Comment #8 – 4/29/11 – 11:23 a.m.

In the ASOP 27 exposure draft section discussing arithmetic and geometric returns, it was noted that the Pension Committee believes that "actuaries might benefit from additional education in this area". This investment topic is outside the traditional education of pension actuaries but recent developments in the pension arena require more and deeper understanding of this and other complicated investment topics. There are many good books on this particular subject, one of which is referenced in the attachment. The [document below] contains a relatively brief description of the issues regarding geometric and arithmetic means as this topic relates to measuring asset returns' actual experience as well as forecasting future asset returns and setting expected investment return assumptions.

Thank you, Elena

--Elena V. Black

Statistical Analysis of Historical Return Series

Geometric Mean and Arithmetic Mean of Historical Asset Return Series

We start with some general theoretical comments with regard to statistical analysis of asset returns and the distinction between geometric and arithmetic means in this context.

The arithmetic mean is a simple average of the elements in the series. For a historical return series, the arithmetic mean is an average of annual returns. The geometric mean of a return series is a compound annual rate of return over the multi-year period. A simple example illustrates the difference between an arithmetic mean and a geometric mean. Let us assume that we invested $100 for a 2-year period into a stock that over the first year returned 20% and over the second year lost 20%. At the end of the first year, we have $120 and at the end of the second year, we have $96. The arithmetic mean is 0% but the geometric mean is the square root of $96 divided by $100 minus 1, or negative 2%. It is clear from this example that the geometric mean is the appropriate measure to analyze a historical return series. Note that in this example, the arithmetic mean overstates the actual return for the two-year period, because our wealth over this period is actually reduced from $100 to $96, while 0% annual return implies that we experienced neither a gain nor a loss.

In general, the arithmetic mean is greater than the geometric mean and the two are equal only for a series consisting of constant terms. The difference between the arithmetic mean and the geometric mean is positively related to the variance of the series. The back-of-the-envelope "rule of thumb" is that the difference between the arithmetic and the geometric means is approximately equal to half of the variance. However, as discussed below, under certain assumptions more precise formulas for the relationship can be developed.

Expected Return on Assets

Annual returns versus compound annualized multi-period returns

When setting economic assumptions related to expected return on assets, a distinction must be made between setting our expectation for a single (annual) period or for multi-year periods, where an expected compound annualized return assumption may be required. The purpose of the assumption and the intended use is the key for this determination. The term "expected" generally relates to the uncertainty of the outcome and has a specific meaning in the context of random variables with probability distributions. Typically, expected return refers to the expected value of future asset returns that either are assumed to follow a specific random variable distribution or are generated using stochastic forecasting models. The 50^{th} percentile of a random distribution can be used as the expected return measure, but, at least in the context of probability and statistics, the "expected value" of the probability distribution has a specific meaning of a probability-weighted average of all possible values for a random variable.

When we refer to 'expected return', we assume that future asset return values follow some probability function either explicitly set by a standard random variable distribution or forecasted using Monte-Carlo simulation models. It is important to understand which expected return is required for a particular economic assumption because the annual returns and the compound annualized returns are different random variables with different probability functions and, thus, potentially different expected values.

Consider the following example. Suppose we want to estimate how much money we can expect in one year. Let us assume, for example, that the annual return variable R follows the normal distribution with mean of 7.5% and standard deviation of 12.0%. We can use Monte-Carlo simulation to forecast the normally-distributed variable R with these parameters and statistically analyze our outcomes in terms of terminal wealth after one year. If we are looking for expected value at the end of the year on an initial investment of $100, depending on the number of simulation trials, we may calculate our outcome mean to fall between $107 and $108. As the number of trials becomes sufficiently large, the confidence interval around the mean will become smaller and smaller and the simulation mean will converge to $107.5. Of course, as we set our simulation to a specific distribution, we already know the answer and the "true" mean of our wealth at the end of the year is $107.5, which we will obtain through a simulation model if we run a sufficient number of trials. In this case, however, there is no need to use a complicated simulation model.

Let us assume next that we want to determine the mean or expected value of our terminal wealth at the end of a 10-year period. Moreover, we would like to determine the expected value of an asset return that, compounded over 10 years, will result in terminal value with this mean. For each 10-year scenario or a "random walk", we start by simulating annual returns from our original normal distribution for the first year, then for the second year and each following year for the 10-year period. We repeat this 10,000 times. Essentially, we are looking at a product of random variables, which itself is a normal variable. Unfortunately, it is a difficult task to determine the probability distribution of a product of normal variables. Furthermore, if we are trying to calculate the expected value of the annualized compound return, then we also need to extract the n^{th} root of the product of random variables, which only complicates the already-difficult problem. In this situation, a stochastic simulation may be the only way to get an answer if we desire that each annual variable is normally distributed.

We used MATLAB software to obtain an answer to the question in this example. First, we simulated 10,000 trials or scenarios, where the annual returns for each year are independent identically-distributed normal variables with a mean of 7.5% and a standard deviation of 12.0%. The expected value of the terminal wealth in our simulation was $206.23 while the mean and the standard deviation of the annual compound were 6.9% and 3.9% respectively. The distribution of the annualized compound returns appears relatively symmetric. In our next step, we simulated 10,000 trials from the normal distribution with a mean of 6.9% and standard deviation of 3.9%, compounded the results for 10 years and calculated the terminal wealth. The mean of the terminal wealth outcome under this second simulation was $206.73. This result is extremely close to the result in the first simulation where we simulated terminal wealth directly. In contrast, if we use the original normal distribution for the annual variable but treat as the distribution of the

10-year compound variable, the expected value of the terminal wealth is $338.66 which is almost 65% higher than the desired result of $206.23. Reducing the variance also decreases the expected value of the terminal wealth, but the only way to obtain the expected value of the terminal wealth of $206 is to reduce the uncertainty around the mean of 7.5% to zero, which is clearly not reasonable for future asset returns' assumption.

In mathematical terms, if the annual return is represented by a random variable R (or annual wealth accumulation W = 1+ R) following a particular probability distribution, then the compound total return over the multi-period of N years, is an N^{th} root of a product of N random variables, generally assumed independent (but not necessarily) and identically distributed. Below are some useful facts to consider when thinking about multiple inputs of random variables or functions of random variables:

- If X is a random variable, then 1 + X is a random variable that follows the same probability distribution with mean (expected value) = mean (X) + 1 and the same standard deviation or variance as X.
- If one random variable is added to another, we can not simply add the probability distributions. The notable exception is a sum of independent normally distributed random variables which also follows normal distribution.
- The product of random variables is even harder to visualize than the sum. As with the sum, the product of random variables, even identically distributed, almost never results in a random variable with the same distribution. One rare exception is the lognormal distribution.

There are a couple ways to obtain an answer to the expected compound (and/or annualized) return problem described earlier. One approach is to utilize a stochastic simulation process (as we demonstrated in our example). The other option, often employed by investment and financial professionals, is to assume that the annual wealth accumulation variable W = 1+ R follows a lognormal distribution.

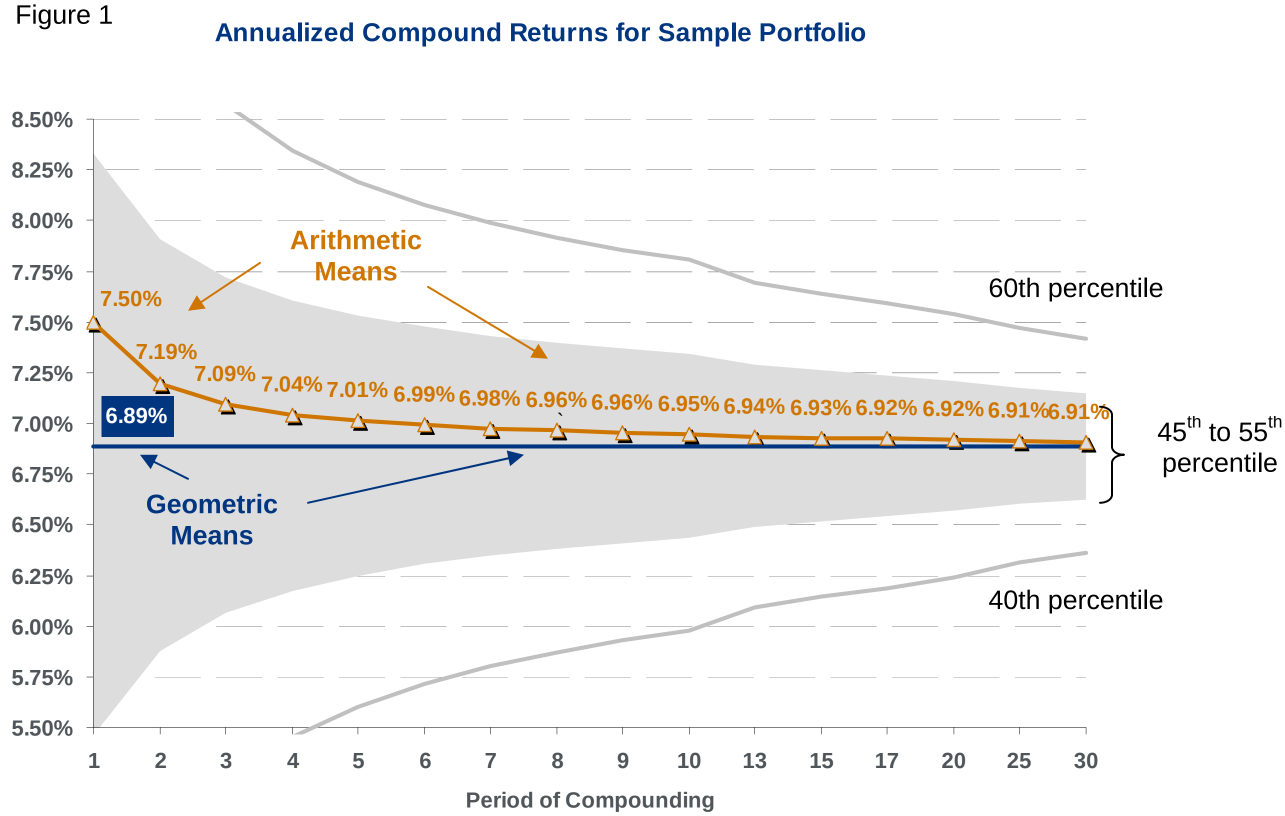
Lognormal Distribution

A random variable W follows a lognormal distribution if it is an exponential function of a random variable that is normal. In the context of the annual wealth accumulation random variable, one can think of the underlying normal variable as the continuously compound annual return. In any event, the magic of the lognormal distribution assumption is that the product of independent lognormal random variables is also a lognormal random variable. Moreover, extracting roots of the exponential function Exp(X) is the same as calculating Exp (X/n) where "1/n" is a constant. A random variable that is a product of a constant with another random variable is a random variable with the same distribution and easily determinable mean and variance.

This approach allows for "closed" formulas for the expected value of a multi-year compound annualized return variable, given the annual return lognormal distribution and the length of compounding period "N". The length N of the multi-year period becomes less important, because as N increases the expected values of the compound return distributions rapidly converge to the median (50^{th} percentile). Finally, it is easy to show that for any N-year period, the median of the lognormal distribution associated with the compound return variable for this period is the same as the median of the original lognormal distribution of the annual return variable. The median or the 50^{th} percentile of the lognormal distribution is also known as the geometric mean and is equal to the exponent of the mean of the underlying normal distribution.

To summarize, if we assume that annual wealth accumulation (1 + annual return) follows a lognormal distribution with the expected value M (the arithmetic mean) and the variance V, we can explicitly calculate the expected value of the N-year compound return variable as the geometric mean (50th percentile) of the same distribution. This would be true for sufficiently large N. The graph in Figure 1 depicts an example of a portfolio with annual returns assumed to be log-normally distributed with the expected annual return (aka arithmetic mean) of 7.5% and the standard deviation of 12%. The geometric mean of this distribution is 6.9%. For each “N” on the “x”-axis (that represents the length of period of compounding) we show both geometric and arithmetic means of the compound returns’ distributions, as well as certain percentiles around the 50th percentile for each distribution. Note that for any compounding period, compound return follows a lognormal distribution with the geometric mean equal to that of the annual return distribution. Recall that we simulated results for N = 10 using MATLAB and assumed a normal distribution for the annual variable. Note how close our simulated results of the expected compound return of 6.9% (with greater precision it was 6.8877%) is to the estimate calculated using the geometric mean under a lognormal distribution of annual returns assumption.

Figure 1



Estimating Parameters for Future Returns Probability Distributions

Whether simulation techniques or explicit closed-formula calculations are used, an important question is how we may estimate parameters of investment returns’ random variable distributions. If the historical market data is used to do the job, or as one aspect of relevant analysis, calculating the means (arithmetic or geometric) of the return series that we discussed earlier, may help in this task. The arithmetic mean of a return series views the series as a sample draw of the annual distribution and corresponds to the sample expected value of the distribution for an annual return variable. The geometric mean measure treats historical return series of a particular length as a random walk sample and corresponds to the estimate of the expected value of a compound return.

As our first simple example demonstrates, the geometric mean of the historical returns represents the backward-looking measure of wealth accumulation over multi-year periods of time and is appropriate for statistical analysis of past returns. Examples of forward-looking compound return measure are yields of fixed income investments or stock earnings' yields.

Care should be taken when using only backward-looking measures to estimate parameters of the assumed probability distribution of future returns. The historical means (arithmetic or geometric) are highly sensitive to the choice of timing and the length of the historical periods. For example, for historical S&P 500 Index *real* returns, the 20-year rolling averages vary from 0.84% (20-year period ending with 1981) to 13.34% (20-years ending with 1999) to 6.5% (20 years ending with 2010). Varying the length of the period also produces a wide range of S&P 500 annualized compound real returns for the period ending with 2010, from -0.9% (the last 10 years) to 7.32% (the last 30 years) to 6.9% (the last 70 years). This can lead to drastically different conclusions depending on the period selected for the analysis. Although we believe the statistical analysis of historical returns has undeniable value, the forward-looking models based on yields and growth expectations should also be considered when setting future investment return assumptions. This is because the yields generally represent market expectations of future compound returns. Most investment and financial professionals use a combination of models to derive capital market assumptions (CMA) for broad asset classes that can be used in the forecasting of future asset returns. These models generally are either based on the statistical analysis of historical return series and/or on the forward-looking yield-and-growth type models. Both types of models use compound return measures, so the capital market assumptions are often presented in terms of geometric means. For forecasting or simulation purposes, the capital market assumptions are typically translated to arithmetic means under an assumption of a lognormal distribution for annual wealth.

A good reference for this topic and other related issues is listed below

D. Pachamanova, F. Fabozzi "Simulation and Optimization in Finance: modeling with MATLAB, @RISK, or VBA" (2010), John Wiley & Sons (2010)