

RECORD, Volume 29, No. 1*

Washington D. C. Spring Meeting
May 29–30, 2003

Session 61PD Benchmarking Investment Performance

Track: Investment

Moderator: ANSON J. "JAY" GLACY, JR.

Panelists: ANSON J. "JAY" GLACY, JR.
SCOTT S. HARTZ
WILLIAM H. PANNING†

Summary: Investment performance increasingly drives insurance product profitability and company risk exposures. Investment professionals labor to navigate challenging and surprising capital markets.

This session addresses how to:

- Objectively measure a portfolio manager's investment performance
- Evaluate relevancy of market indices, such as the Lehman Aggregate, as performance benchmarks for insurance liabilities
- Successfully benchmark performance against insurance liabilities

MR. ANSON J. "JAY" GLACY: I'm with New England Asset Management, the insurance investing arm of Gen Re, which is owned by Berkshire Hathaway, the diversified holding company. I'll discuss something that's become more and more important in recent years and that's standards for measuring and presenting the investment performance of life insurers.

Scott Hartz is the senior managing director and senior portfolio manager in the bond and corporate finance group with John Hancock, where he oversees the \$45 billion dollar general account bond portfolio. Scott is an FSA and a CFA. Scott will start off by helping us assess and evaluate investment performance on the single side of the balance sheet, without consideration of liabilities. Bill Panning is

* Copyright © 2003, Society of Actuaries

†Mr. William H. Panning, not a member of the sponsoring organizations, is Executive Vice President at Willis Re in New York, NY.

executive vice-president at Willis Re in New York and has responsibility for analytic services, including actuarial and financial modeling.

Let me start out very briefly by discussing the Association for Investment Management and Research (AIMR) whom most of you probably know best for their sponsorship of the chartered financial analyst (CFA) exams. They are also very active in corporate governance issues and lobbying efforts. They were in the thick of things with the Enron debate of last year. They also propagate performance presentation standards, which they initiated in the early 1990s once they realized there was a need for standardization in this area. In the mid 1990s, AIMR developed what we know as performance presentation standards (PPS) for North America, which were very well received. Towards the end of the 1990s, AIMR realized that with the globalization of not just business, but also investing, there was a need for it to develop a global standard or a core set of global principles that would govern investment performance presentation. That's what happened in 1999 when they approved global investing performance standards (GIPS).

The performance presentation standards that we operate under in North America are a sub-set or a version of GIPS. Two big differences between GIPS and PPS in this country are that we require fee schedules to be disclosed if you want to achieve PPS compliance. We also require, in the United States and Canada, that you disclose 10 years worth of investment performance where GIPS only covers five.

In the early 1990s, AIMR, as I mentioned, tried to address the standardization problem. Essentially the goal was to solve the selectivity problems that afflicted the measurement of investment performance, such as cherry picking, the convenient choice of time periods and survivorship bias.

These are the main selectivity issues that they intend the PPS to address. From AIMR's standpoint, the benefits of standardization include an apples-to-apples measurement basis, which allows any investment company to be measured against any other investment company and any portfolio manager to be measured against any other. Good standardization creates a fair playing field for competition. Finally, AIMR is very interested in globally promoting a self-regulatory solution. They intend to preclude government action in this area by establishing good, strong and voluntary standards for investment performance measurement.

The key goal of PPS is to achieve full and fair representation of investment performance and full disclosure of it. Another goal is to ensure uniformity in methodology to enable comparability. To promulgate minimum required standards worldwide is also a goal, but they also put some recommended standards on top of it that you can adhere to if you desire. Again, the key notion is to foster self-regulation on a global basis.

There are several key concepts of GIPS or PPS. Compliance is important, but not at an individual product or composite level. You must include all fee paying and

discretionary portfolios in at least one composite. The firms themselves who are trying to achieve compliance with GIPS or PPS have the flexibility to define what discretion means. Usually discretion in this context means that the investment manager has power to act on behalf of his or her client. Composites are nothing more than collections of portfolios with similar asset allocations with similar investment objectives. Finally, the definitions that you use for investment performance should be well documented and applied consistently.

GIPS requires that portfolios be evaluated at least monthly. In our firm, we do our performance measurement on a daily basis. GIPS requires the use of time-weighted total rates of return, like an internal-rate-of-return in actuarial parlance. You'll hear practitioners mention the so-called Dietz method, which is commonly applied to calculating total rates of return. A composite that you build from individual portfolios will reflect aggregates of investment objectives and asset allocations. They must be asset-weighted and cannot exclude cash. Trading expenses must be deducted, unsurprisingly. You have to disclose the treatment of derivatives, which is an area that I personally expect to see more build out in the PPS in the coming years. Finally, you need to disclose performance as net of fees.

GIPS, the core global standards, require you to present at least five years worth of performance. In this country and in North America generally, you have to present at least 10 years to achieve PPS compliance. You have to report at least annual returns for all years. There's a "dispersion" co-efficient that measures how dispersed the portfolios that aggregate to the composite are.

There's a compliance statement that you have to follow word by word to vouch that you have indeed achieved compliance with these standards. If you measure against a benchmark, the actual performance of the benchmark over the five- or 10-year period must be shown. Finally, an audited verification report has to be issued attesting to compliance. Typically, that's done by a third or an independent party.

The steps to compliance are very simple, very direct and logical. First of all, you need to construct composites with relevant and comprehensive data. Then compute the returns using the time-weighted method. Disclose the results and create presentations including them. Declare your compliance by issuing the statement that AIMR requires. Hopefully you also maintain compliance over the years.

Why did the AIMR feel it was necessary to transition from North American-only standards to the GIPS standards? They recognized the increasing globalization taking place, the need for ethical standards worldwide and that global managers and global clients want one set of standards. They don't want a hodge-podge of competing measurement standards. The benefits for businesses include fair cross-border competition amongst countries, the facilitation of business development and improved efficiency and cost savings in investing activities.

Scott is going to give a perspective on evaluating investment performance from the

asset-only perspective

MR. SCOTT HARTZ: As Jay said, I have been involved in managing the John Hancock bond portfolio for the last 10 years or so. Part of that responsibility has been to look at, think about and measure the investment performance. I'm going to talk about a couple of different ways to measure performance: a total rate of return standard and a more yield based measure. I'll talk about why we drifted away from total rate of return and then I'll talk a bit about our yield-base system and how we overcome some of the problems people think of when they think of yield-based systems.

As an investment professional, a CFA charter holder, we've had the AIMR the performance standards, which require total rate of return measures, drilled into us. I have to say that total rate of return is the right way to measure investment performance. There's been no debate at AIMR, there's been a lot of work around this and no one has ever suggested that it be done any other way. This is a standard for most investment managers, mutual funds and pension funds. As part of my job, I also manage about \$2 billion of money for pension funds in private placement bonds. We certainly use total rate of return there. We wouldn't think of doing it any other way. It captured all aspects of return and it's clearly the right way to go.

Despite that, we've transitioned to a yield-base system. I'll have to spend a bit of time explaining why. Part of it has been our transition to a public company. Hancock went public in the year 2000. Part of my premise will be that it's very difficult for a public company to use total rate of return performance measurement.

One reason to avoid total rate of return measurement at insurance companies is that assets are difficult to value. When you're doing total rate of return, you need to value the assets on a regular basis. As Jay mentioned, his firm does it everyday. Just to give you the landscape of what assets we're talking about at insurance companies, they are predominantly fixed income, bonds and commercial mortgages. Bonds are difficult to value. The former head of our department was fond of saying the bond market is like a rug market. What he meant is the guy selling the rugs knows a lot more about how much they're worth than the person buying them. The person buying them is apt to get fleeced. That was a word of advice to us buyers of bonds. But it also show you how hard it is to value bonds. Bonds are traded in an over-the-counter market, not an exchange trade market. You cannot pick up the *Wall Street Journal* and say I've got Kodak, 6% bonds, 2021, how much are they worth? You will not find it in there. There are some isolated instances where you will. Treasuries are in there, but by and large, corporate bonds are not. If you want to buy a corporate bond you have to call up three different brokers and ask them to give you an offer on the bond or a bid if you want to sell it. It's very hard to value.

With that said, there are some standards and there are some pricing services. With

public corporate bonds we all use pricing services that value them. However, the prices aren't always so good. I just wish I could trade on those prices, because you can make a lot of money on the bad prices that come out of that. But if all of our assets were public bonds, the pricing, on average, wouldn't be too bad. Most of the larger insurance companies also have large private placement portfolios, including commercial mortgage portfolios. There's no real pricing service that will price those. You have to develop your own internal pricing system and hence, you don't get very good market prices. That's the first problem.

Our accounting systems are designed to produce statutory filings and when we tried to measure total rate of return in the past we found this was probably the number one problem. We ended up having to develop our own new system. We looked through a bunch of vendor products, which probably have gotten a lot better over the years. At the time they were not very good. We developed our own system. It was challenging. The raw data still has got to come out of your accounting systems. Although they have gotten better over the years, that's still a major problem.

Benchmarking is very difficult. It can be done, but it's difficult. We probably have about 20 different portfolios within our general account each for a different business. To create a benchmark for each one is very difficult. Think about a long-term care product. How do you create a benchmark that matches up to those liabilities? It's a very hard thing to do.

Performance attribution is very difficult, but it is very important and it's complicated. Our accounting systems are inadequately equipped to handle it. You ask questions, such as what if treasury rates hadn't changed last month? I want to keep those rates the same and see what my total return would have been. That's how you start breaking things down and getting performance attribution. Our accounting systems are just not up to that challenge.

These can all be overcome, but it's very time consuming, and it's very expensive. We tried to do it and we never got exactly where we wanted to go.

Total rate of return measurement can cause a short-term focus. I would suggest that the last thing we need is more short-term focus. If you do these things daily, monthly or quarterly, you start needing to look at your returns in a very short period of time, and you need to do a lot of explaining when prices on certain bonds have moved a certain way. You start to spend a lot of time and energy explaining your short-term results.

You also lose a competitive advantage. I think it's a great competitive advantage when insurance companies are not tied to a total return measurement. Most of the rest of the world is. I've talked to a lot of total return managers at mutual funds, pension funds etc., and they will often times avoid a particular bond or sector because its short term price fluctuation is high. Maybe you're getting paid a fair amount for that, but they just can't stand the volatility. What does it do to their

returns versus their index in the short term?

As an example, back when long-term capital management exploded, they were investing in CMBS. Because of their need to sell, the prices on AAA rated CMBS went down quite a bit. They became very volatile. A number of total return managers I know had just got out of that sector because they couldn't stand the volatility. We weren't concerned with the short-term volatility. We actually ended up buying more because they were cheap. It's a competitive advantage that we're not stuck with these temporary swings in market prices.

Another reason total return doesn't work for us is that business unit customers could care less. As I said, we've got about 20 different portfolios and what do they care about it? For the most part, their goals are growing sales and growing earnings. Let's say as investment managers we thought corporate spreads were going to widen. So we sold all our corporate bonds, brought treasuries and lo and behold, it happened. We outperformed some sort of corporate benchmark that we're tied to. Our total returns would be great but our business unit managers would be saying, "Yes, but last quarter you didn't earn any spread on your assets. My earnings went down and I'm getting yelled at by senior management. You're not really tied into my metrics." That has been a big problem for us over the years.

Going public had a big impact on us. Shareholders who follow our stock now play a very large role in our company. They also don't care very much about total rate of return. They're more interested in things like earnings per share, return on equity and product spreads, which are basically return on assets. They come in and we talk to the security analysts on a regular basis within the confines of the full disclosure regulations. They ask us a lot of questions about how we manage the portfolio. A key question they ask us a lot is, "What are you investing in now? Are you investing in bonds that create enough spread so you can maintain the profit, the product spread for the return of assets in your businesses?" That's what they are concerned about. If we show total rate of return numbers, they'll look at them and be mildly interested, but it's not really what they care about.

Now, the number one reason to avoid total rate of return at insurance companies is that the senior management is not interested in it. By senior management I mean particularly the CFO, but also the CEO, and the board of directors, who all are our bosses. We've shown them total rate of return numbers over the years, and they have not been interested. They are interested in GAAP earnings. They want to make sure they hit their earnings per share (EPS) targets and they're afraid if they miss them by a penny that our stock will go down by 10%, someone will move in and buy us and they'll be out of a job. That is a little over simplification, but it feels like that's what drives their behavior. I revert back to my prior example on how we can do a great job in total rate of return and hurt the company on a short-term basis. I would argue that in the long-term it's going to definitely improve the earnings and improve the economic value of the firm, but you have to live through the short-term to get to the long-term and they're concerned that we won't.

What does senior management care about? They care about earnings as I said, as well as earnings per share. Earnings can be reported in a number of different ways. There are three that are the most common. There's operating income, which is the one analysts tend to care most about and tend to put a multiple on to get your share price. Net investment incomes affects the GAAP operating income. There's a lot of focus on net investment income.

GAAP net income is the second choice. That would include realized capital gains and losses, which have largely thought to have been unrelated to the basic business. They'll move around a lot, but should not amount to much over the long term. Because they're volatile, it's hard to value a business when you include those. The earning stream is too choppy and they should wash each other out over the long-term. In the last couple of years it's been nothing but losses in insurance company portfolios because of the economic environment we've been in. The analysts are starting to look at this and say, "Well, maybe GAAP net income is a better measure." There's been more concern on senior management's part about the realized gains and particularly the losses.

The final component, which I think really is the definition of GAAP income, would include what's called other comprehensive income (OCI). For the investment portfolio that means you include unrealized gains and losses on securities, bonds and common stock. It wouldn't include commercial mortgages and some other asset classes. It doesn't really capture all the unrealized gains and losses.

The accountants basically want us all to move to a mark-to-market system on both the assets and the liabilities. This has just been one step in that direction. FAS115 requires the mark-to-market of securities. Ultimately, I do see a FASB requiring us to mark both sides of the balance sheet to market. At that point in time I think total rate of return will be the right way to measure performance. Until then, we're driven more by GAAP operating income.

So, we care about GAAP operating income, hence we care about net investment income. We need to figure out what component of net investment income (NII) is under the control of the investment manager and what is not. Obviously we want to measure those that are. The uncontrollable factors would be things like cash flow. We've all, for the most part, had a great increase in cash flow lately because the fixed products have been selling better than the variable products. Net investment income has been going up significantly. Of course, the expenses from the insurance side, what we pay out on our products, has been going up as well. We can't just measure NII, because cash flow will drive where it goes and that's not in the control of the investment manager.

Interest rates are another factor. At our shop, we pretty much try to immunize ourselves against interest rates. Others may take some interest rate risk, and so what we measure will be a function of your style. But let's say assets and liabilities stay the same from the beginning of the year to the end of the year. You're going

to have some turnover in the investment portfolio and interest rates have been doing nothing but going down over the last however many years. NII is just going to gradually drop each year if the portfolio is not growing. That's obviously not in the control of the investment professional.

Finally, credit spread is really just another component of interest rates. It is an important one to think about. Our basic business model is that we raise money at our claims paying rate, AA rate, and we invest in more BBB type rates. We've earned that spread. When that spread is narrow, it's more difficult to produce good returns. It is easier to produce good returns when it's wide. But again, that is not under the control of the investment professional.

What is in control? The primary thing is the spreads we achieve on our new investment relative to what's available in the market place. The second would be investment delay. If the product folks bring in a bunch of cash and we just sit on it and don't invest it for a while, I'll assume we've hedged it. We're maybe not sitting in a money-market fund, or maybe we bought treasuries or derivatives to hedge. We're going to earn a treasury-type rate with no spread and that's going to hurt earnings. That is under our control. That's been a big concern in the last several years.

What do you compare these things against? What do you benchmark them against? That is a big part of what we're supposed to be talking about today. There are two different things you can benchmark against. We're a bit schizophrenic in our shop. Sometimes we look at one and sometimes we look at the other. The corporate folks and senior management are always trying to push us back to the ROE targets that we have at the company. We price for a certain amount of defaults and that's built into our pricing model, which generates the ROE of the company. They want to keep comparing our investment results for what we had priced for. Now I don't know about you, but ROE targets have changed maybe once in the last three years, whereas the investment markets have been through incredible upheavals in the last three years and change on a day-to-day basis. That's not very market driven. The investment professional will continually argue to compare ourselves against some sort of peer group or the market as a whole. The problem is this is often difficult to define.

I'm getting to my yield-base approach and getting to some more specifics as to what exactly we do. We start by looking at our acquisition spreads relative to "target." We've moved to a peer group base target. The ACLI has some very nice surveys of where insurance companies are buying private placements and buying mortgages, and we'll compare our spreads on new investments to those. In the public market, we'll compare them to either similar sorts of surveys or the new issue spreads in the market. That's the starting point for our system. We have thought about, and at times looked at, comparing our new issue spreads to those based on ROE model and our issuance cost, because if we issue a product at some sort of rate, and we put all our pricing assumptions, that means we got to buy

investments at this other rate. We could compare our investments to those rates. But again, it gets to all those sort of problems about our ROE not changing with market conditions. Obviously it is very important with all of these is to adjust for the credit quality of the new investments and their average lives.

A lot of work goes into adjusting this new issue rate. There are a lot of other things that will affect the investment portfolio, our GAAP earnings and are under the control of the investment managers, and we need to look at each of these. These get us back closer to a total rate of return management. In fact, as we looked at our system over the years we've said, "Well, total rate of return really captured this. It's not captured on the ROE system, so let's figure out how we can adjust our yield basis to capture it." If in fact, it's something that's under the control of the investment managers.

We calculate actual defaults. We price for X amount of defaults. Is that what we should compare it against? That's what hits our earnings. Or we should compare it against the amount of defaults in the market in general on a risk-adjusted basis? In the last couple of years obviously the results are very diverse based upon whether you're looking at market-type defaults or long-term assumptions on defaults.

There's no real market basis to compare investment delay to, but in pricing, you may assume if you're investing in private placements it takes you a month or two to find an investment so you price for that and you compare how you did relative to that.

This works well for buy and hold investment strategies. This is probably appropriate for private placements and commercial mortgages, but on public bonds there's trading involved. Trading actually brings up a whole other topic, which I'm not going to address. But if you're doing much trading, you're changing the yield on your portfolio as interest rates in the market are changing, and that will influence your earnings, not necessarily economic value. That's a problem for your GAAP result. If you buy a bond at a spread of 150 over treasury and you sell it the next day at a spread of 200 over treasury, you've destroyed value. That's something you want to pick up. We do look at all our sales and compare the sales to those same new issued targets we're looking at for acquisitions.

In the last couple of years we've had huge amount of downgrades in the portfolio and that causes us to need more capital to back our portfolio. It also limits the amount of risk we can take in a new investment, so there's real cost to the company. We will look at those and charge for those.

As interest rates come down if your bonds have pre-payment options in them that allow borrowers to buy pay off, you have to reinvest. The current interest rate environment is going to be a cost to the business. We look at that and charge for the pre-payment experience.

Investment expenses are benchmarked against peers. It's just another thing we'd want to look at. I want to conclude here by just reiterating my basic thesis, which is that total rate of return is the right way to measure investment performance. It doesn't fit well with the accounting standards we need to adhere to. I think those accounting standards will evolve over time and move us more in the direction of looking at total rate of return as the right way to measure investment performance.

MR. WILLIAM PANNING: Benchmarking investment performance really consists of creating investment equivalents of liabilities. Liability benchmarks have other uses as well that I won't really talk about in detail, but I want to at least mention them. I'm sure they are probably obvious to most of you.

First, liability benchmarks can be useful in pricing a product, particularly a complex product. It mystifies me how products can be priced without doing something like this. Second, they can also be very useful in risk analysis in determining what could happen under extreme scenarios to the firm. As we all know, market conditions can vary enormously. So having an investment equivalent of a liability can be extremely valuable when we are trying to figure out just what kind of stress the firm might be subject to under a long-term scenario. In both cases, the translation of a liability into an investment equivalent makes it possible to place a value on the liability that can be compared with the value of the assets that are being used to fund the liability and to measure the relative risks of each. Although these uses of liability benchmarks are important, I will not talk about them. I'll focus on just the investment aspects.

My main topic is how to go about constructing a liability benchmark. I'm going to begin with a very simple real example that deals with pension liabilities. Later on, I'll allude to some of the problems in dealing with high convexity liabilities, where we have essentially granted options to policyholders. I'll then deal with performance measurement and performance attribution, and I certainly agree with Scott in that these are both very difficult. I'll focus on risk because normally when we try to assess whether someone has outperformed a benchmark we find it useful to know how much risk they took in doing so. Then, finally, I'll have some summary remarks.

First, let's consider how to build a benchmark for the very simple case of a defined benefit pension obligation that includes both active and retired lives. This is about the simplest case I can think of, because these liability cash flows are not interest-rate-sensitive for the most part. What you receive from the pension actuaries is a set of projected cash flows over time. These overall cash flows are re-forecasted annually, but year-to-year changes in these forecasts are typically not significant. In any case, for any given year, you have a set of forecast liability cash flows that you know you're going to be measured against.

Now the particular circumstances of this example are such that, if one were to do this exercise today, one might come out with a different solution for creating a

benchmark, simply because markets have changed in the last couple of years. My example is about four years old, and the treasury market was a little different back then. Despite all of that, the implications of this case study are, I think, absolutely valid, even though today we might use a market other than the treasury market to implement it.

The objective in creating liability benchmark is simply to translate projected cash flows into a portfolio of real securities that can be priced and/or purchased. A liability portfolio is thus a hypothetical portfolio of real securities. This sounds like something that's obvious, but I don't think it is. In our industry there's a strong tendency to create complex liabilities for which there is no corresponding investment portfolio. I'll address that again later. To the extent that we create something that's not investable we automatically create a problem for the investment professionals who are going to somehow be measured against those liabilities.

In this particular case where we've got non-interest sensitive cash flows and we translate them into real securities, the value of this portfolio, whether or not we actually purchase it, will be treated as the value of the liabilities. The performance of this portfolio will be what the actual portfolio that funds the liabilities will be measured against. In this particular case, treasuries are especially useful because of their liquidity. Most other asset classes are simply not available across the whole range of maturities and are not sufficiently liquid to be useful. I hasten to add, though, that the treasury market has its own quirks. On-the-run treasuries, which consist of the most recently issued treasury securities, are the most liquid. They are extremely liquid, but have widely spaced maturities. If you have annual cash flows, but have anywhere from two years to even 20 years between different treasury securities that are on the run, trying to interpolate in between becomes a real problem. This situation is not necessarily impossible, because there are other treasuries besides the most recently issued ones. They are a little bit less liquid, but the treasury market is still the most liquid market. Nonetheless, treasury securities are absent in some maturity ranges. Moreover, there are a great many treasuries in some maturity ranges of the yield curve that have call features that are undesirable.

As a consequence, in this particular case, we looked at treasury strips. These are securities that are somewhat less liquid than regular treasuries. They simply consist of zero coupon bonds created by placing treasury securities in a portfolio and then issuing a portfolio of new securities, each of which matures at a single point in time and has no cash flows between the present time and the maturity date. There are two types of treasury strips: those created from principal payments and those created from interest payments. The fact is that principal strips have the same problem as ordinary treasuries in having maturities that are not equally spaced. Interest strips, on the other hand, appear on a very regular basis throughout the whole maturity range of the yield curve that, four years ago extended out to 30 years. An interest strip can be created for any date on which a treasury bond pays

interest. As a consequence, there were interest strips for every quarter extending out to 30 years. Because of their greater availability and convenience, we decided to use interest strips rather than principal strips for creating this particular benchmark.

Now one of the problems we ran into is very common in creating benchmarks. This is the fact that projected liability payments were forecast to occur once each year going forward, whereas treasury interest strips have maturities that are spaced quarterly. Our actuaries were very reluctant to try to fine-tune liability payments more than annually. This sounds like a simple problem, although there would be a variety of fancy ways to solve it. Our solution was simply to create quarterly payments by dividing projected annual payments by four, and then essentially to match these quarterly payments with interest strips having the same maturities and face values.

Now why did we convert annual to quarterly cash flows in such a simplistic fashion? The answer—and I want to strongly stress this—is that we tried to avoid fancy procedures. The reason for avoiding fancy procedures is that when you're actually calculating total returns or other measures having to do with performance attribution, you often find that your procedure is one of the factors that affects returns. You don't want to get yourself in that situation. It often becomes something that you trip over. Trying to keep everything simple becomes very important. People want to know not only how you performed, but also *why* you outperformed or under-performed. You don't want one of the answers to be that we adopted this particular procedure for massaging cash flow projections, and that resulted in a five point gain or a five point loss. That's not really a very convincing answer for most of the audiences that care about performance attribution. So the first lesson I want to stress is to keep benchmarking as simple as possible.

There was another problem that illustrates yet another practical issue concerning benchmarks. The liability cash flows on these particular products continued well beyond the longest treasury. Projected liability cash flows went out to something like 100 years. There's no treasury that long, or anything else except a few extremely long corporate bonds that are too idiosyncratic to really bother with. So our problem was how to use treasury strips to represent cash flows longer than 30 years.

Our solution was designed to be simple. We took the yield on the longest strip, the 30-year strip. We discounted all cash flows that occurred beyond 30 years by that yield and then essentially used that number to create an artificial cash flow at the 30-year point. In other words, we calculated the present value of all cash flows beyond 30 years and represented this as a cash flow at the 30-year point. Again, the rationale was simplicity. If we don't adopt this particular procedure, we have to come up with some fancy procedure for pricing cash flows that occur beyond 30 years. Such procedures do exist, but I don't think there is any industry standard for doing so.

Let me quickly summarize the benchmark and then point to a couple of problems we ran into even though we tried to keep everything simple and straightforward. The benchmark in this case was a real asset portfolio consisting of treasury interest strips with maturities spaced quarterly for 30 years. The face amount of each strip included in the benchmark was one-fourth of the projected pension liability cash flows for that year. The face amount of the 30-year strip was augmented by the present value of the cash flows occurring beyond 30 years, discounted at the 30-year strip yield. That sounds like a very straightforward benchmark, about the simplest that one could possibly create, given a liability stream that is very transparent.

There are a couple of implications of what we did. One is that there's nothing perfect about this. I would argue that no procedure for constructing a liability benchmark is perfect. There are always tradeoffs among different imperfect procedures. Second, I think one of the virtues of what we did was to keep the procedure as simple and transparent as possible, which helped us to explain what happened and why, even if only to ourselves. The problem with complex procedures is that they can fool us as well as others.

Now let me address high-convexity liabilities for a moment. What happens when you deal with products where the firm has granted various kinds of options to policyholders? In essence, you and the firm are short options. The classic problem in our business is that you are short options when you create the liabilities and you are short options when you buy securities. You're always short options. In this particular case, the problem of trying to construct a benchmark portfolio that is truly investable may simply be insoluble. Instead, what you may have to do is construct a benchmark that matches the liability derivatives, by which I mean the first derivative with respect to interest rates and perhaps the second derivative as well. You want to have a portfolio that has the same interest rate sensitivity or duration, and the same convexity (the change in duration that results from a change in interest rates) as your liabilities. That's a trickier problem.

One way you can solve this problem is to essentially include in the benchmark some actual derivatives. That is one alternative. You may have to buy options or something that creates the convexity that you're short.

Complex liabilities, such as equity-linked annuities, can be approached in two ways. One is to find a combination of options that matches the liability. Another is to create a dynamic strategy that synthesizes the liability. These two ways of dealing with complex liabilities are, in fact, equivalent. In both cases the real risk is model risk, which is the fact that we have only an imperfect understanding of how clients will behave under different circumstances. Model risk is something over which the investment manager has no control.

The key point is to create a portfolio that closely replicates the liabilities. In the case of very complex liabilities this may be impossible. If so, then it's preferable to

use transfer pricing, a procedure that separates the liability into two components, one of which is investable and one of which is not. For example, the product pricing may assume that policyholders will not be efficient in exercising options that have been granted to them. That's really not in the investment manager's control, and not something that the manager can hedge. A transfer pricing approach essentially allows the investment manager to construct an investable benchmark that incorporates assumptions about policyholder behavior, but deviations from those assumptions become the responsibility of another part of the organization.

Let me give a simple example of how transfer pricing might have worked in the recent past. There were numerous firms that offered equity-linked annuities at prices that presumed that policyholder attrition rates of 50% over a seven-year period. The problem with these models was that they assumed such an attrition rate over all scenarios. In reality, however, when equity markets returned 30% or more in three consecutive years, policyholders had enormous incentives to keep their policies. Firms whose hedging had assumed high attrition found themselves with actual liabilities that far exceeded what their models had projected. This was model risk at work.

In these circumstances, a transfer pricing solution would have required the investment manager to assume attrition rates like those incorporated in the pricing of the product. The financial consequences of deviations from this rate, which were substantial, would have been attributed to the product managers rather than to the investment managers, since the product managers were responsible for the risk that customers' behavior would deviate from what they had projected in pricing the product.

The simple example that I presented earlier described how we constructed the benchmark. The performance target for the asset manager was simply the total rate of return of the benchmark, plus a given spread. In practice, there are pitfalls in determining whether an investment manager has actually achieved that objective.

One pitfall consists of pricing problems. Scott covered these so I won't say much about them, but one of the solutions for pricing problems is to try to use a single pricing service whenever possible in the hope that there will be consistency across different types of assets. You know that pricing services make errors, but you hope they are consistent errors across all securities. We used the same pricing service for the actual asset portfolio as for the treasury interest strips in the benchmark. That doesn't solve all problems, but it helps. Pricing is one of the inherently difficult and vexing problems that you run into.

A second problem is that sometimes the benchmark itself can create a problem. In the particular case I described, where we used treasury strips, the 30-year strip was heavily weighted because, as you may recall, it included the present value of all cash flows beyond 30 years. But here is where we encountered a very subtle

problem in the market. At that time, the 30-year bond was actively issued, and the 30-year interest strip was created from the interest payments on this bond. Suppose that we bought some of these strips. Now consider what happens six months later when a new 30-year bond is issued. This new bond is very liquid, and the old 30-year bond becomes less liquid. As a consequence, the old 30-year interest strip becomes less liquid and a new 30-year strip comes into existence. From a liability standpoint, nothing significant has occurred, since the old liabilities that are 30+ years are now mapped onto the new 30-year strip. But from an investment standpoint, it has been necessary to sell the old 30-year strip, which now has a maturity of 29.75 years, and buy the new 30-year strip. This results in a loss of value that is unavoidable in the real world.

By contrast, the benchmark would not suffer any such loss, because the cash flows longer than 30-years were always discounted and then invested in a 30-year strip. There's an implicit assumption in the benchmark that they would be costlessly transferred from the old strip to the new strip. But the costless transfer of that sizable amount of money could not, in reality, be done. As a consequence, we found ourselves losing out to the benchmark every month simply because of the way we had constructed it.

The solution in this particular case was very simple. We changed the benchmark. Instead of using the 30-year strip to match the present value of all cash flows beyond 30 years, we put one-fourth of the total into each of the four longest strips. This doesn't actually eliminate the problem entirely, since the market behavior of the 30-year strip also affects the strips that precede it in maturity. But we changed the benchmark to dilute the problem. Before we made the change, the behavior of the 30-year strip was the single largest factor in our total return, especially during months when a new long bond was issued. Although we didn't know it at the time, one reason for this was that Long Term Capital was taking huge positions in the longest strip, which exaggerated the problem with our benchmark. After we made the change, their actions had far less impact on our results. This particular problem, although narrow, has a broad implication: Don't be afraid to use market knowledge in constructing a benchmark. You don't want your benchmark to be highly sensitive to the peculiarities of the markets you're dealing with or to a few securities or circumstances. You'll find that if it is highly sensitive to a few securities you end up forcing yourself to buy them simply as a defensive maneuver.

I'd like to turn now to performance attribution. Typically, whether you out perform or under perform as an investment manager, you'd like to know why and so would everyone else. Performance attribution is essential, but it's an imperfect procedure for finding the answer. Scott alluded to this and I'd like to explain a little bit more about why this is so difficult.

It's more difficult than just a matter of finding good prices. I've taken a segment of a yield curve with the spot rate—the yield of a zero coupon bond—with a maturity

of T and the spot rate of a strip with the maturity of T minus one, whatever that is, it could be a year. A year later the yield curve has shifted down and also flattened out. We'd like to do some performance attribution on this. We asked ourselves, "What were the factors that lead to my making money?" I'm going to concentrate only on the fact that this yield curve shifted. Let's assume that we're invested totally in treasury strips, so we don't have to worry about things like spreads, exercise of options and all of that other stuff. All we have to worry about is the performance of treasuries. We'll see that there's some real ambiguity in what we mean by a performance attribution, even in this very simple case.

What do we know about a treasury strip? When we consider the initial yield curve we anticipate two things happening. First, we anticipate that because the yield curve and spot rate curve are typically upward sloping we're going to make money simply from rolling down the curve. In other words, since the treasury curve is sloped upwards, we're going to start at a higher point than where we end. Over the course of the year you think that the yield on your strip will drop. That means the total return should be pretty good.

The second thing that happens is that the curve itself is moving. In this case, when the curve moves down I make even more money on a total return basis, because the curve has dropped. I want to separate out those two components of my return. How much of my return is due to the roll-down effect, which would have occurred without any market changes? How much is due to the fact that the curve itself dropped? One way to answer these questions is to look at the difference between the T -period yield and the $T-1$ -period yield on the beginning curve and use that yield difference to calculate the roll down effect, and then use the difference between the beginning $T-1$ -period yield and the ending $T-1$ -period yield to calculate the effect of a change in the curve itself.

There is another possible way to do it as well. That is to use the difference between the beginning and ending T -period yield to calculate the effect of curve shift, and then to use the difference between the ending T -period yield and the ending $T-1$ -period yield to calculate the effect of roll down. These two procedures can give quite different answers. In my example, the curve shifted down and flattened, so the two procedures give answers that are extremely different.

My point here is that there's nothing about those two procedures that says one is correct and one is not. It's simply arbitrary. This is one of the problems that classically occurs with performance attribution. The problem becomes even more complex when other factors such as spreads are included.

Still another problem with performance attribution arises because the various factors in performance affect returns in a multiplicative fashion. This creates a problem for people like CEOs who expect things to be additive. Of course, you can use the logarithm of total return, in which case the components do become additive, but here the problem is that most senior executives are not very

comfortable with logarithms either.

All solutions to these problems are somewhat arbitrary. That doesn't mean that you can put in any number you want. There are limits on how far things can be arbitrary. Most of the solutions involve adopting a convention, a rigid set of rules that prioritizes the order in which the effect of different factors is calculated. I might say I'm always going to calculate roll down before I calculate the effect of curve shift. That's one solution, that's the way I'm always going to do it. I have consistency from month to month, or year to year.

Now let me turn to the issue of relating performance to risk. Suppose we constructed a benchmark. Suppose we measured how that benchmark performed, we measured how the asset manager performed, and suppose that the asset manager is outperforming the benchmark. We might want to ask whether he took a lot of risk in the process. It's not necessarily a good thing to outperform a benchmark if you took fantastic risk to do so. But if measuring performance is difficult, measuring risk is even more difficult.

I'd like to give an example to show what the problem is. Although in this example I focus on equities, the example has implications that are true for almost any kind of portfolio. It's just simpler to demonstrate them with an equity portfolio.

In my example there are three portfolio managers that all start out at the same time with portfolios of 50% stocks and 50% cash but follow different strategies. The first one follows a buy-and-hold strategy. The second one follows a buy-low-sell-high strategy. That means that as stocks go down he buys more and as they go up he sells. The third manager follows a momentum strategy, and buys stocks as they rise and sells as they fall. This is just the opposite of buy-low-sell-high. Now I've constructed this example very carefully, so the behavior of these last two, the buy-low-sell-high and the momentum player, are mirror images of one another in response to market behavior. I've used a very realistic market simulation to explore the results for these three managers. The total number of simulations was about 50,000, each simulating daily investment results for a calendar year.

Now conventional wisdom would tell us that there are only two things that affect performance: the composition of the investment portfolio and the behavior of the market. This would tell us that the three managers are taking the same amount of risk. But what I'm going to show you is that this isn't true if your time horizon is longer than overnight. For longer time horizons, investment strategy becomes a third principal factor that affects performance. And strategy not only affects return, it also affects risk.

Let's look at the results here. The return for momentum was the highest. It was lowest for the buy-low-sell-high. Let's look at the standard deviation, which is a conventional measure of risk. For the buy-low-sell-high player, the standard deviation was 7.9%, almost equal to the actual return. For the momentum player,

the standard of deviation was 14% considerably higher than the 8.4% return. On a risk-adjusted basis the buy-low-sell-high guy certainly looks good. That certainly fits with conventional wisdom.

Let's also look at some other measures of risk as well. Let's consider, for example, value at risk measures. I've constructed three. We've got a 95% loss value at risk measure, a 99% loss, and then I've got the worst-case loss. If we look at these, we see a somewhat different story. We're looking at extremes now. Remember, standard of deviation is symmetric in that it takes into account extremes on both sides of a distribution – extreme gains as well as extreme losses. By contrast, these values at risk measures only consider the losses.

The rank order for the three managers is completely the opposite of that for the standard deviation. Whereas the standard of deviation was highest for momentum, all the value risk measures are highest for buy-low-sell-high. It completely reverses the order. If you ask which portfolio manager could lose me the most, it's clearly the buy-low-sell-high. In a worst-case scenario the buy-low-sell-high strategy produced a loss of almost 25%.

We can see the reason for this if we look at the return distributions. The buy-low-sell-high guy has a very long tail going out to the left. I mentioned replicating options with dynamic strategies. A buy-low-sell-high strategy is very similar to dynamically selling a call. You have a lot of downside risk and not a lot of upside potential. The buy-low-sell-high is clearly the highest in terms of the percentage of tail value off at the left extreme. The return distribution for the buy-and-hold strategy is in the middle. It looks like a conventional lognormal distribution, which it should. The momentum strategy has much more extreme tail. There's a lot of upside here. When the market goes down, the momentum strategy guy is pretty much out of the market by the time you get to really extreme negative scenarios because he sold off. That strategy comes close to dynamically buying a call.

The lesson from this example is that performance risk really has three components. One is the composition of the portfolio, especially differences between the liability benchmark and the assets actually held. The second is the volatility of the market conditions. The third and very important one is the strategy used in managing assets. It's non-trivial. It doesn't show up when you look at Wall Street firms. They use value at risk measures, but they do it on a one-day basis. It's essentially an overnight measure, so that strategy plays no role. For insurance companies that have time horizons that can stretch out extremely long, strategy becomes extremely important. If benchmarking is used for analyzing risk in a product, strategy needs to be specified. It can't just be implicit in the analysis.

It's difficult to measure these components of performance risk over relevant time frames. We should always compare performance on a risk-adjusted basis. The only tools that we have right now are scenario simulations and exposure limits for measuring and constraining risk. There's a lot of work that needs to be done in

coming up with fully adequate measures of risk, particularly given all the limitations in performance attributions and performance measurement. This is an area that our industry needs to address and is addressing.

In summary, liability benchmarking is useful to asset managers and to the rest of the firm in three ways. First, it translates liability cash flows into real assets, which the firm is short. It therefore provides a basis for identifying opportunities to increase return because you know that you're short the securities in the benchmark and you have to compare to them the expected performance of the securities that you're long. Second, it provides a partial basis for performance attribution. And third, you can analyze the risk of the benchmark relative to the risk of your actual portfolio and estimate whether you are being adequately compensated for the incremental risk you are taking.

Because you are long on your real portfolio and short on the benchmark portfolio, you are very much like a hedge fund.

Benchmarking is both an art and a science, and using it for performance attribution remains difficult. Combining performance measures with risk measures needs to be done in ways that are, as yet, imperfectly understood.

FROM THE FLOOR: I was curious that you alluded to the appropriateness of using the Sharpe Ratio to justify the supposedly superior performance of the hedge fund market. What are your feelings about that?

MR. PANNING: I think the Sharpe Ratio is a valuable tool, but it's limited in one important respect. If you actually look at Sharpe Ratios in my strategy example, the buy-low-sell-high manager looks the best. Typically, the Sharpe Ratio subtracts the risk free rate from a portfolio's total return, and then divides that number by the standard deviation of portfolio return. So essentially the ratio compares performance, relative to the risk free rate, to risk, as measured by standard deviation. But the problem here is that returns are not necessarily symmetrically distributed, and I may care a lot more about how much I could lose than how much I could gain. This is why, in the example I presented, the standard deviation gives one result while the value-at-risk measures give a very different result. So a Sharpe Ratio it doesn't tell me much about upside risk versus downside risk, and that is in my view its biggest limitation.

The second thing that the Sharpe Ratio doesn't tell you much about are the tails of the return distribution. Long Term Capital Management is a good example of this problem, since they were doing things that had a low probability of loss, but when such a loss occurred it would be very large. Value-at-risk measures are more sensitive to tails.

FROM THE FLOOR: Would a value-at-risk measure be more appropriate for hedge fund?

MR. PANNING: Yes. Actually, I would consider value risk as a family of measures. Although value at risk itself has some limitations, I think that it is superior to many other risk measures because it focuses on downside risk.

MR. HARTZ: I agree with that. I've managed some money for pension funds and they do look at Sharpe Ratios quite a bit, and not so much at value at risk. Maybe that is because they've got a diversified pool of managers and they are not so worried about the downside in any one manager and hence diversify that way. Certainly the Sharpe Ratio is used extensively in the investment world. Although I agree with the comment that it is more appropriate for the life companies as the value at risk is more important for determining how much capital you need to hold and so forth.

MR. GLACY: Bill, when you did these 50,000 simulations, were they based on current market conditions?

MR. PANNING: I did that simulation about a year ago. There's one important thing that should be explained. I alluded to the fact that I tried to simulate the market realistically. There's one specific feature of these market simulations that is especially important, and that is simulating changes in market volatility in addition to simulating returns for a given level of volatility. This was really a two-stage simulation, in which I first modeled and simulated volatility and then simulated the return for a given volatility drawn from the distribution of volatility.

Now the importance of this procedure is that volatility does not fluctuate randomly from day to day or week to week. Basically if the market goes up, volatility tends to decline. If the market drops, particularly if it has an extreme drop, volatility tends to go up. That has an important implication for the difference between momentum players and buy-low-sell-high guys. Because if the market goes up and volatility is going down as a consequence, which is typically the case, the momentum player is buying into the market precisely when volatility is low, whereas, the buy-low-sell-high guys are selling. Now when the market drops volatility tends to spike upward, and that's when the buy-low-sell-high guy is buying. He's buying into volatility. He's increasing his exposure to volatility.

Now if it's the case that the expected return on the market is the same in low volatility and high volatility conditions, and I haven't found any statistical evidence that it's different, that means that the implicit Sharpe ratio for the guy who was doing the momentum strategy should be low, because he's buying at the time when volatility is low. He should have better risk adjusted returns than the guy that's doing buy-low-sell-high, who is buying when volatility is high. I just thought I'd mentioned this because I think it adds a little bit to the discussion.

FROM THE FLOOR: It struck me that it sounded like a slam-dunk for the investment manager to perform against a benchmark liability made up of

treasuries. I assume it's all in the spread to the benchmark. That is where the challenge is. Could you define how the spread is determined?

MR. PANNING: You're right. Our target was the benchmark return plus a spread. But it didn't feel like a slam-dunk to me because I used only treasuries to outperform the benchmark plus spread target. I actually had six portfolios, each with different spread targets. At that time, corporate spreads to treasuries were a lot lower than they are today, and our typical spread target was about 75 basis points over the benchmark. This assumed something like an AA-rated or A-rated portfolio. In fact, though, we used treasuries almost exclusively. We tried to identify specific treasury securities that we considered mispriced and we put a lot of money on them. These were total return portfolios, so we could do that.

You're right. It's all in the spread that is added to the treasuries you use in the benchmark. The reason for not using other securities is that they're not sufficiently homogenous to give you the same understanding of what is going on with the liabilities you are funding. Using a variety of securities makes your benchmark itself have all sorts of weird options and other characteristics. This makes it really hard to understand. Instead, you need some sort of homogenous security class to use in order to construct the benchmark. Then you add the spread to give yourself a challenge and the firm a profit.

MR. HARTZ: I might add that in 2002 treasuries outperformed corporate by a huge amount. Over the long term, it's a slam-dunk to outperform treasuries with a corporate bond portfolio. But in a short run, you can get very different results. In creating a benchmark, I would suggest that if you're adding a spread you want that spread be a bit dynamic based on what market spreads are doing because you are being measured against it, unless you're being asked to call that sort of thing. If you're being asked to move in and out of treasuries into corporate, if that's part of your strategy, then it is less important. But for the most part, we are asked to continually take corporate risks. You'd want to have a spread benchmark that would move around with market spreads.