On The Use and Misuse of Downside Risk

It is easy to understand and easy to calculate... incorrectly.

Frank A. Sortino is Director of the Pension Research Institute (see Biography)

Hal J. Forshey is Professor of Mathematics at San Francisco State University (94132).

The authors wish to thank Les Balzer of Lend Lease Corporate Services, Australia, and Dimitri Rastopolous of Merrill Lynch, for their helpful comments.

Downside Risk can provide additional insights to standard deviation when it is calculated correctly. This paper deals with errors that are introduced when it is calculated improperly. The most frequently used and least reliable procedure for calculating downside risk considers only those historical returns that fall below some minimal acceptable return. Errors are reduced by using simulation procedures to generate a discrete distribution of annual returns from monthly data. Simulation results are further improved by fitting a curve to the data that allows the distribution to be skewed. Integral calculus is then required to make the calculation.

Introduction

When the first article on downside risk was published by the staff at the Pension Research Institute [1980] there were no applications of downside risk in the investment community. Now, a growing number of practitioners are using downside risk in various portfolio management applications. We want to support this movement. However, based on our continued research on
the subject, we would like to call attention to some critical problems. Balzer [1994] and Harlow [1991] have examined various ways to conceptualize downside risk. Both concluded that deviations below some set value that must be earned at minimum in order to prevent bad outcomes is preferred to downside probability or shortfall risk. We concur with this finding and will focus here on the problems attendant to estimating and calculating downside deviation (DD) in asset allocation and performance measurement applications. For the remainder of this paper downside risk shall be synonymous with DD.

**The Basic Problem:**

Before we make an investment, we don't know what the outcome will be. After the investment is made and we want to measure performance, all we do know is what the outcome was, not what it could have been. To cope with this uncertainty, we assume that a reasonable estimate of the range of possible returns as well as the probabilities associated with those returns can be estimated. Some of these uncertain returns are undesirable and therefore crucial to the calculation of downside risk. The other uncertain returns are associated with reward, and we want to be able to make some intelligent decisions about the trade-offs between risk and reward for different investments we are considering. It stands to reason, that the more accurately we can describe this shape of uncertainty, the better will be our investment decisions. In statistical terms, the shape of uncertainty is called a probability distribution. Quantifying risk assumes an ability to obtain reasonable estimates of the location point of the distribution, which is usually the mean, and the dispersion about the mean, called variance. If the distributions are not symmetric, it would also help to have estimates of skewness and kurtosis in order to manage this uncertainty. Even better, computer graphics capabilities now make it possible for the decision maker to see the complete probability density function, which is far superior to a few summary statistics (see Exhibit 1). Some models allow the decision maker to change the shape in either a subjective or objective manner in order to ensure consistency in the decision making process.
We believe that distributions are fairly stable within economic scenarios [see Van Der Meer 1989], and reasonable estimates can be made of asymmetric shapes that will allow professional managers who apply them to outperform those who assume they are simply bell shaped. Some would argue that the shape of uncertainty is not stable; like a cloud, it keeps changing shape, and is therefore unknowable. Others might say it is difficult enough to estimate the mean and variance; attempting to estimate higher moments is pretentious or superfluous. To paraphrase Les Balzer, there are no laws of the universe that apply to financial markets, only our rudimentary models of how these markets behave. The theory on which we base our models does not attempt to describe a process that follows some immutable law of the universe. However, to the extent that models capture enough of what is really going on, users should be able to earn returns in excess of passive strategies or other active managers who choose not to use such models, without taking more risk. Past articles have offered evidence to support this claim [Sortino et. al. 1990, 1994]. We present one example here that has not been published before.

**Simulation Results**
How accurate do the estimates of asymmetric shapes have to be in order to beat a passive index or a model that assumes the shape of uncertainty is bell shaped? Sortino and Van Der Meer conducted a study at Shell Oil, Netherlands in 1988 to test the veracity of two asset allocation models. It was assumed that the investor was unable to forecast the risk and return characteristics of any asset (e.g., the mean and standard deviation each year for stocks, bonds, and T-bills). Instead, he made economic scenario forecasts (e.g., growth, stagnant, and recession) and accepted the historic statistics (first, second, and third moments) for these periods as inputs to the models. For example, if the forecast was for economic growth, the average return for stocks in all growth periods was used as the expected return on stocks. To further impair the investor's foresight, it was assumed he could only correctly forecast the first year's scenario and then held that position for three years. The investor used two optimizers in an attempt to beat a passive market mix. In Exhibit 2 (Optimizer Excess Returns) we plot the excess return that each optimizer earned over and above the return on a passive market mix (60% stock, 35% bonds, 5% T-bills) for an average risk-averse investor.

Exhibit 2

Because results are sensitive to the year that a study begins, we increment the beginning date each year of the study: the first two bars are for the period 1960 to 1988, and the second two bars are for the period 1961 to 1988. The mean-variance optimizer was only able to beat the market mix in five out of the 26 intervals. The downside risk optimizer outperformed the market mix in eighteen of those intervals. However, only in five intervals were the excess returns greater than one hundred basis points, and for the entire interval 1960 to 1988, the downside risk optimizer only earned an excess return of 3 basis points.
The point we wish to make is: perfect forecasts of the distributions are not required to do a little better than a passive mix or a mean-variance model, but GIGO is still the first law of computer modeling. The roll of models is to ensure that decisions are made in a manner that is consistent with the beliefs of the user. The beliefs have to do with the estimates of the shape and location of the distributions. There might only be a few people who are good enough at describing the uncertainty they are coping with to be worth the fees they charge. But those are the ones worth searching for.

Problems Problems:

We now focus on what could go wrong. What if the estimate for the location point is wrong? Returning to Exhibit 1, let's assume an investor must earn 8% in order to accomplish her goal. She wants to estimate the risk of falling below the minimal acceptable return (MAR) of 8% in order to make comparisons with other risky assets. Suppose the true location point for the mean is 16%, not 12% as Exhibit 1 indicates. All three shapes (A, B, and C) should shift to the right. Therefore, the current location so drastically overestimates the downside risk that the choice of shape becomes irrelevant. Similarly, if the true location for the mean is 8%, all three shapes drastically underestimate the downside risk.

Standard deviation does not suffer from this problem, it is independent of the location point but dependent on the shape. If the estimates for the mean and standard deviation are as shown for B and the true shape is C, standard deviation far over estimates risk, no matter where the true location point is. If the estimate is as shown for C and the true shape looks like B, standard deviation far under estimates risk. Active management implies an ability to manage the uncertainty associated with investments. If one cannot obtain a reasonable estimate of the location point, it is unlikely that an estimate of the shape will prove reliable. In which case, we see no justification for active management, let alone the use of quantitative techniques.

Now, let's suppose the location is correct and the true distribution looks like A in Exhibit 1. If we say it looks like B (bell shaped), we will far underestimate the amount of downside risk, but standard deviation is not affected. However, if we say it looks like B and the true shape is like C, we will far overestimate the amount of risk, whether we use downside risk or standard deviation. Again, to the extent that one can describe the shape of uncertainty (probability distribution) they are dealing with, they can successfully manage that uncertainty with quantitative tools. Tools that accommodate a wide range of shapes should produce better results than models that force the shape to be symmetric.
Discrete Versus Continuous Distributions

The final problem has to do with the technique for calculating downside risk if we are reasonably close on the location and the shape. Calculation error is due primarily to measuring only what did happen (discrete) instead of what could have happened (continuous).

Exhibit 3

Japan

Calculations are further flawed by the choice of differencing interval (e.g., annual instead of monthly). Exhibit 3 presents a crude picture of the shape of uncertainty based on annual data for the FT-Actuaries Japan index. The Exhibit indicates there were no down years for this index during this ten year interval. Should we conclude that there was no downside risk during this period of time? Of course not. First of all, it is ludicrous to think that ten observations of annual returns are adequate for describing the uncertainty associated with equity returns in Japan. Using a histogram of monthly data begins to unveil the shape hidden in the data (see Exhibit 4). The
The second flaw in calculating downside risk from a discrete sample is less obvious and will require some development before presenting an example.

Exhibit 4

**Historical Frequency**

Most practitioners we have observed calculate downside risk from a discrete distribution of monthly returns. That is, each time a return falls below the MAR it is squared, summed, and divided by the number of observations, before taking the square root. If the annual MAR was 10%, there were 45 observations below the monthly MAR of .8% for the Japan index. This method of calculating DD assumes that the worst return that occurred in the past is the worst that can occur in the future. A statistician would tell you that thirty-four years of monthly data are needed to produce statistically significant results in this manner, but it doesn't take a statistician to see the error in this logic.
To reduce this error, one should fit a curve to the monthly histogram of data to create a continuous distribution and then use integral calculus to calculate DD. The discrete method tells us only what did happen. The continuous method gives us insight into what could have happened. Suppose we assume that the distribution is bell shaped, and we therefore fit a normal curve to the data in Exhibit 4. We see the normal distribution does not fit asymmetric data well. The block of returns in the right tail are ignored because the normal curve ignores skewness. Also, the returns in the center go clear through the curve indicating a high degree of kurtosis, i.e., it is more pointy than a normal distribution. There are too few observations (blank spaces) below zero. This is referred to as the end point sensitivity problem and will cause the calculations from a discrete distribution to underestimate the amount of downside risk.

A curve that allows for a greater variety of shapes is the lognormal, but like its cousin the normal, it is a two parameter family of distribution. This means that two numbers, the mean and standard deviation, are all that is needed to determine a member of the family. For example, the standard lognormal is always skewed in a positive direction with a long tail. In this sense, the lognormal is just as restrictive as the normal. Yet, most of the prior work done in downside risk analysis assumes a two parameter distribution that is either normal or lognormal.

To make up for this deficiency, the lognormal family can be expanded by including equations that allow one to flip the distribution and slide it either side of zero. In this way we can incorporate negative returns and negative skewness. This results in a three parameter family, which can be taken to be the mean, the variance, and an extreme value. These three values will determine the skewness and kurtosis (third and fourth moments) of each member of the family. If the estimates of the parameters are good, fitting a three parameter lognormal distribution to the empirical data will be more accurate for calculating DD than using the normal or the standard lognormal. We will refer to this method of curve fitting as the 3-point method.

Another way to estimate the shape of the underlying distribution is to use simulation. But simply selecting monthly returns randomly and calculating the deviations below some monthly MAR is not optimal. What we need are annual returns from which to calculate DD. A better procedure, called the boot strap, can generate thousands of annual returns from ten years of monthly data (see Exhibit 5). The grey line fitting the data is a three parameter lognormal. This is not a perfect fit, but it is far better than the normal. It allows for skewness. It has very few blank spaces in the left tail, which would cause underestimation, and it allows for more pointiness (kurtosis) than the normal.
We now compare calculations of DD for the monthly data on Japan from 1980 to 1990 for a 10% MAR.
Using the discrete method in Exhibit 6, we get a monthly downside deviation of 2.74% as opposed to 3.2% with the 3 point curve method; indicating the discrete method calculation is 27% less than the 3 point curve method. If this monthly DD is annualized (see Exhibit 7) in the traditional manner, multiplying by the square root of 12, the annualized DD in discrete time (9.5%) becomes larger than fitting the 3 point curve to annualized data points (DD = 7.5%). The monthly DD in discrete time was smaller than the 3 point curve DD, but when annualized, the discrete DD is larger than the 3 point DD. What happened? It is incorrect to annualize downside deviations the way one annualizes standard deviations and can lead to gross exaggerations of the downside risk.

The boot strap method (Exhibit 5) indicated that it was possible for the Japanese stock market to decline 42.7% in one year and there was a 25% chance returns would be below the MAR. In fact, the loss was 39% in 1990, the year following the data period. Was the probability assigned to this outcome high enough? It is impossible to know the true underlying distribution. For a ten year holding period, the probability of a 42% loss is probably very small.

Data Errors
Another serious error is caused by using data that is not representative of the underlying distribution. The asymmetrical distribution (Blue) in Exhibit 9 was generated with five years of quarterly data ending June of 1994, using the bootstrap method. The symmetric distribution (Red) in Exhibit 9 has the same mean and standard deviation. If we assume the asymmetric distribution to be a more accurate description of the uncertainty associated with the hedge fund strategy of investing in distressed securities than a normal distribution, we come to the ludicrous conclusion that there is no chance of earning a return less than 7%, and the expected return from this strategy is 30%. The use of a symmetric distribution would improve the calculation of DD, but the probability of a negative return is still very small. Reason dictates that the location point should probably be shifted to the left and/or the 10th percentile lowered. A model that allows one to perform this task would add a great deal of credibility to the calculation of risk, whether it be DD or standard deviation.

The first thing one should do when an investment strategy or asset offers high return with little or no risk, is question the data. Let me preface any remarks with the acknowledgement that Van Hedge Fund Advisors Inc., who provided the hedge fund data for Exhibit 9, are just as concerned as we are that their data could be misused. To begin with, there are not enough observations to justify the use of the bootstrap. We consider five years of monthly data and 10 years of quarterly data to be sufficient. In addition, the data was generated as the average return from a pool of managers. Very possibly, the high returns of some managers offset the low returns of other managers in some quarters. An old saw says, "if you're in boiling water up to your waist and a block of ice from your waist up, on average the temperature is just right, but you'll be very uncomfortable." So should we all be very uncomfortable with averages. They can hide a lot of misery.

Next, question the methodology for generating the distribution. Yes, you can generate 2000 annual returns with the bootstrap. But, one experience repeated fifty times does not make one a person of great experience. A reasonable estimate of the true underlying distribution of the hedge fund strategy in Exhibit 8 might be generated by using the bootstrap with five years of monthly returns of each manager. But then again, if negative returns never happened, the bootstrap can't produce them. For inherently risky strategies that have had a run of luck, like borrowing money to buy inverse floaters in the hope that interest rates will continue to decline, other procedures are called for than discussed here.
If all else fails, question the veracity of using quantitative methods. Anyone who thinks they have found a way to get high returns with little or no risk will someday come face to face with the risk they didn't know how to calculate.

**The only correct measure of risk**

There isn't any universal risk measure for the many broad categories of risk. Within one category, standard deviation captures the risk of not achieving the mean; beta captures the risk of being in the stock market; DD captures the risk of not achieving the MAR necessary to accomplish some goal. They all provide useful information. None of them contains all the information necessary to manage risk in every situation. Additional categories might include: credit risk (e.g., counterparty default), market risk (e.g., lack of liquidity or currency devaluation), operations risk (e.g., human error or systems failure), and settlement risk (e.g., counterparty reneging on a contract). No single measure can capture all of these risks. We do not yet have all the pieces of the portfolio management puzzle. Even more reason to use all of the pieces we do have.
In Summary

- Estimation errors affect the validity of standard deviation and downside deviations for both symmetric and asymmetric distributions.
- The correct calculation of DD is more difficult than standard deviation.
- Fitting a continuous probability density function is superior to discrete sample calculations.
- Beware of averages. What you don't see can hurt you.
- No one risk measure is the be all and end all.

We believe DD is a valuable measure of risk that should receive broader use. The above comments were made in an effort to prevent its misuse, and/or, over-reliance on its completeness as a risk measure.

References


Kaplan, Paul D., and Lawrence B. Siegel, "Portfolio Theory is Alive and Well". Journal of Investing, Fall 1994, p 22.


Van Der Meer, Robert, and Frank Sortino, "Describing Uncertainty is the Key", Pensions and Investments, November 27, 1989.

Endnotes