Portfolio Optimization: The Robust Solution

Dan DiBartolomeo

Prudential Securities Quantitative Conference

21st December 1993

Investment practitioners who use mean-variance optimization techniques for portfolio construction are often disappointed in the results. As many users of such algorithms swear at them as swear by them. The most widely noted complaint is a lack of robustness in the optimization results. Small changes in informational inputs can have a seemingly overly dramatic impact on the resulting suggested portfolio. Most of these optimization systems use a "factor" representation of the variance-covariance matrix of expected security returns.

It must be considered that the process of optimization is merely that of finding a maximum of a particular mathematical function. Assuming that the algorithm in question is getting a mathematically correct result, the answer obtained must be the "right" solution. Why is it, then, that so many practitioners often find the results somewhat unintuitive at best, and ridiculous at worst?

There are four separate and distinct possible problems associated with usage of mean-variance optimization algorithms that are not well considered by practitioners. They are:

I) The expected returns on the investment assets may not be normally distributed, calling into question the validity of variance as a measure of dispersion of future returns.

II) The usual mean-variance utility function is not an adequate representation of the true preferences of the investor (or their agent).

III) Informational inputs to the algorithm do not adequately express the investor's beliefs about future events affecting asset returns and the probabilities of those events.

IV) The optimization algorithm takes the informational inputs that are forecasts of future events as parameters of known probability distributions, when, in reality, they are estimates of parameters of unknown probability distributions. We are simply ignoring estimation error.

While all four problems contribute to the difficulty of using optimization algorithms, we believe the last of the four is the most difficult around which to frame meaningful discussion.

I. The Non-Normality Problem

The first problem that is raised with MV optimization is that historic return distributions for some asset classes (and possibly individual securities within those classes) are skewed and leptokurdic thereby calling into question the basic validity of mean-variance techniques that are parametric statistics implying a normal distribution of the variable.

The long term return time series for most asset classes are unimodal leptokurdic. Given the mean and standard deviation of the series, there are more points than expected near the mean and more points than expected in the extremes of the tails. This implies that there are fewer than expected events of intermediate magnitudes. This fact has been used by critics of MV techniques to argue that asset class returns are not normal and therefore are not appropriately analyzed by MV methods. We would argue that such kurtosis only shows itself significantly when looking at periods far longer than the practical time horizon of almost all investors (say 30 years). We would argue that the kurtodic appearance of long series is due to the fact that over any shorter period, the distribution is roughly normal, but that the variance of the successive short term distributions is not fixed. Example: If I have a five year period with a normal distribution of returns with a low variance, followed by another five year period of normal distribution with high variance, the combined ten year period will be unimodal leptokurdic. This is easily illustrated by overlaying two transparencies, each with a normal curve of different width (variance). The resultant combined outline will have the observed shape. (Rosenberg, "The Behavior of Random Variables with Nonstationary Variances and the Distribution of Securities Prices", working paper, University of California at Berkeley, 1975) The current popularity of GARCH models (Generalized AutoRegressive Conditional Heteroschedasticity) is an indication of the wide acceptance of this view of asset return distributions.

The second attack on the normal distribution is that return series are often very skewed, again raising the issue of the validity of MV techniques. While skewness is indeed significantly observed over long time periods, such as the last 30 years of the S&P 500, this ex-post information is not, by itself, sufficient for the ex-ante forecasts needed in asset class allocation decisions. If we take the monthly return series of the S&P 500 for the thirty years ending May 1991 (diBartolomeo, "Estimation Error in Asset Allocation", Northfield Conference, Boston, 1991) and break it into ten three year segments to reflect a realistic investor time horizon, we will observe that the skewness values for the ten sub-periods are unstable and random. There are four changes of sign as compared to an expected value for changes of sign of 4.5 (nine period to period junctions, with a 50/50 chance of sign change at each junction). When considering this same issue for individual equity securities, we are aware of no evidence that historic skewness values have been shown to exhibit significant persistence.

An additional argument against the skewness issue being generally meaningful is as follows: If the distribution of returns is not normally distributed, then prices of assets cannot be log-normally distributed. While there is some discussion of skewness of returns, little in the theoretical literature suggests any successful challenge to the concept of log-normally distributed prices (for background, Alexander, "Price Movements in Speculative Markets": Trends or Random Walks", Cootner, ed., *The Random Character of Stock Market Prices*, MIT Press, 1964).

One situation where the skewness argument does hold against traditional mean-variance optimization is when a portfolio skewness can be meaningfully forecast, such as portfolios encompassing stock options, and other derivative instruments with truncated pay-off patterns.

II. Inadequacy of the Mean-Variance Utility Function

Most optimization algorithms operate on a conventional mean-variance utility function, expressed as follows:

$$U = E - (S^2/T) - C$$

where U = the investor's expected satisfaction with the portfolio
E = the expected return of the portfolio relative to some benchmark portfolio
S = the expected standard deviation of returns relative to the benchmark portfolio
T = a positive scalar expressing investor tolerance for risk
C = transaction costs associated with taking the portfolio from its present security mix to whatever mix of securities which will maximize U

The most obvious attack on mean-variance techniques is that using variance (S^2) as a measure of return dispersion implies that investors are equally averse to unexpectedly good (positive) portfolio returns as they are averse to unexpectedly poor (negative) returns. This argument is irrelevant if the likelihood of unexpectedly good outcomes and the the likelihood of expected bad outcomes are unequal. If they are equal, then the "downside" dispersion is merely one-half the total dispersion and this constant proportion will be reflected in investor's choices of the value of T, the risk tolerance parameter. Only if the distribution of future returns is expected to be skewed (as discussed earlier) does this aspect of the problem merit concern. Note that a lack of skewness does not imply that the distributions must be normal, merely that they are symmetric.

The other possible problem associated with the techniques is the inability of investor's to establish an appropriate value for T, the risk tolerance parameter. For the purpose of equity portfolio optimization, the choice is theoretically obvious. When we select a benchmark against which we are measuring our portfolio performance, we are implicitly stating our preference for the value of T, which is embodied in the presumed return and dispersion characteristics of the benchmark. In practice, an agency risk problem arises in the selection of a value for T; because of the business risk of being fired for poor performance, hired investment managers and fund sponsors are often much more conservative in their selection of T, than would be economically dictated by the benchmark index that they, themselves, have selected.

III. Failure to Correctly Express Expectations about Optimization Inputs

Optimization algorithms require input information on which to operate. There are two common ways in which practitioners fail to properly express their beliefs. The first error is to fail to provide complete information. For example, many firms have investment decision processes which "screen" for attractive stocks but do not provide explicit return forecasts for each of the stocks passing the screen. In essence the algorithm takes all portfolio candidate stocks as having equal expected returns. Even if this is what we actually believe, a less obvious problem arises. In holding only certain securities in a portfolio, we will have negative active weights on those securities that are constituents of the benchmark but are not included in our portfolio. However, the magnitudes of those effective "short" positions are an accident of the weights that each security represents in the benchmark portfolio. The extent to which our portfolio is different from the benchmark is not optimal unless all of the securities in the benchmark but not in the portfolio just happen to have expected returns and marginal contributions to expected variance that are combinatorially proportionate to the negative active weights. This is, at best, a wildly heroic assumption.

This problem is easily dealt with by changing the paradigm in which the portfolio optimization problem is couched. Practitioners generally think of an optimization problem as involving two portfolios, the investor portfolio and the benchmark portfolio. A better construct involves three portfolios, the investor portfolio, the benchmark portfolio, and the "difference" portfolio. The difference portfolio is that set of security holdings which when added to the investment portfolio give us the benchmark portfolio. It is really this difference portfolio that we are optimizing. The final (and optimal) investor portfolio is merely the portfolio that results from subtracting the optimal difference portfolio from the benchmark portfolio. Once we have focused our attention on the difference portfolio it becomes obvious how much necessary input information is overlooked by typical practitioners.

A second failure of the information inputs is to not express them in maximum likelihood magnitudes. For example, let's assume we have a security valuation model that produces explicit return forecasts and that for a particular stock the current is 20% return per annum in excess of the average return per annum for other stocks. Let us also assume that this valuation methodology has produced forecasts in the past that have been 8% correlated with subsequent observed performance. If we use the 20% return forecast as our informational input, we are leaving out the very critical piece of information regarding the probable reliability of this value. As a "quick and dirty" answer we could use 1.6% expected excess return. This would represent an 8% probability of our forecast being correct and a 92% probability of it being incorrect, in which case we would assume that this stock is no different from the average of all other stocks, i.e., an expected excess return of zero. Phillipe Jorion (Bayes-Stein Estimation for Portfolio Analysis, *JFQA*, Sept. 1986) and Richard Michaud (The Markowitz Optimization Enigma: Is Optimized Optimal?, *FAJ*, Jan. 1989) have both published on how to improve optimization results by preprocessing input forecasts with statistical methods such as Bayesian adjustment techniques.

Most practitioners using optimizers create their own return forecasts while the risk forecasting aspect of the problem is left to a risk model provided by the vendor of the optimizer. It is common practice to intentionally use an unrealistically low value for T, the risk tolerance parameter, as a way of expressing a lack of confidence in our return forecasts, as opposed to a more formal Bayesian statistical approach. Unfortunately, this technique not only impacts relative magnitudes of the return and risk terms in the utility function, but also impacts the relative magnitude of the risk and transaction cost terms. As such, it can lead to unrealistic solutions for the optimal portfolio.

IV. Failure to Include Estimation Error

As simple mathematical algorithms, optimizers take the input information as "certain" as opposed to being subject to error in estimate. While the Bayesian methods considerably improve reasonableness of the results by contributing to the reasonableness of the inputs, they do not represent an explicit process to include estimation risk in the actual optimization process.

In the example in the section above, we obtained a 1.6% expected excess return forecast as a reasonable value to use as an informational input. It is still, however, really an estimate of the long-term mean of the probability distribution of future excess returns for that security. While the expected standard deviation of returns is also an input to the optimization process, the expected standard deviation expresses how each distinct future time period (as a single draw from the probability distribution) will likely vary from other future time periods. It in no way expresses the fact that our estimate of future long-term mean may simply be wrong. Or that our values for the future standard deviations of security returns may be wrong. Or that our estimates of correlation among future securities returns may be wrong.

Optimization algorithms as conventionally used are systematically overconfident because they do not explicitly consider estimation error. The risk estimates provided are a downward biased estimate of the overall uncertainty inherent in the problem. The lack of robustness in optimization results arises from this overconfidence problem. The algorithm thinks it that all inputs are fully reliable and hence it should act on even the most minor change.

Our firm utilizes a process suggested by Bey, Burgess and Cook (Measurement of Estimation Risk in Markowitz Portfolios, University of Tulsa, unpublished) in 1990. The first step in the method is to attempt to quantify the potential estimation error in each of the inputs to the optimization problem. One approach to quantifying the magnitudes of potential errors in estimate is to use resampling methods such as "bootstrapping" on return and variance elements of historic asset (or factor) returns and "jacknifing" on the elements of the correlation matrix among asset (or factor) returns.

Another possible approach to defining the magnitude of potential estimation errors is to consider the level of disagreement among market participants. One could consider the dispersion of analyst's estimates of future company earnings and earnings growth rates. For securities that have listed stock options, one could compare the implied volatilities of the various option contracts for the dispersion of return volatility estimates. Another consideration could be the level of short interest as an indication of violent disagreement among market participants regarding the likely returns from a particular stock.

We begin the actual optimization process by calculating a traditional efficient frontier and selecting specific points along the frontier that correspond to specific values of the risk tolerance coefficient. We next repeat the process of calculating the efficient frontier 1000 times, each time randomly perturbing the input information by magnitudes derived from the resampling or substituted processes. By observing the same points along each frontier, we have 1001 points for each value of the risk tolerance coefficient that we had selected along the original frontier. In essence, each known point along the efficient frontier is now represented by a region consisting of 1001 related points (related by having the same risk tolerance coefficient). We then calculate the Euclidean distance between each of the additional 1000 points and the original point. To establish a confidence interval, we then eliminate the 10% of the 1000 points that are most distant from the original optimal portfolio. The remaining 900 points represent a region in mean-variance space to which we are indifferent. Given the formally admitted uncertainty of the inputs, we are indifferent from a risk/return standpoint to holding any portfolio that falls within the region. We can also calculate composition of a portfolio that falls at the density center of the region; the portfolio composition least likely to be driven outside the region by estimation error in the inputs.

The empirical result of such work has been very intuitively appealing. In short, the "density center" portfolios tend to be close to the original optimal portfolios but slightly biased toward equal weighting of the assets. Consider a spectrum with traditional optimization at one extreme, where all input information is considered certain. At the other extreme, all input information is devoid of any meaningful predictive content. In the latter case, where no information is meaningful, all assets are fungible and an equal weighted portfolio will result (or equal active weighted relative to a benchmark). By formally introducing estimation error into the process, we change the portfolio composition toward equal weighting.

This last observation has an important implication for factor models used to predict portfolio risk. One of the traditional measurement criteria of any factor model is its ex-post explanatory power. To the extent possible, we would like to have a model that explains a great deal, leaving very little asset return variation to be considered as unexplained residuals. As far as statistics go this is perfectly sensible, as long as the model is unbiased.

In comparing two similar factor models with different levels of explanatory power, the natural first tendency is to prefer the more powerful model. However, once we have implemented a factor model as part of an optimization algorithm, we need to address the estimation error problem. If we limit ourselves to traditional optimization, the weaker model will tend to produce "optimal" portfolios that also tend to be biased toward equal weighting relative to the portfolios produced by a similar but "better" risk model. In the weaker model, a greater portion of the expected variance is unexplained residual. The weaker model has a proportional balance between explained and unexplained variance that is more similar to the better model with estimation error explicitly included. The weaker model is systematically less overconfident than the model of higher explanatory power and produces portfolios more like the "density center" portfolio. For the purpose of the defining similarness among portfolios we use the J-statistic. The J-statistic is defined as:

$$J = 1 - ((\Sigma_{i=1 \text{ to } n} (abs(W_p - W_b)))/2)$$

where W_p = decimal weights of securities in portfolio
W_b = decimal weights of securities in benchmark
n = the number of securities in the union of the portfolio and benchmark

Two other implications arise out of the explicit consideration of estimation error. To the extent that we are indifferent to a range of portfolios given any particular set of inputs, we should not expend transaction costs to rebalance a portfolio unless we have moved out of the region within which we are indifferent. The other implication is that the substantial computational complexity added to optimization problems if we seek only solutions with round lot position sizes is unlikely to be worth it. The process of rounding off positions is not apt to move our portfolios out of the region of indifference.

There are also many efforts to include estimation risks as part of the objective function in a wide variety of linear and quadratic programming problems. An excellent overview of the industrial applications of these techniques appeared in *SIAM News*, November 1993.

Conclusions

Explicit inclusion of estimation error is a necessary but rarely practiced step to obtaining optimization results which are mathematically correct, intuitively appealing and robust. Lesser problems relate to the correct expression of input information and the selection of realistic risk tolerance parameters. While the theoretical attacks on mean-variance optimization are meaningful, they are important in only a minority of circumstances.