required by stochastic testing. It really depends on your software.

I'm going to use a case study to demonstrate the impact of grouping. Here, I'm working with the diversified GMIB portfolio, but I've extended the inforce model to include a range of assets and policy sizes. The average market to guaranteed value (AV/GV) is about 0.9. In reality, many portfolios would be lower than that right now. As before, this model has all kinds of moving parts: interest rates, equity returns and dynamic behavior for GMIB election.

We're going to look at three models. The seriatim model (no grouping) with roughly 40,000 policies will be the benchmark. The first grouped model uses a naïve grouping technique and achieves 99 percent compression, compacting the portfolio down to 400 cells. I also built another model with roughly 2,000 cells using a much more intelligent algorithm—this will be referred to as the "smart grouping."

The grouping criteria are pretty obvious things that you would normally consider: attained age, policy duration, market-to-guaranteed ratio and so forth. However, the smart algorithm doesn't group every policy. It looks at the entire dataset, all 40,000 policies, and does not group outliers. So it calculates some statistics on the portfolio and chooses not to group the largest policies, the oldest lives and those that are deepest in the money. There is some judgment involved, but the so-called outliers are modeled seriatim. The technique assumes that outliers are going to have a large impact on results, which in fact they do.

The conclusion is that the groupings can be really effective. Actuaries have used cell-based models for many years, but you have to be cautious. Compared to grouped models for traditional products, more thought and testing are needed to produce a representative model. Unfortunately, unless you take some time to understand the impact of grouping, you might never know.

The results are shown in Chart 7. The benchmark is the solid line—the results from the seriatim model. At CTE90, the standard for the AAR, the ratio of the smartly grouped model to the seriatim model is about 0.99. The results are very close. The ratio of the CTE90 results for the naïve model, however, to the seriatim model is about 0.93—a 7 percent error. That's probably still pretty close, but you should do some analysis to recognize this and perhaps even adjust for it in your modeling. Still, all three models produce similar values. I'm showing not just the CTE90, but several CTE levels so you can get a picture of the whole tail of the distribution.

Next, let's look at a number of scenarios. By now, you can see the theme that's developing: we're looking at a number of practical issues that not only impact model results, but are really important considerations because they strongly affect run time. That's really the issue that we're trying to overcome here. It is always true to say that running a given business model over more scenarios should give you a truer picture, but there are practical constraints. To reduce run time, not only can you group your inforce policies into cells, but of course you can also reduce the number of scenarios.
I'm going to connect the results to sampling error because that's a very important

issue to understand when you start running fewer scenarios under Monte Carlo simulation. I'm going back to my case study here: the seriatim model for the diversified GMIB portfolio. We have a universe of 10,000 RSLN scenarios as the standard. That is, we assume that modeling over the 10,000 scenarios gives the "right answer." Then we'll look at two approaches.

First, I'm going to go into those 10,000 and randomly sample 1,000 scenarios. The second technique also picks out 1,000 scenarios, but does so more strategically and intelligently. There are different ways of doing this. Jeff showed you a distance measure in his presentation that combined with a pivoting strategy to produce a representative set of non-equally weighted scenarios. I will use a special case of that distance measure, called *S*, that is frankly easier to calculate. We will see how well it works.

The second approach is a form of stratified sampling, a very common statistical technique for reducing sampling error. You first need to calculate the significant measure *S* for every scenario using the definition below. Notably, it is solely a function of the scenario, not the results from cashflow modeling. Sort the resulting values, and then stratify into the desired number of samples. Finally, pick the mid-point of each stratum as the representative scenario.

$$S = \sqrt{\sum_{t=1}^{H} \left(\prod_{k=1}^{t} \frac{1}{AF_k} \right)^2}$$

Inside the square root we have the sum of squared discounted values, where AF_k represents the accumulation factor for period k along the given scenario. You have to decide what horizon, H, to use. Do you use 30 years? 10 years? 15 years? I used H = 60. Since I'm using a quarterly time-step in the model, that equates to a 15-year horizon. Finding the optimal value of *H* would depend on the underlying business, but be careful to choose *H* large enough so that a majority of the cashflows occur within the horizon. Note that the modeling will still project out the business for the entire length of the scenario set. In my testing, I've consistently used a 30-year forecast period.

Before we look at sampling error and the impact of stratification, first let's consider estimator bias. The good news is that the CTE is an unbiased estimator, or more precisely, it's asymptotically unbiased. This means that the more scenarios you run, the closer you get to the true value and furthermore, the expected error in the estimator approaches zero. That's a good thing. There are all kinds of estimators that are not unbiased or even asymptotically unbiased, and they can become problematic. However, even though the bias is small, there can be a lot of measurement error when running fewer scenarios. We call this "sampling error." What we will see is that sampling error can be very significant when you're running fewer than 1,000 scenarios unless you take great care in selecting the sample.

From the modeling, the sampling error at CT90 with 1,000 scenarios is roughly 1 percent of the account value. That can be significant. So you may be thinking "What does sampling error mean? Why do I care about it?" Sampling error is really your potential mis-estimation of the quantity you truly want. It's noise in the process due to sampling, and it can be positive or negative and change through time. It also varies with the composition and risk profile of the business. We'll see at the end of my presentation that the AAR can be a pretty volatile quantity, and part of that volatility is from the method itself and part of it is from sampling error.

Chart 8 illustrates the impact of stratification for the various sample sizes shown along the X-axis. However, before we examine it, let's jump to the conclusion: stratification seems to work pretty well. Jeff concluded that as too, using his distance measure, and he showed you an example. I'm using a simpler measure and I'm coming up with the same conclusion. However, there can be problems. The problem is that this particular form of stratification depends on a relative distance or significance measure that looks only at the scenarios (inputs) and not the model results (outputs). So the measures used for stratification assume there is an extremely high correlation between the scenarios and their impact on potentially hundreds of thousands of policies. That correlation usually is quite high, but not always. Sometimes you can't tell what the impact of a scenario is going to be without actually running the model.

As a word of advice, I would encourage companies who are considering running models, and those getting more familiar with stochastic techniques, to recognize sampling error, measure it and then do something about it.

There are a number of ways to control sampling error. One approach is to use some form of variance reduction technique—you've probably heard that mentioned before. There's also the use of a control variate. The control variate approach is very practical and intuitive, and an easy way to improve results without a lot of fuss. You might want to call it a "reference portfolio" instead of "control variate," but I'll describe what that means in a moment. The key point for now is that the control variate technique uses model outputs, not only the scenarios, to improve the quality of the estimator. Until then, let's flip over to Chart 8.

The X-axis shows the number of scenarios, from 100 scenarios all the way up to 10,000. That's the universe—10,000. The left Y-axis is the CTE90 for the AAR, expressed as a percentage of account value. The solid points show the CTE90 for the two sampling techniques: random sampling (squares) and stratified sampling using the significant measure to pick scenarios (circles). However, perhaps the best quantities to look at here are the lines connecting the open squares and the open circles; these are the ratios to the actual result, expressed on the right Y-axis. The actual (true) result is the one based on 10,000 scenarios. So that's why these lines converge to one.

At 100 scenarios and pure random sampling, you could get a result that is 76 percent of the true result. That's a lot of measurement error—being off the mark by 24 percent. On the other hand, the stratification approach, where we're a little more intelligent about how we select scenarios, does a much better job. It actually overestimates the true value by about 7 percent. Again, these lines do converge to one as you run more and more scenarios. The important point to recognize here is that there's going to be a threshold for the number of scenarios you can reasonably accommodate in your models. So, run as many scenarios as you can. Whatever that limiting number, try to do the best you can to reduce sampling error. If the number is 100, you have a fair bit of work to do, and you'll have to be quite clever. If the number is 1,000, you can get really great results. But even 1,000 scenarios might be too challenging. So you're going to have to consider some of these techniques: stratification, variance reduction and use of a control variate.

Charts 9 and 10 offer further insight into sampling error by looking at the variance of the CTE estimator for various sample sizes. Chart 9 runs the model over randomly selected scenarios, while Chart 10 shows the results for the stratification method using the significance measure.

The number of scenarios is shown on the X-axis and the CTE90 AAR on the Y-axis. The circles show the average CTE when the model is run 10,000 times using the specified number of scenarios. Therefore, for a given sample size (number of scenarios), we have an empirical distribution for the CTE90 estimator (a sample of size 10,000). From these results I've calculated the sampling error, the standard error of the CTE90 estimator. Then, I've shown a 90 percent confidence interval around the expected value. That is, we're 90 percent sure that the true value lies between the upper and lower bounds. Unfortunately, the band is extremely wide when the number of scenarios is less than 1,000. It can be incredibly wide. At 100 scenarios, it's over 5 percent of assets. You're not very certain about your result when the confidence interval is that wide. Surprise, surprise: when you run more scenarios, the confidence interval narrows, but not as quickly as one might hope.

Chart 9 shows the random sampling approach. Chart 10 is for the significance measure stratification technique. We still see the same funnel shape—narrowing as the number of scenarios increases—but is it any better? The answer is "no, not really." Relative to random sampling, this simple stratification method doesn't appreciably reduce sampling error. It *does* get you to a result more quickly (that is, there is less bias in the estimator), but you need to consider other techniques for better variance reduction.

I'm not going to get into a lot of detail here, but there are ways to understand sampling error for this risk measure, the CTE. I presented a paper on this subject with a colleague John Manistre at the Stochastic Modeling Symposium last month in Toronto. The title is "Variance of the CTE Estimator" and it's available on the Canadian Institute of Actuaries (CIA) Web site.

The key result is the following for the variance of the CTE estimator:

$$VAR\left(C\hat{T}E\right) \approx \frac{VAR\left(x_{(1)}, \ldots, x_{(k)}\right) + \alpha \cdot \left(C\hat{T}E - x_{(k)}\right)^2}{k}$$

When you look at this quantity, it makes some intuitive sense. It says that the variance of the estimator has two components. The first component is the variance of the order statistics—those values that are used to estimate the CTE. However, CTE(α) is defined as the conditional expectation above the α-quantile. When the underlying distribution is unknown, the α-quantile is also unknown and must be estimated from the limited sample data available. This also affects the uncertainty in the estimator and is represented by the second term in the formula.

MR. WINSTON WISEHART: What is "K" in this equation?

MR. HANCOCK: "K" is the number of elements in the averaging to obtain the CTE. For example, if you have a 1,000 scenario sample and you want CT90, "K" would be 100.

You can read our paper. Therein, we talk about biased sampling techniques and they work quite well. However, they are more complicated.

Now, I'm going to show you something simple that can often dramatically improve the quality of statistical estimators. Earlier, I referred to this as the "control variate" or "reference portfolio" approach.

The essence of this technique is to compare the statistics from your model to known or highly certain quantities. These "known quantities" would have little or no sampling error. For example, if you can calculate a quantity in closed form, you remove sampling error from the "reference" or "control" variate.

In the real world, with thousands of policyholders owning rather complicated products, we can't do any of this in closed form. That's why we're running simulation models in the first place. So the challenge here now is to set up what I call a reference portfolio and use that as your standard for comparison.

Here is the method in five simple steps:

1. Design a "control" or "reference" portfolio whose statistics (for the metric of interest) are highly correlated to those for the actual portfolio. That is, the "reference" should be a close proxy for the actual portfolio in terms of how it reacts to the scenarios used for simulation.

2. Run the control portfolio and calculate the desired statistic (e.g., CTE90 for the present value of lowest accumulated surplus) to the highest possible accuracy. This could involve simulation over a large number of scenarios. Use "CA" to denote the value of this statistic.

3. Run the control portfolio over the valuation scenarios (those that will be used to test the actual model) and calculate the desired statistic. Use "CS" to denote this value.

4. Process the actual portfolio of interest over the valuation scenarios. Calculate the desired statistic. Let "FS" denote this value.

5. Use the quantities CA and CS to improve the estimate of FS.

In practice, the reference portfolio is usually a handful of policies that are representative of the actual business to be tested. We run the reference portfolio over a large number of scenarios—perhaps 10,000 or more—and calculate the AAR at CTE90. This is the value "CA." Then run the reference portfolio over the valuation scenarios, perhaps as few as 500 or 1,000 and calculate the AAR. Call this "CS." Now, run the actual portfolio over the valuation scenarios and calculate the AAR. This is the value "FS." Finally, estimate the AAR as FS × (CA ÷ CS). It's really pretty simple.

There's also another way to do this that I feel compelled to mention. I'm sure some companies have even attempted to do this. You can run your reference portfolio, or even your full portfolio, over a number of scenarios (as many as you can manage) and get an empirical distribution. Then, fit the empirical results to a probability distribution by estimating the parameters from the sample. You may even decide to fit only the tail of the distribution if that is where your interest lies.

For example, you may decide to fit the tail values to a translated gamma or Pareto distribution. Those are nice distributions for fitting these types of results. Estimate the parameters through some statistical technique—the method of moments or maximum likelihood—and then calculate all your quantities analytically based on that fitted probability distribution. As actuaries, we don't do this very much. We tend to just stop when we construct the empirical distribution based on sample results. You can go that next step and gain a lot of insight from it.

Here's the control variate approach in action (Chart 11). I ran the model to obtain results for all 10,000 scenarios for my control. The AAR at CTE90 is 11.02 percent of assets. That's what I'm saying is the true answer. When CTE90 is calculated over the 500 "subset scenarios," we obtain 10.42 percent of account value. We can use the ratio of these two quantities to improve the CTE90 estimate for the actual model, what I call the "full portfolio" run over the "subset scenarios." The "subset" scenarios are merely 500 paths selected for valuation using the stratification technique discussed earlier (i.e., where the significance measure is used to stratify the sample).

Running the full (actual) portfolio over the 500 valuation scenarios, we obtain CTE90 = 9.76 percent of assets. We then adjust this result based on the two measures for the reference portfolio. It's simple, but often very effective. Here, the final estimate would be 9.76 percent × (11.02 percent ÷ 10.42 percent) = 10.33 percent. To see if this worked, I ran the actual model over all 10,000 scenarios to obtain the "true" CTE90 of 10.41 percent of assets. It worked! 10.33 percent is closer to the true value than the original estimate of 9.76 percent. Unfortunately, it might not always work this well, but it's worth exploring. The really nice thing about this technique is that it doesn't require any knowledge of fancy statistics.

I have a few more things I'd like to go through before I finish. The next on the list is the impact of dynamics. The LCAS proposal does talk about the need to consider dynamic behavior for lapses and option election (e.g., GMIB annuitization).

In this example, we'll look at a GMIB portfolio. The key behavioral component is election of the guaranteed benefit; that is, the decision to annuitize at guaranteed rates. I still have my reference model, which is the seriatim portfolio with stochastic interest rates and equity returns and full dynamic behavior for GMIB election. I'm going to compare it to two other models. The first does not have stochastic interest rates. It has fixed rates equal to the guaranteed interest rate (3.5 percent). The second model assumes non-dynamic annuitization rates fixed at 5 percent per year, regardless of asset performance or the underlying value of the GMIB.

Recall, the base model dynamically simulates election of the GMIB as a function of the "in-the-moneyness" of the guaranteed benefit. It's an anti-selective model whereby election rates increase as the guaranteed benefit becomes more valuable.

Let's look at the results in Chart 12 and try to draw some conclusions. This shouldn't be a surprise to anyone—the behavioral dynamics have a huge impact on the results, absolutely massive. As with previous slides, I've shown results at various confidence levels, not merely the AAR at CTE90. The line with the open circles shows the ratio of the model without behavioral dynamics (i.e., flat 5 percent election rate) to the fully stochastic/dynamic reference model. At CTE90, the ratio is 0.57, so the difference between those two quantities is over 40 percent! You obviously can choose different models, but the point here is that dynamics have a huge impact.

So the takeaway is that you need to build reasonable dynamics into your models. Don't take this task lightly. Come up with something that's reasonable and sensible. I've seen some companies go overboard in the dynamics, and come up with very extreme behaviors. Behavioral modeling is difficult at the best of times; with an absence of experience data, you need to take great care. I'm just trying to demonstrate that that impact is potentially quite large.

FROM THE FLOOR: Was there a maximum annuitization rate?

MR. HANCOCK: Yes. This model allowed the annuitization rate to go up to 20 percent per year.

Now, when I compare the stochastic interest rate model to the model with fixed interest rates, the differences are a lot smaller, but still material. At CTE90, the fixed interest rate model "understates" the true AAR (as measured by the fully dynamic/stochastic model) by about 10.5 percent. However, this difference is very much a function of the current interest rate environment, and the fact that I chose fixed yields to be equal to the guaranteed rate. If, for example, we turned the clock back to 1990, with interest rates at 9 percent and used that as the fixed interest rate, you would see very different results. You'd probably see very little value under the GMIB. So, the conclusion is that you don't have to run a stochastic interest rate model, provided that you are very careful on how you select the interest rates for projection. You have to pick something that's related to the current environment, but suitably conservative and sensible for a long-term projection because it has a big impact on results. For example, grading from current rates to 3.5 percent over five years would not be unreasonable.

Time horizon is another issue I'd like to address. The LCAS proposal talks about building and calibrating a model, running out the policies, calculating the accumulated value of surplus for each scenario, etc., and all that takes time and effort. We talked about reducing that time and effort by decreasing the number of scenarios, grouping policies and so forth. Well, the projection time horizon is important too. You don't have to run things out for 60 years, maybe 30 is good enough, maybe even 15 is acceptable. Here are the CTE90 AAR results for the seriatim GMIB model with stochastic interest rates and full behavioral dynamics.

With regard to time horizon, Chart 13 gives you a picture of how the AAR changes as you alter the horizon in your projection model. I've gone out 30 years.

The magenta line (data points are represented as solid squares) is the actual quantity expressed as a percentage of account value—that is, the CTE90 as a percentage of starting account value. The right Y-axis here is for the other line—the blue line (data points are marked with open diamonds). This is the ratio of the results for a given projection horizon relative to the results for the full 30-year projection. Hence, the ratio converges to one. So how do you read this table? For a 15-year horizon, you would end up with a result that is about 90 percent of the result for a 30-year horizon.

That's actually a pretty nice thing to see, meaning that you don't always have to run a 30-year model. You might be able to get away with running a 15-year model, which could potentially run much faster than a longer projection. Maybe even a 10-year horizon is acceptable, provide you adjust the results. Of course, you can't go too low. My suggestion to you is to do some of this analysis on your own, not for the whole portfolio perhaps, just for a representative subset. Try to understand this relationship. Then you can use this relationship to modify the end result. For example, if I decided on a 15-year horizon, at the end I could just divide the modeled AAR by 0.9 to gross it up. I've saved myself a lot of time and effort, but still managed to obtain a very good result.

There are a whole host of other important implementation issues, but I'm not going to go into great detail. Here are the highlights:

The random number generator is critical. We're talking about scenario analysis here—stochastic simulations. The backbone of that is the random number generator. Don't let it be the weak link in the chain of events, because if it is, it could potentially invalidate all the hard work and analysis that comes thereafter. So take some care in choosing the random number generator. There are many great reference materials out there on this subject.

Proxy funds. As a practical matter, you will need to map all your investment funds, potentially dozens, to a smaller number that you can manage in the model. Maybe six or seven different asset classes will suffice: bonds, diversified equity, international equity and so on. This is a very important decision, and I know a lot of companies are familiar with the process. They might use regression analysis by taking a look at actual fund returns versus index or benchmark returns as a way of understanding how the actual funds would map to proxies or market indices. It's a fairly straightforward technique, but involves as much art as science. I would suggest that you avoid the temptation to be unduly optimistic. And by that I mean: avoid the temptation of giving too much credibility to the good historic data for your portfolio. Fund managers can change, styles can change; don't overestimate the positive benefits of active fund management, etc. Take care to use volatility as a guide, not necessarily the returns themselves.

Market correlations are extremely important. Running a model that either assumes all asset classes are independent (uncorrelated) or 100 percent correlated is, in my opinion, unacceptable. The 100 percent correlation is extremely conservative; the no correlation approach is almost ludicrous. How could you assume that there is no correlation between asset classes? But the correlations are probably not one! Correlations estimated from historic data can have high standard errors. This leads to parameter uncertainty. As a consequence, you should err of the side of conservatism.

Finally, the proposal talks about fixed expenses in the projections. This not only covers per-policy expenses, but all allocated costs to this line of business, including overhead.

My final topic is the volatility of the AAR. As before, I'm going to introduce the subject and provide some demonstrations through a case study.

Here, I've calculated the AAR at CTE90 for a portfolio of business, complying with the standards for calibration. I've compiled the model results at each quarter-end over the historic period December 1995 to December 2002 inclusive. The underlying portfolio is comprised of a 5 percent roll-up GMDB product with the underlying accounts fully invested in the S&P 500. The population is stationary with an average age of 65. So I'm assuming as lives exit through death or lapse, new lives come in to maintain the demographics. I have also assumed a deposit growth rate of roughly 5 percent per annum. Money is flowing in, so this will introduce a measure of dollar-cost averaging in terms of the market highs and lows. Overall, I've tried to make this example as realistic as possible.

Chart 14 gives you a hint of what this RBC standard will look like when applied over time, using the last seven years as an example. As you can see, the AAR is pretty sensitive to small changes in market to guaranteed ratio. The solid line on this chart shows the aggregate ratio of account value (market value) to the underlying GMDBs. Use the right Y-axis to read these values. The other two lines show the AAR at CTE90.

So, the AAR is relatively volatile. As the ratio of the market value of the assets to the guaranteed value changes, so does the AAR, and sometimes quite significantly. They're highly negatively correlated.

However, time diversification is real and it's significant. What do I mean by time diversification? I mean the fact that new money is coming in over different market cycles. This is the dollar-cost-averaging effect. The AAR volatility would be considerably higher for a single deposit in December 1995. About the worst thing you can do is sell a ton of business in one year and then close the books. It's better to actually take on more business, maybe at a slower pace, but do it through time, because the diversification element is quite significant. Those new deposits coming in at the market lows have a big impact on the aggregate exposure.

Let me go back to Chart 14. The market-to-guaranteed ratio is one at the end of 1995, just for reference. The account value is equal to the guaranteed value. The guaranteed value starts rolling up at 5 percent a year. Interestingly, at June 2002, the account-value-to-guaranteed-value ratio is about the same. It's almost one again. Now you think to yourself, "the market fell over this time, so how can that be?" Yes it did fall, but you had money coming in over the whole period. However, even though the market-value-to-guaranteed-value ratios are about the same at these two dates, the AAR is very different. The value at June 2002 is about half of what it was at the start, expressed as a percentage of assets. In dollar terms it's larger, but it's expressed as a percentage of the underlying market value. This again demonstrates the power of dollar-cost averaging to reduce exposure. The fact that you have money flowing in through time, especially during market downturns, results in significant relative diversification.

For reference, Chart 14 shows two types of guarantee adjustments on partial withdrawal: dollar-for-dollar and pro rata by market value. You can see they're moving in lock step. The "dollar-for-dollar" privilege is higher, but they move the same way. Remember, the "dollar-for-dollar" provision is more costly when markets fall because for a given withdrawal, the relative adjustment to the account value is larger than the decrease in the GMDB. This increases the persisting exposure. While this example is somewhat exaggerated because it assumes all money is invested in a single market (the S&P 500), it does demonstrate the potential RBC volatility if the underlying exposure is unhedged.

MR. WISEHART: I'm from Swiss Re. I'm doing my professional development paper on this topic and I want to thank you guys for writing it for me, but I'm a little unhappy that I have to do some non-trivial calculations to get there. On that last slide, Geoff, you showed the AAR going down when the market was going up, and going up when the market was going down. If I'm trying to explain the amount of dollars to senior management that they have to put up, and I tell them, well, the market just dropped in half and I have to put up more in AAR, now that the market is down, than I did when it was up. They're going to say that doesn't make any sense. You guys must be actuaries, why don't you go back and do your models again? I'm wondering what your comment is to that?

MR. HANCOCK: There's a very simple answer to that. All my stochastic simulations used what's called a Markov model; that is, a model that satisfies the Markov property. As implemented, the regime-switching model satisfies the Markov property, as does the lognormal model. Simply, the Markov principle states that only the current state of the environment affects the evolution of the process in the next period, not the past. In short, the models are state-dependent, but not path-dependent. So the fact that you say that you're at a market low, or the market's gone down, and my requirement keeps going up, it's because those future prospective simulations don't care what's happened in the past. Indeed, there will be scenarios that project significant future downturns regardless of recent actual events. If you use a model, however, that does reflect what's happened in the past, it will tend to give you the more intuitive result that you're looking for. For example, mean-reverting models are typically non-Markov. However, there's a potential problem with non-Markov models: it's hard to adjust the parameters to satisfy the longer-horizon requirements of the calibration table.

MR. JAMES MALIN: My question is for Geoff. You mentioned a couple of times that the RSLN model would automatically fit the calibration. But I think the model also has to take into account the transition probabilities, because at the extreme, you could have zero transition probability to the second regime and that would be a lognormal model. Are the transition probabilities set in the calibration paper or is there some guidance on that?

MR. HANCOCK: Well, the transition probabilities are not set. The ones defined in the paper are the maximum likelihood parameters. So we took the historic monthly S&P 500 returns and fit them to the six-parameter regime-switching model. Mary

Hardy has well described this technique in her excellent paper, "A Regime-Switching Model for Long-Term Stock Returns" that appeared in the April 2001 *North American Actuarial Journal*. It's just maximum likelihood estimation. You come up with those six parameters that best fit the observed data. Those are the parameters that were used in designing the calibration table. So all I've suggested is, if you fit your regime-switching model using maximum likelihood parameters to the same data set, you'll get the same parameters and, by definition, they'll fit the calibration. If, however, you choose to fit your regime-switching model to different historic data, you will get different parameters that may not meet the calibration.

MS. MARY HARDY: Thank you for your compliment. I'm not going to talk about the regime-switching model. I was going to ask a quick question on the stratified sampling methods that you both discussed. I just wonder whether these methods remove some of the path dependencies and if the methods will work for a path-dependent product like a ratchet as well as it does for a product that's less dependent on the path, such as a straight GMDB.

MR. LEITZ: That's a great point Mary, and that is true in fact.

MR. HANCOCK: Yes, Mary is quite right. Stratification techniques that use relative distance measures based only on the scenarios, and not their impact on results, can be a challenge to implement. You have to be careful or the sampling won't have the desired effect. In some cases, you really should run some sample policies, maybe a control or reference portfolio to understand how those scenarios actually affect results and doing that is much more important when you have path-dependent options. In my opinion, the most practical and effective stratification techniques rely on a *combined* analysis of the scenario inputs and sample output.

Chart 1

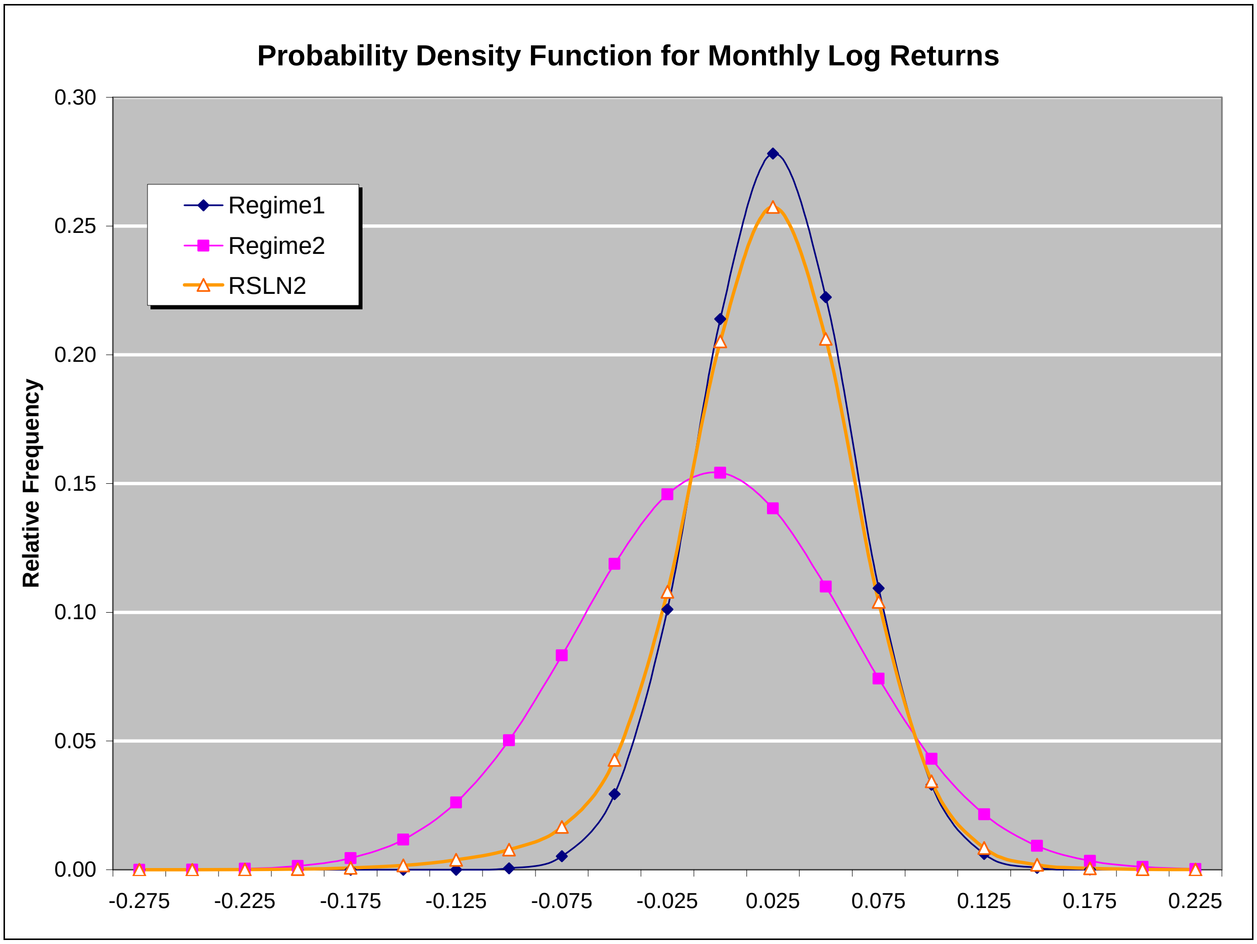

Probability Density Function for Monthly Log Returns
Relative Frequency
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Regime1
Regime2
RSLN2
-0.275
-0.225
-0.175
-0.125
-0.075
-0.025
0.025
0.075
0.125
0.175
0.225

Chart 2

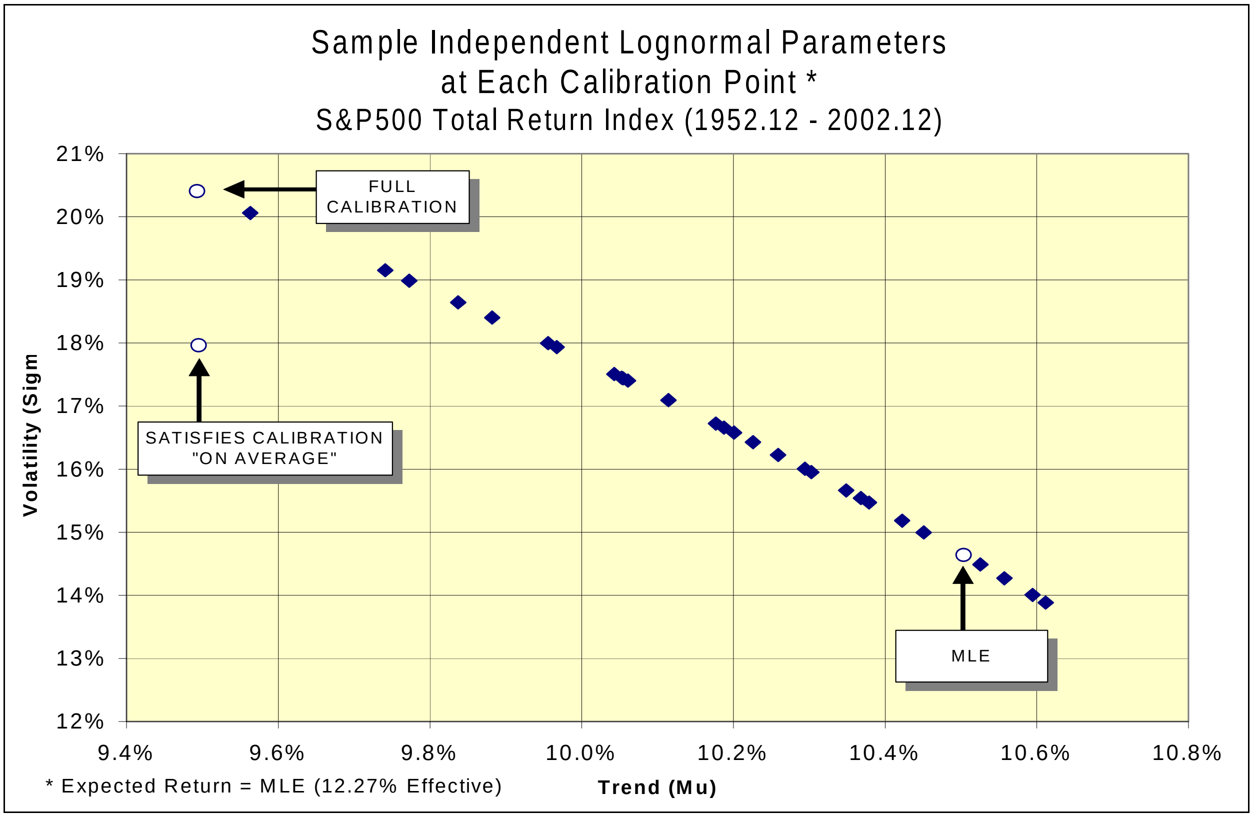

Sample Independent Lognormal Parameters
at Each Calibration Point *
S&P500 Total Return Index (1952.12 - 2002.12)
21%
20%
19%
18%
17%
16%
15%
14%
13%
12%
Volatility (Sigm
FULL
CALIBRATION
SATISFIES CALIBRATION
"ON AVERAGE"
MLE
9.4%
9.6%
9.8%
10.0%
10.2%
10.4%
10.6%
10.8%
* Expected Return = MLE (12.27% Effective)
Trend (Mu)

Chart 3

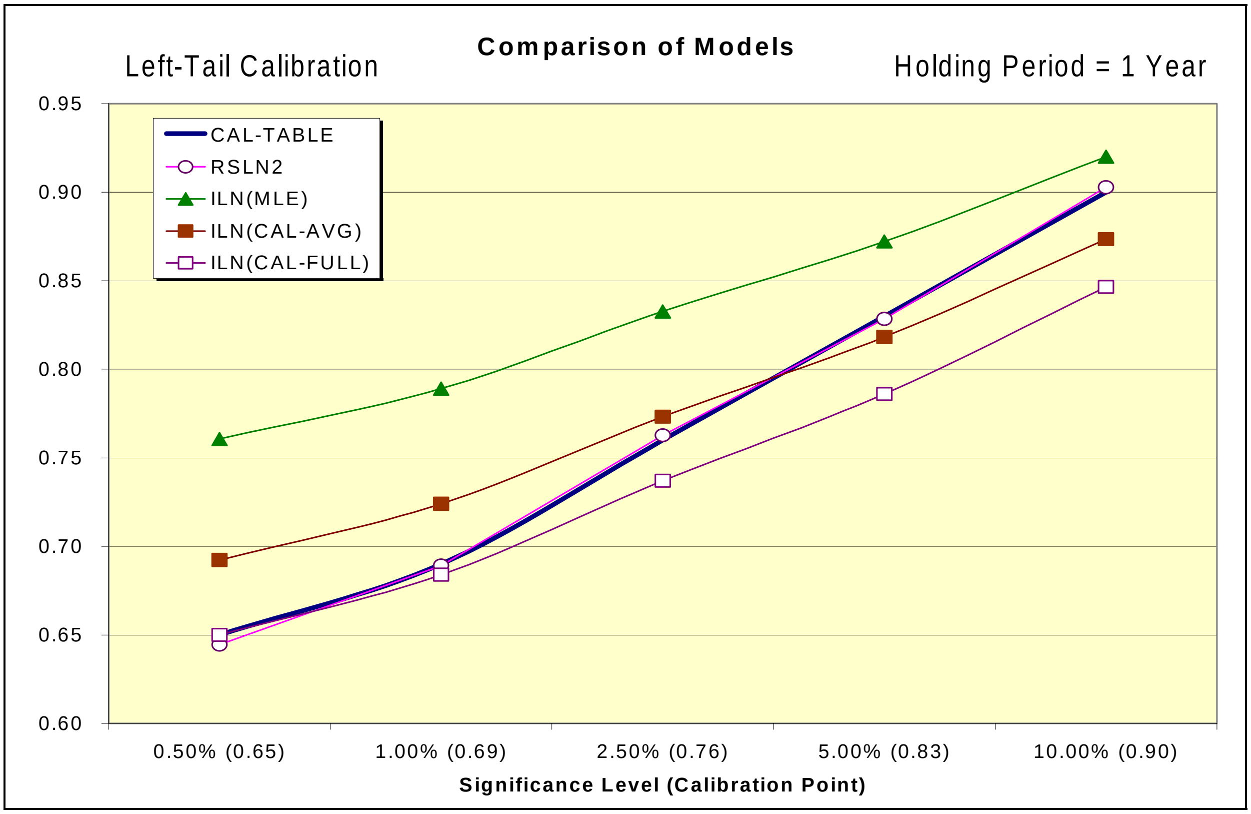

Comparison of Models
Left-Tail Calibration
Holding Period = 1 Year
CAL-TABLE
RSLN2
ILN(MLE)
ILN(CAL-AVG)
ILN(CAL-FULL)
0.95
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0.50% (0.65)
1.00% (0.69)
2.50% (0.76)
5.00% (0.83)
10.00% (0.90)
Significance Level (Calibration Point)

Chart 4

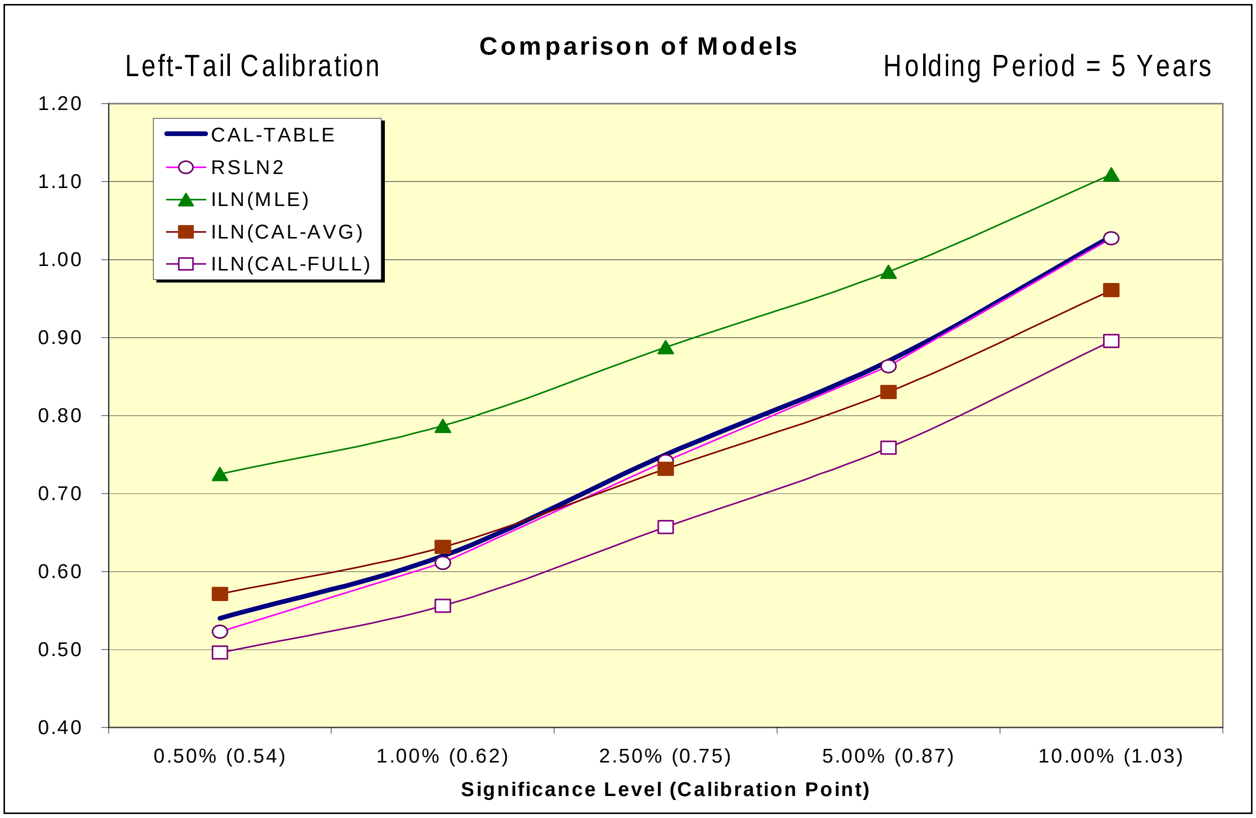
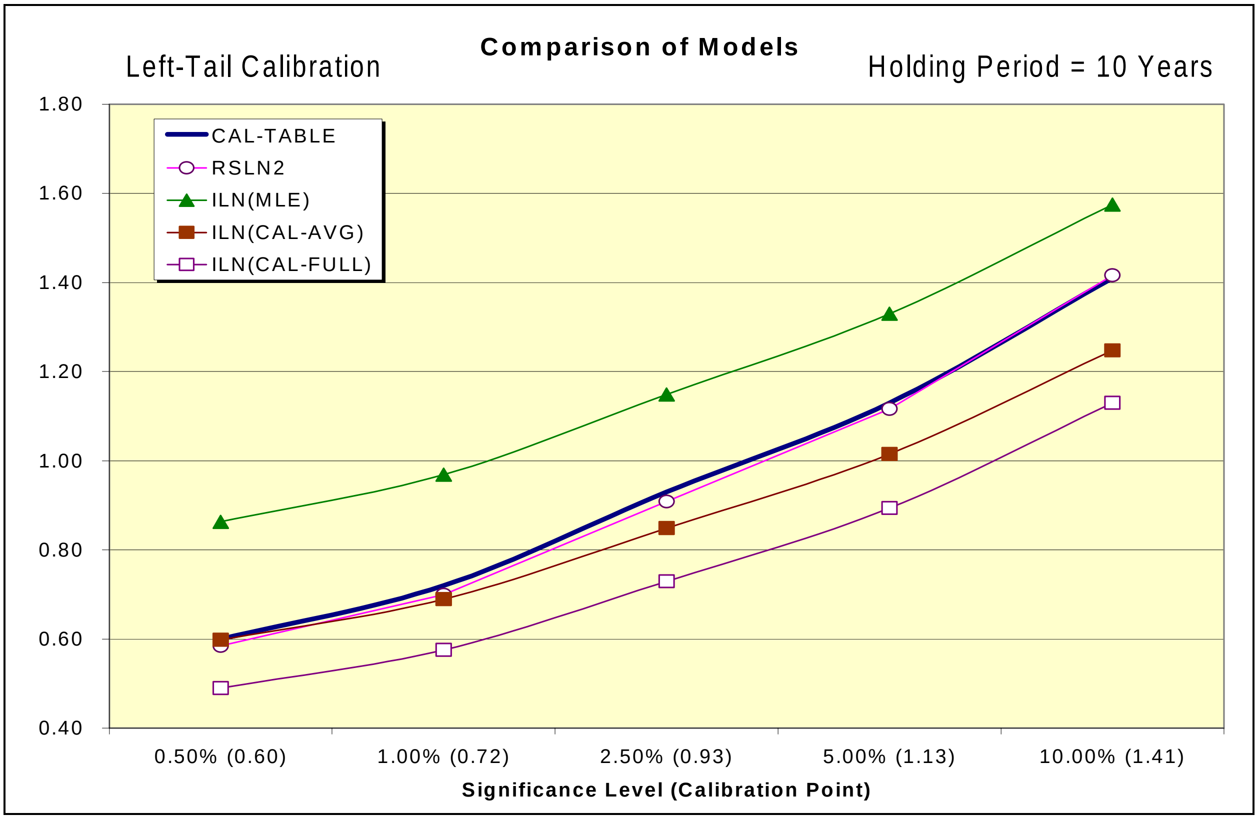

Comparison of Models
Left-Tail Calibration
Holding Period = 5 Years
1.20
1.10
1.00
0.90
0.80
0.70
0.60
0.50
0.40
CAL-TABLE
RSLN2
ILN(MLE)
ILN(CAL-AVG)
ILN(CAL-FULL)
0.50% (0.54)
1.00% (0.62)
2.50% (0.75)
5.00% (0.87)
10.00% (1.03)
Significance Level (Calibration Point)

Chart 5



Comparison of Models
Left-Tail Calibration
Holding Period = 10 Years
1.80
1.60
1.40
1.20
1.00
0.80
0.60
0.40
CAL-TABLE
RSLN2
ILN(MLE)
ILN(CAL-AVG)
ILN(CAL-FULL)
0.50% (0.60)
1.00% (0.72)
2.50% (0.93)
5.00% (1.13)
10.00% (1.41)
Significance Level (Calibration Point)

Chart 6

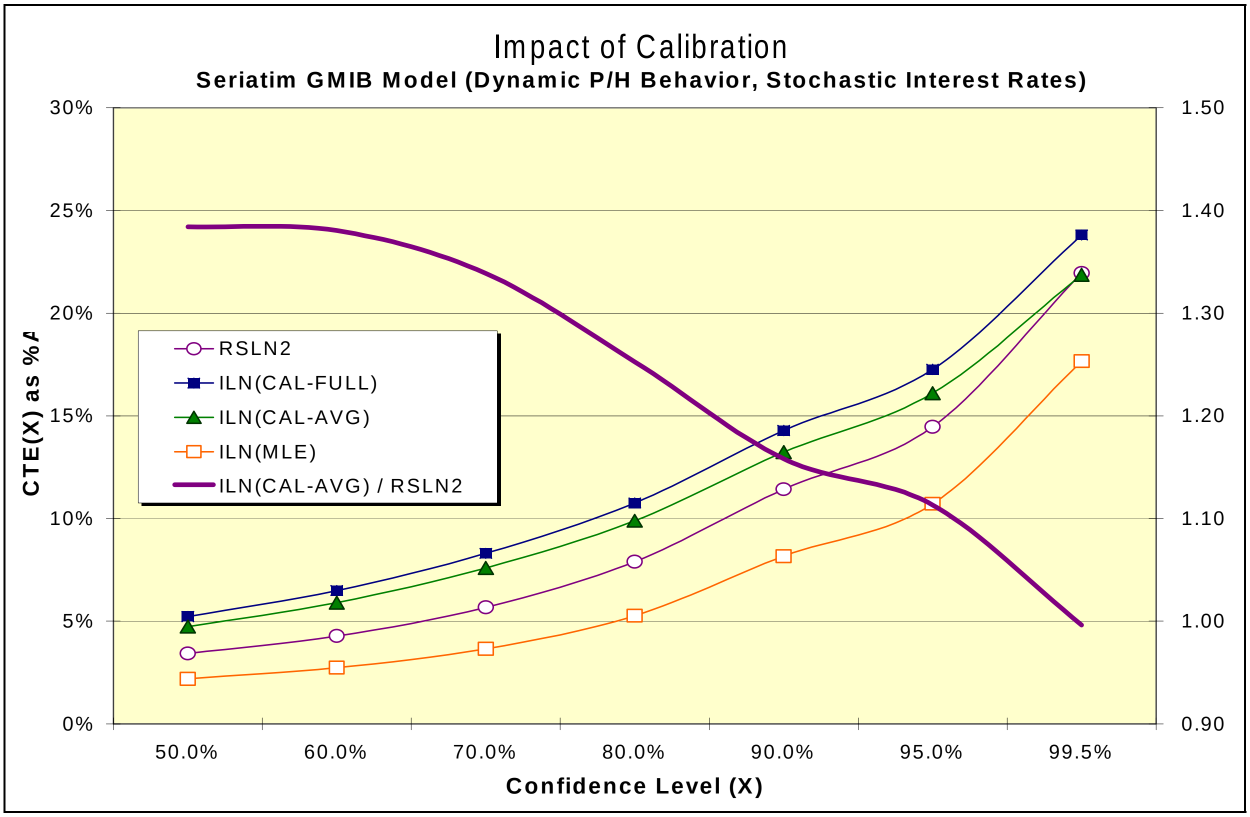

Impact of Calibration
Seriatim GMIB Model (Dynamic P/H Behavior, Stochastic Interest Rates)
30%
25%
20%
15%
10%
5%
0%
1.50
1.40
1.30
1.20
1.10
1.00
0.90
CTE(X) as %A
RSLN2
ILN(CAL-FULL)
ILN(CAL-AVG)
ILN(MLE)
ILN(CAL-AVG) / RSLN2
50.0%
60.0%
70.0%
80.0%
90.0%
95.0%
99.5%
Confidence Level (X)

Chart 7

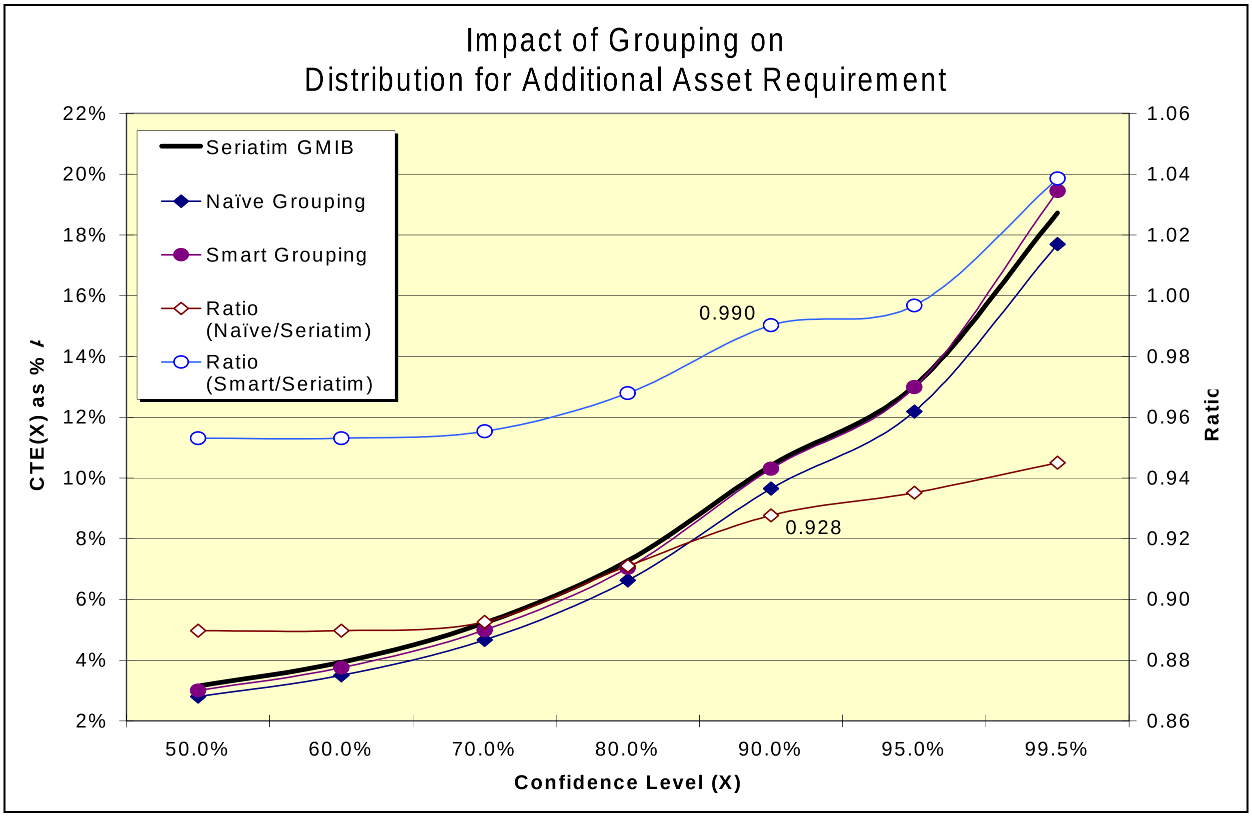
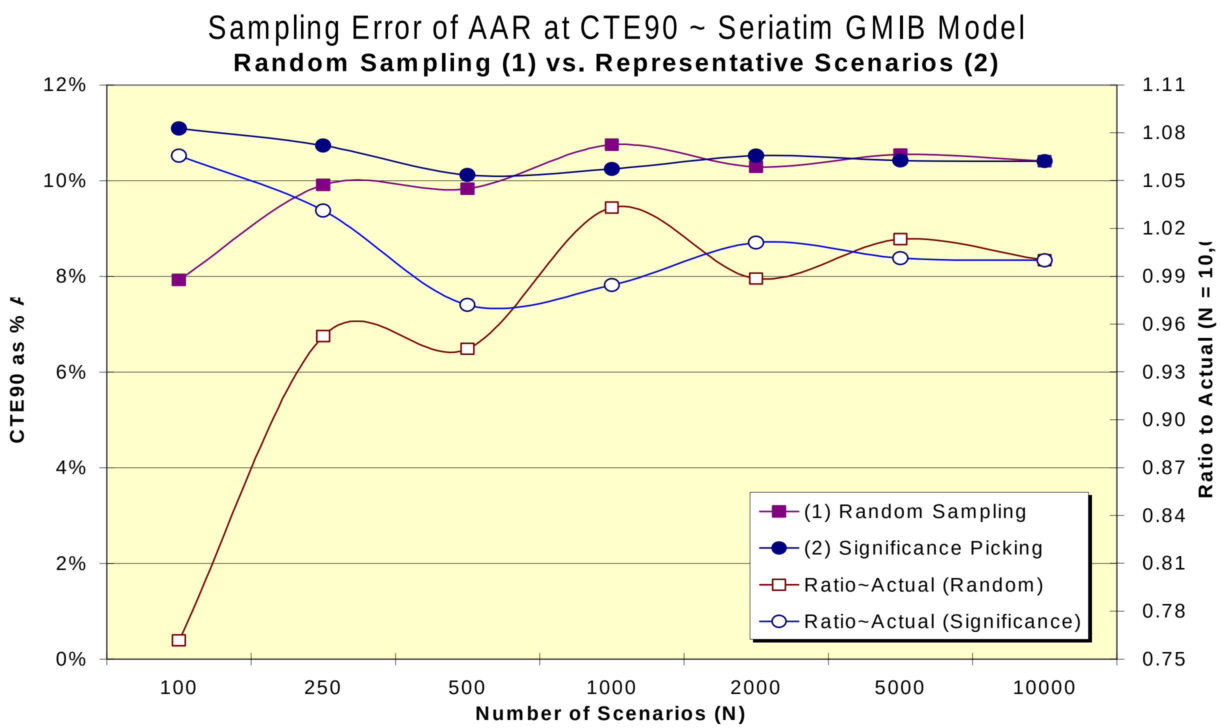

Impact of Grouping on
Distribution for Additional Asset Requirement
Seriatim GMIB
Naïve Grouping
Smart Grouping
Ratio (Naïve/Seriatim)
Ratio (Smart/Seriatim)
0.990
0.928
22%
20%
18%
16%
14%
12%
10%
8%
6%
4%
2%
1.06
1.04
1.02
1.00
0.98
0.96
0.94
0.92
0.90
0.88
0.86
CTE(X) as % /
Ratio
50.0%
60.0%
70.0%
80.0%
90.0%
95.0%
99.5%
Confidence Level (X)

Chart 8



(1) Randomly selected from a universe of N = 10,000 RSLN2 scenarios such that % starting in regime 1 = invariant probability.
(2) Stratified selection from universe based on significance measure for each scenario. Invariant probabilities maintained.

Chart 9

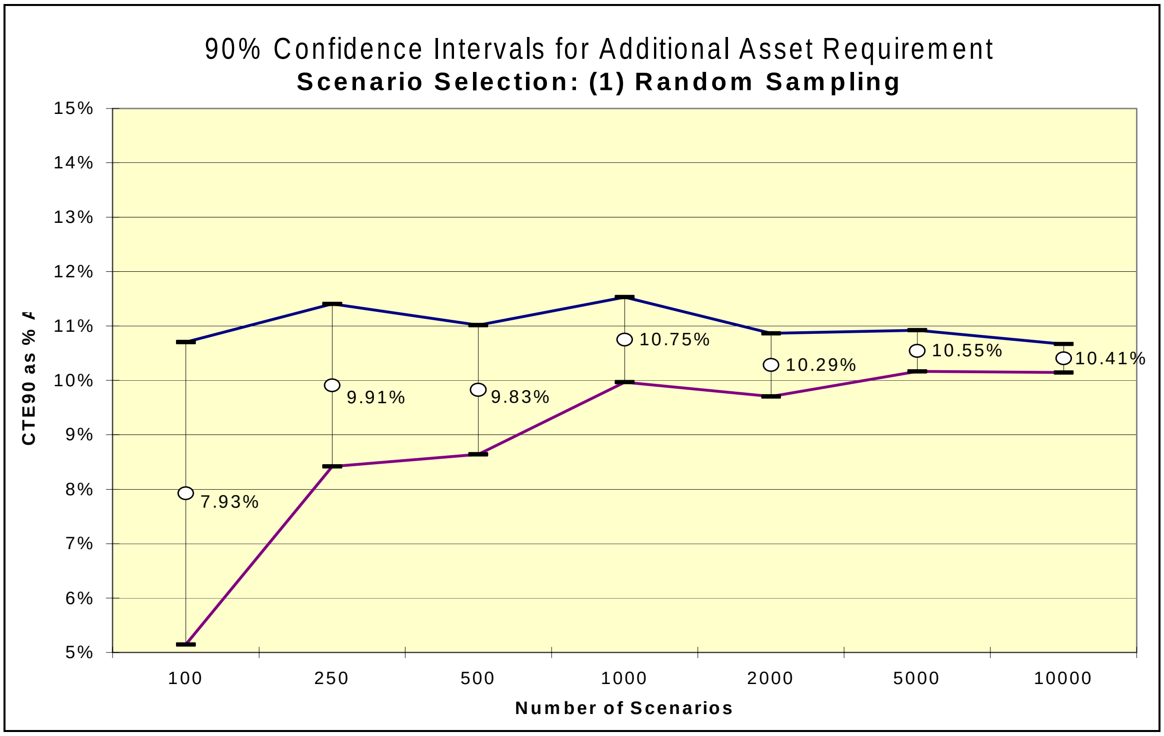

90% Confidence Intervals for Additional Asset Requirement
Scenario Selection: (1) Random Sampling
15%
14%
13%
12%
11%
10%
9%
8%
7%
6%
5%
CTE90 as %
7.93%
9.91%
9.83%
10.75%
10.29%
10.55%
10.41%
100
250
500
1000
2000
5000
10000
Number of Scenarios

Chart 10

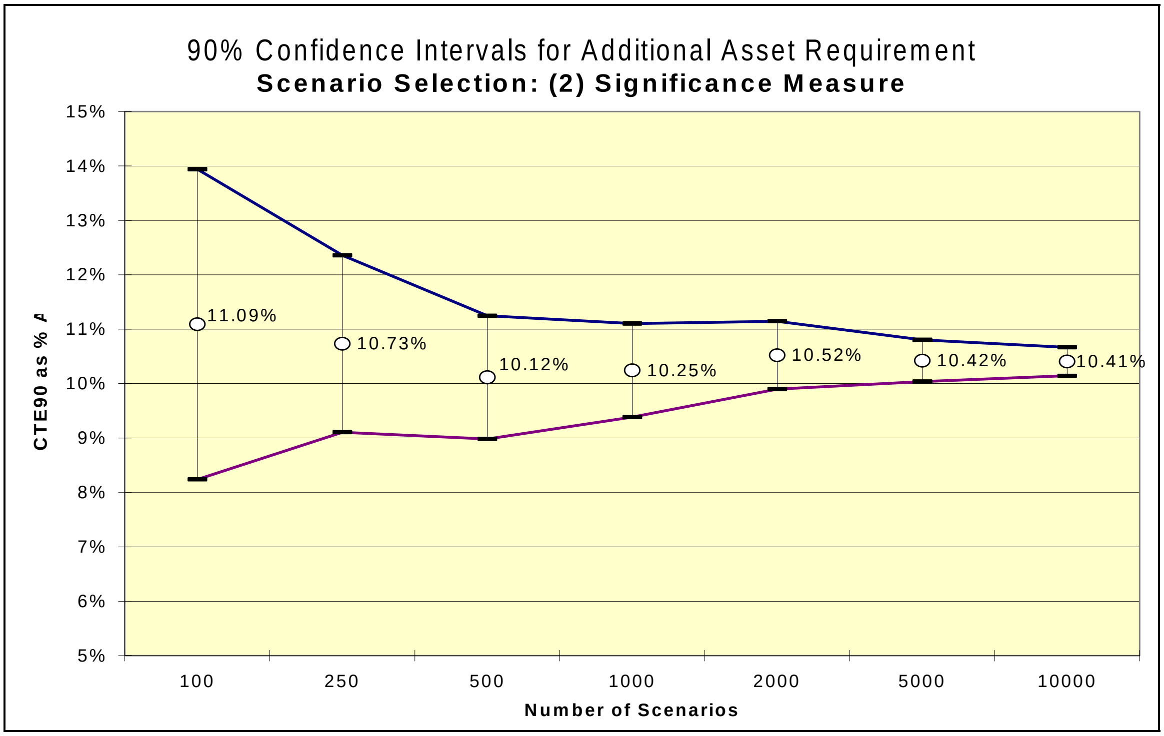

90% Confidence Intervals for Additional Asset Requirement
Scenario Selection: (2) Significance Measure
15%
14%
13%
12%
11%
10%
9%
8%
7%
6%
5%
CTE90 as % A
11.09%
10.73%
10.12%
10.25%
10.52%
10.42%
10.41%
100
250
500
1000
2000
5000
10000
Number of Scenarios

Chart 11

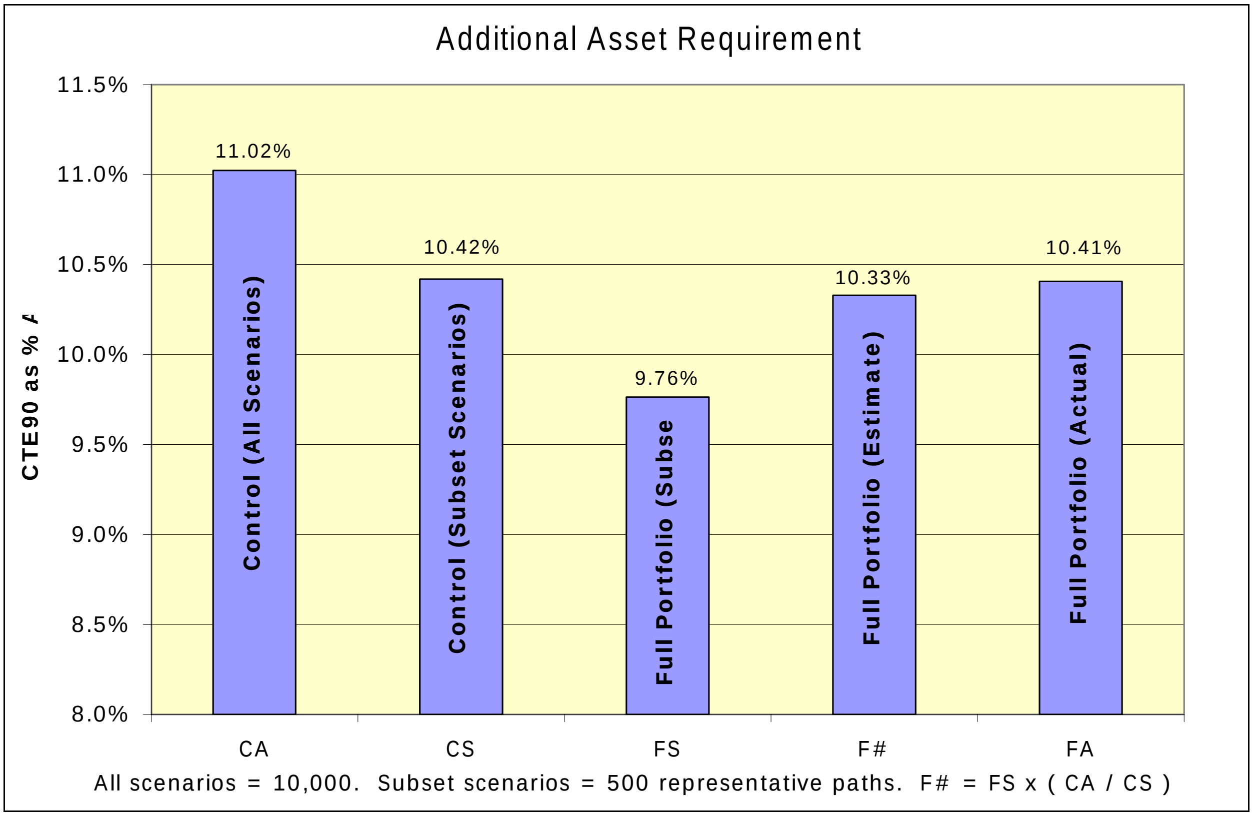

Additional Asset Requirement
CTE90 as % A
11.5%
11.0%
10.5%
10.0%
9.5%
9.0%
8.5%
8.0%
11.02%
10.42%
9.76%
10.33%
10.41%
Control (All Scenarios)
Control (Subset Scenarios)
Full Portfolio (Subse
Full Portfolio (Estimate)
Full Portfolio (Actual)
CA
CS
FS
F#
FA
All scenarios = 10,000. Subset scenarios = 500 representative paths. F# = FS x (CA / CS)

Chart 12

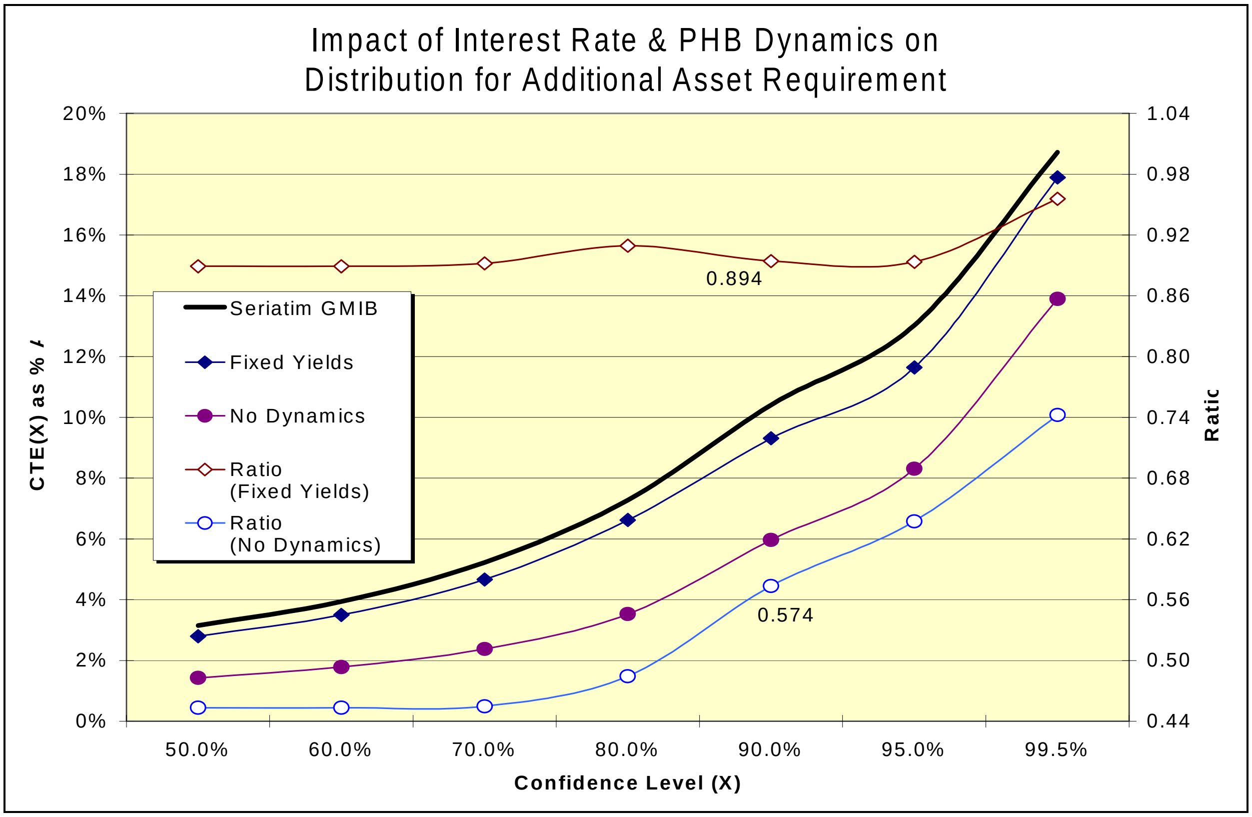

Impact of Interest Rate & PHB Dynamics on
Distribution for Additional Asset Requirement
20%
18%
16%
14%
12%
10%
8%
6%
4%
2%
0%
CTE(X) as % /
1.04
0.98
0.92
0.86
0.80
0.74
0.68
0.62
0.56
0.50
0.44
Ratic
0.894
0.574
Seriatim GMIB
Fixed Yields
No Dynamics
Ratio
(Fixed Yields)
Ratio
(No Dynamics)
50.0%
60.0%
70.0%
80.0%
90.0%
95.0%
99.5%
Confidence Level (X)

Chart 13

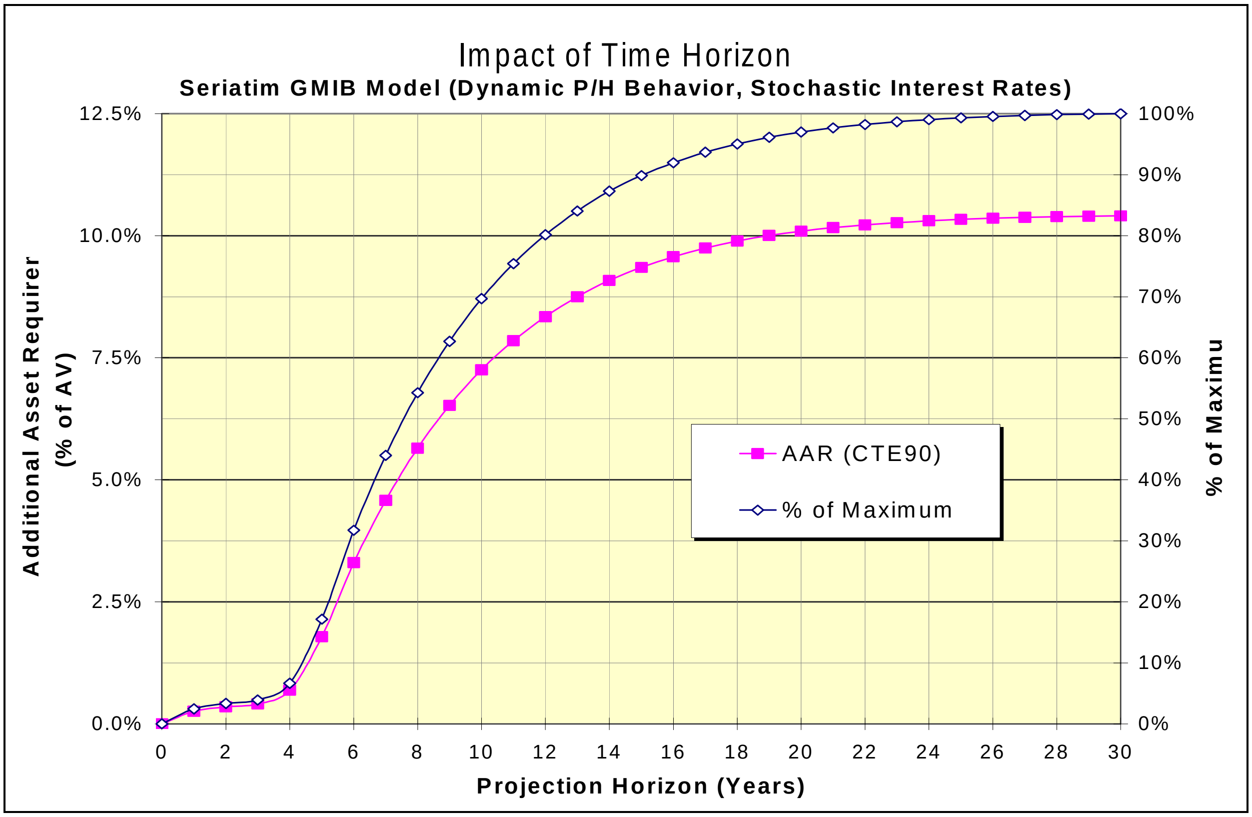
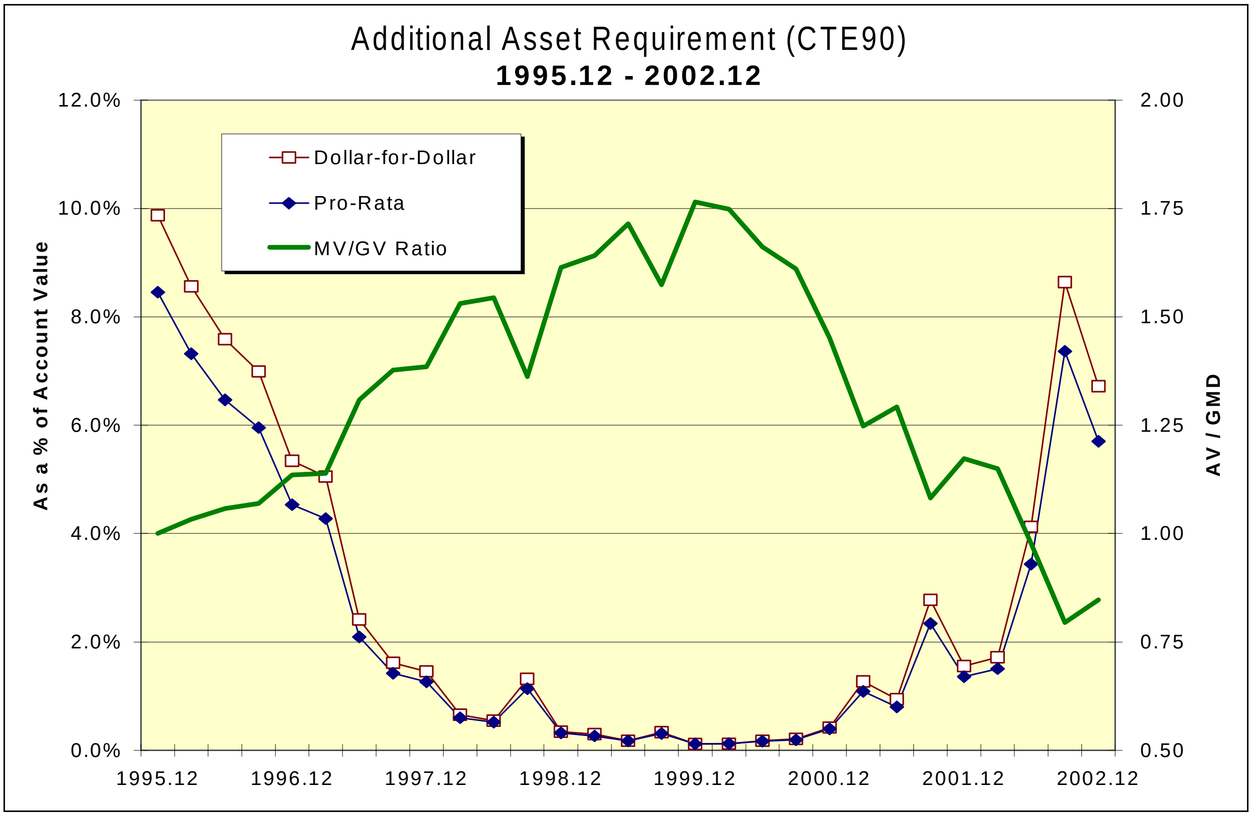

Impact of Time Horizon
Seriatim GMIB Model (Dynamic P/H Behavior, Stochastic Interest Rates)
Additional Asset Requirer
(% of AV)
% of Maximu
12.5%
10.0%
7.5%
5.0%
2.5%
0.0%
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
0
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
Projection Horizon (Years)
AAR (CTE90)
% of Maximum

Chart 14


Additional Asset Requirement (CTE90)
1995.12 - 2002.12
Dollar-for-Dollar
Pro-Rata
MV/GV Ratio
As a % of Account Value
AV / GMD
12.0%
10.0%
8.0%
6.0%
4.0%
2.0%
0.0%
2.00
1.75
1.50
1.25
1.00
0.75
0.50
1995.12
1996.12
1997.12
1998.12
1999.12
2000.12
2001.12
2002.12