Implementation of Equity Return Forecasting Methods

(everything you wanted to know about alpha but were afraid to ask)

Dan diBartolomeo

December 20, 1998

Introduction

The most difficult aspect of modern portfolio management is the forecasting of expected returns for the various securities in which one may invest. The need to perform this difficult task arises from the idea that investors will be most satisfied by maximizing the expected risk-adjusted returns on their portfolios, net of related costs. Stated formally, this is the familiar mean-variance utility function of Levy and Markowitz (1979).

$$E[\text{Utility}] = E[\text{Return}] - E[\text{Risk}]^2 / T - E[\text{Cost}] \qquad (1)$$

Where

E[ ] is the expectation operator
Utility is the utility of the portfolio for the investor
Return is the return on the portfolio per unit time
Risk is the standard deviation of the return per unit time of the portfolio
Cost is cost per unit time of owning the portfolio such as transaction costs, taxes and fees
T is the investor's marginal rate of substitution between risk and return

In the case of a portfolio within a single asset class, it is customary to recognize the investor's opportunity to use inexpensive passive management such as index funds. We can transform our utility equation to a relative form where returns and risk are measured relative to some benchmark index

$$E[u] = E[a] - E[s]^2 / t - E[c] \qquad (2)$$

Where

u is the utility of the portfolio relative to the benchmark index
a is the relative return of the portfolio per unit time often called "alpha"
s is the standard deviation of relative returns per unit time often called "tracking error"
c is the cost per unit time of owning the portfolio less the cost of owning the index
t is the investor's marginal rate of substitution of relative risk and relative return

For our discussion, alpha will be the annual rate of return on a stock, less the annual rate of return of the benchmark index. The alpha of a portfolio is a simply a weighted average of the alphas of the individual stocks in the portfolio. There are many strategies that are employed in equity portfolio management in order to forecast the benchmark relative returns or "alpha" of different stocks. The focus of this paper is efficient ways to implement methods for return forecasting. It should be noted here that our use of the term alpha is strictly limited to indicate returns relative to a benchmark, rather than to indicate returns that are relative to consensus investor expectations set by the Capital Asset Pricing Model (Jensen, 1969). The reader is referred to diBartolomeo (1997) for discussion of the risk aspects of the problem.

Forecasting Stock Alpha

In forecasting security returns, we can take one of two broad approaches. The first is to directly forecast the expected return using some form of dividend discount model (see Farrell, 1985). Once we have an expected return for each security, we can compute the expected return of the benchmark index. To obtain the relative expected return (alpha) of each stock, we simply subtract the expected return on the benchmark index from the expected return of each stock.

The second approach is to find some observable characteristic of a stock such as Price/Earnings ratio and use statistical methods to infer the expected relative return. The inference takes the form of a mapping between the distribution of the characteristic and the expected future distribution of stock returns. It is widely assumed in finance that stock returns are normally distributed (Richardson and Smith, 1993). This second approach is often called the "Alpha Forecasting Rule of Thumb" (Grinold and Kahn, 1995). The general form is:

$$\text{Alpha} = E[IC] * (\text{Score} * E[\text{Volatility}]) \qquad (3)$$

IC is the correlation between the investor's forecasts and subsequent returns often called the information coefficient. The Score is the standardized (z-score) form of the observable stock characteristic. Volatility is the standard deviation of the return distribution over which the investor is trying to make predictions.

An example may make equation (3) intuitive. Let's assume that both our characteristic and the stock returns are normally distributed. We can also make the heroic assumption that our characteristic is *perfectly* predictive of future stock returns. In such, a case there is a direct one-to-one correspondence between points in the characteristic distribution and the related point in the distribution of future returns. A stock that is two standard deviations above the mean in the characteristic will have a future return that is two standard deviations above the mean of the return distribution. Hence, the product of the Score times the Volatility is equal to alpha for IC = 1.

For a moment, let us consider the opposite situation where our characteristic has *no ability* to predict future stock returns. In such a case, we would have to assume that the expected return of each stock was no different than the expected return of any other stock. Hence, the expected return of each stock would be the same, and the expectation of alpha (relative return) would be zero for the case IC = 0.

In essence, the information coefficient (IC) is used to scale the perfect prediction expected return (Score * Volatility) to reflect the reliability of the forecasting characteristic. This is an informal application of Bayes' Theorem in statistics (for discussion, see Mosteller, Rourke and Thomas). There are three popular ways to estimate the correlation between two sets of data. The first is the traditional Pearson correlation coefficient (arises from a standard ordinary least squares regression) that gives the actual correlation between the two data sets. The second is the Spearman rank correlation coefficient that approximates the correlation as the correlation of the rank positions of the related elements of the two data sets. The third is another form of rank correlation called Kendall's Tau coefficient (for discussion see Gibbons, 1971). Of these three methods, Spearman rank correlation is most often used in practice. It is considered more robust (less affected by data outliers) than the Pearson method but is much easier to calculate than the Kendall Tau. The issue of robustness of regression results in stock return studies is discussed in Knez and Ready (1997). They found dropping even 1% of the data sample negated the conclusions of some well-known studies, such as Fama and French (1992). Note that the information coefficient can also be applied to forecasts of relative returns arising from a dividend discount model.

The "Rule of Thumb" can be looked at in two different ways. The first method would be to assume that we wish to predict the future relative returns between securities directly by analyzing the cross-section of stock returns. In such a case, we get

$$\text{Alpha}_{it} = E[\text{IC}_t] * (\text{Score}_{it} * E[\text{Volatility}_t]) \qquad (4)$$

Where

Alpha_{it} is the expected relative return of stock i at time t
IC_t is the information coefficient across the distribution of relative stock returns at time t
Score_{it} is the cross-sectionally standardized characteristic of stock i at time t
Volatility_t is the standard deviation of the cross-section of future stock returns

In this method, we are forecasting where each stock will end up in the future distribution of returns based on where it is located in the present distribution of the characteristic (Score_{it}). Our term is a measure of the expected dispersion across the distribution of returns in the future. Since we have only one distribution of future returns with dispersion equal to Volatility_t, we have only one value for the information coefficient IC_t.

A second, somewhat more elaborate approach is described in Grinold and Kahn. In this approach, we consider the process of forecasting one stock at a time.

$$\text{Alpha}_{jt} = E[\text{IC}_{jt}] * (\text{Score}_{jt} * E[\text{Volatility}_{jt}]) \qquad (5)$$

Where

$Alpha_{jt}$ is the expected relative return of stock j at time t
IC_{jt} is the information coefficient for the distribution of returns of stock j at time t
$Score_{it}$ is the time series standardized characteristic of stock j at time t
$Volatility_{jt}$ is the standard deviation of the benchmark relative returns of stock j at time t

In this method, we standardize our characteristic within a time series of observations for the characteristic value for stock j (e.g., where is the P/E of stock j at time t, relative to the time series of the P/E value for stock j). We find the information coefficient by computing the correlation of the time-series standardized score against the time-series distribution of the observed benchmark relative returns for stock j. The volatility term is the dispersion of the relative returns for stock j.

Using the method of equation (4) we are obtaining forecast alphas directly for each stock based on their relative exposure to the predictive characteristic we have chosen. Using the method in equation (5), we are first determining whether expected returns at time t are favorable or unfavorable for each stock relative to its own history, and then making a comparison between the stocks.

In theory, the time series method (5) should produce better results because it takes stock specific information about volatility and information coefficient into account. Unfortunately, obtaining reasonable estimates of stock specific information coefficients requires long time series of data history that may be unrepresentative given the high rate of new information arrival in financial markets. A common error among practitioners is to use the cross-sectional scores and information coefficients of equation (4) with the stock specific volatility of equation (5). This error can lead to alpha estimates which are biased in favor of high volatility securities because the high stock specific volatilities will lead to higher magnitude alphas, if we do not take into account that higher volatility stocks are usually harder to predict (i.e., lower ICs). Another common error is to use cross-sectional variables such as market capitalization that are not even close to normally distributed. In such cases, a transformation of the original characteristic (e.g. logarithm of market capitalization) may be normally distributed.

Among practitioners it is common to use an alpha prediction model that is the combination of more than one separate alpha prediction models. For example, we may have an alpha prediction model that uses both P/E and Price/Cash Flow ratios. These ratios are highly correlated and their respective information coefficients arise from largely the same information. Contrast this with another case where have one model based on some balance sheet characteristic, while having another model that scores stocks on an analysis of their price trend. In the first case, the information coefficients for each model must be adjusted to reflect the extent that the models are correlated in their forecasts. The information coefficients of uncorrelated factors are additive in their squares.

$$IC_m^{\,2} = SUM_{j\,=\,1\text{ to }n}\,[IC_j^{\,2}] \qquad (5a)$$

Where

IC_m is the information coefficient of the combined model
IC_j is the information coefficient of the jth model
n is the number of models

For correlated factors, we must adjust the information coefficients to reflect the extent to which each model "overlaps" the other models. This will allow us to judge the independent predictive strength of a each model. Details of the adjustment are presented in Grinold and Kahn.

An Alternative Approach

A common alternative approach to the use of information coefficients is to use fractile analysis. For example, we could break a set of stocks into quintiles by their rank on some characteristic and observe the differences between quintile average subsequent returns. Under certain assumptions about the uniformity of the relationship, fractile analysis will lead to similar results to the methods described herein (Buckley, 1996). However, if those uniformity conditions do not hold, fractile methods are often not robust.

Let us assume a universe of 1000 stocks that we break into quintiles of 200 stocks by some characteristic. We do this for each month during a period of time. We observe that in the month subsequent to each revision in the rankings, the top quintile stocks have higher average returns than the lowest quintile stocks. If the time-series distribution of the spread in returns between the quintiles was statistically significant, we might believe that our chosen characteristic is a reliable predictor of stock returns.

However, it might be that the consistently good returns of the top quintile are due to just one or two stocks producing extremely high returns, or that the consistently inferior returns of the lowest quintile are due to just a couple stocks producing extremely poor returns. If quintile membership changes over time and trading costs are sufficiently high, investors will prefer to hold a sample of the top quintile rather than the entire membership of the top quintile. If the fractile return spreads arise from just a few members, sampling methods are unlikely to capture the apparent benefit of the characteristic. One simple step that can be taken to reduce the danger of this problem is to compare median, rather than mean, returns within the fractile groups.

A Real Life Example: OWNRELVAL

Since April 1991, our firm has computed approximately fifty characteristic variables that could be used for alpha forecasting at the end of each month. One of these variables is called OWNRELVAL.

$$ORV_{it} = ((E_{it} / R_t)/ P_{it}) \ / (AVG_{T = t-1 \text{ to } t-5}((E_{iT}/ R_T) / P_{iT})) \qquad (6)$$

Where

ORV_{it} is the OWNRELVAL for stock i at time t
P_{it} is the price of stock i at time t
E_{it} is the earnings per share of stock i for the twelve months ending at time t
R_t is the decimal form risk free interest rate at time t
AVG is the average operator

The logic of the model is simple. Let us assume that the future earnings per share of a company will be constant in the future. In such a case, since there is no variability in the earnings stream, it is appropriate to discount the future earnings to present value using the risk-free rate. In such a case, the intrinsic value of a stock is simply the earnings per share divided by the risk-free rate. The model operates by taking the ratio of the current intrinsic value to current price and comparing that ratio to the "normal" (five year past average) of the same ratio.

Let's do a numerical example of making an alpha forecast with OWNRELVAL. In our hypothetical case, we are only interested in stocks within the Standard and Poors 500 and we anticipate holding each stock an average of six months. As of November 1997, our database had information for seventy-four semi-annual periods (monthly overlapping, going back to 1991) for which the values of OWNRELVAL were computed at the beginning of each period and the stock returns observed during the semi-annual period.

Using just the sample of stocks in the Standard & Poors 500, the times-series average of the cross-sectional information coefficients was .046 (T-statistic of 3.76, corrected for overlapping observations). During these same periods, the average cross-sectional volatility of the Standard & Poors 500 stocks was approximately 25% per year. If we assume that the future will be similar to the past, the expected values of IC and Volatility are the averages of the respective past values.

$$Alpha_{it} = [.046] * (Score_{it} * 25) \qquad (7)$$

To examine the alpha for particular stocks we merely substitute the appropriate score values. Using the S&P 500 as the entire stock universe, $Score_{it}$ values at the end of November 1997 for IBM, Unocal and Cummins were .063, -2.510 and .475, respectively. Hence the expected relative return (alpha, % per annum) is .07 for IBM, -2.89 for Unocal and .55 for Cummins. These alphas are relative to the expected mean return of the S&P 500. We would also account for the fact that the S&P 500 when used as a benchmark index is capitalization-weighted. We would compute the alpha of the S&P 500 as the capitalization weighted average of the alphas of the member stocks. This value would then be subtracted from each stock alpha so that the alpha of the S&P 500 index was exactly zero by mathematical construction.

Turnover Considerations

In the hypothetical example above, we computed our information coefficients based on the correlation between forecasts and subsequent six-month returns. Obviously, we could have computed the information coefficients using any particular time horizon we might choose. Obviously, it is of little value to estimate information coefficients over short periods likes a week or a month, if we are constrained by trading costs to trade stocks infrequently, with average holding periods of years. Conversely we deprive ourselves of profitable opportunities if we do not act on reliable short-term forecasts if we have them.

Let's start with the relation between the risk/reward ratio of active management (IR) and the information content (IC) of one's forecasting models.

$$IR_{available} = IC * BR^{.5} \qquad (8)$$

Equation (8) is referred to as the “Fundamental Law of Active Management” in Grinold(1989). We'll define BR as the "breadth" of one's strategy. For the moment, let us consider this the number of independent buy/sell decisions one makes. Hopefully this makes intuitive sense: the risk/reward tradeoff one can achieve is linearly related to the effectiveness of one's forecasts and related to the square root of the number of opportunities one gets to implement a forecast. Let’s also define:

$IR_{desired}$ is $alpha_{gross}$ / tracking error

BR is the number of stock ranking decisions per year

For active management to be successful in adding value over a passive strategy,

$$IR_{available} > IR_{desired} \qquad (9)$$

For this example, let’s assume that we are an active manager with the following goals with respect to our active management strategies:

Desired Excess Return (Alpha) Net of Costs: 2% annually

Acceptable Tracking Error	3.5% to 4% annually
Acceptable Turnover	60% to 100% annually
Size of Coverage Universe	1400 stocks
One way transaction costs	1%
Portfolio construction	40 stocks, equally weighted

Let's first go through the exercise, assuming that transaction costs are zero, short-selling is allowed and there is no constraint on turnover. For tracking error and turnover we will use the midpoints of the ranges provided. We assume that IC = .04, close to our previous example. Then $IR_{desired}$ = 2 / 3.75 = .533.

To determine BR, we assume a 1400 stock universe with a model updated monthly. Since the components of IR are in annual terms, we'll use 1400 * 12 for BR. That is, we have 1400 stocks to pick from and we get to revise our picks 12 times per year. Therefore BR is 16800. The square root of 16800 is 130. As such, the left-hand side of our basic equation is $IR_{desired}$ = .533. The right hand side of $IR_{available}$ is 5.2 (equal to .04 * 130).

Now we'll look at alpha considering transaction costs. We can express the required gross alpha to get our desired net alpha as:

$$IR = alpha_{gross} / \text{tracking error} = (alpha_{net} + 2 * \text{turnover} * \text{transaction cost}) / \text{tracking error} \quad (10)$$

Assuming the midpoint turnover of 80%, alpha(gross) is 3.6 (2 + 2 * .8 * 1). So now our $IR_{desired}$ has moved from .533 to .96.

Our next consideration is what happens to BR. In the real world, we have limited short-selling. In institutional portfolios we often cannot actually do a short-sale, but we are being measured against a benchmark index (against which we can underweight). This is like a short-sale on a relative basis. As we are measuring stock weights relative to benchmark, the sum of the active weights must add to zero. However, there is a limit on the magnitude of each of the relative short-sales (we can't go below zero absolute weight). Since most stocks in the universe are only a tiny portion of total market capitalization, our ability to do a "relative short-sale" is very small. In effect, the short-sale portfolio must be an extremely diversified portfolio of very small positions. It is therefore unlikely to add much relative return to the portfolio, as the Central Limit Theorem guarantees that the sample mean of a very large sample will approach the population mean.

We might therefore assume for the purpose of alpha generation that our universe of 1400 stocks has been cut substantially since we cannot take meaningfully sized negative positions in stocks we believe will underperform. One extreme approximation would be to assume that our universe of stocks has been cut in half to 700. A more exact calculation would also take the benchmark into account as we can only do a "relative short sale" on those stocks that are in the benchmark, irrespective of whether they are in our coverage universe.

We also have the constraint of reasonable turnover. While we can produce stock rankings every month (or every day), we certainly could not implement them if they required 1200% turnover, due to excessive trading costs. BR is limited by the number of buy/sell decisions that we are willing to implement per unit of time. If our portfolio turned over once a year, we could make 700 (no short sales) such decisions. At 80% annual turnover, we can make 700 * .8 or 560 decisions per year. Now lets look at the numbers again. So BR is now 560 and the square root of BR is about 24.

Our $IR_{desired}$ is now .96 and our $IR_{available}$ is now (.04 * 24) or .96. With some realism thrown in, our hypothetical model is exactly on the borderline of adequacy to produce the desired risk/reward tradeoff.

Another thing to think about is the construction of the model itself and its relationship to BR. For example, if we use a model that predicts industries rather than stocks, the decisions to buy or sell each of the stocks in a given industry are no longer independent. As such, the "breadth" of the strategy is reduced. Our estimate of BR should reflect only the independent decisions we make. For numerical examples, see Buckley.

It should be noted that the number of stocks in the portfolio and the weighting scheme did not enter into the calculations. These items are important only to the extent that a portfolio with desired tracking error can be constructed in this fashion from your universe. There is always the possibility that no portfolio exists that can meet both the desire for expected alpha and the desire to control expected tracking error within reasonable bounds. Assuming a no short-sale constraint, the highest expected alpha can always be obtained from a one stock portfolio.

Estimating the reasonable range for portfolio turnover targets is complex. As turnover increases, the investor’s costs increase and hence the required IR, but the number of opportunities to implement the predictions you make is also increased. Which effect is stronger depends on the other aspects of the problem. Another view of this problem is presented in Grinold and Stuckelman(1993).

Conclusions

The most important aspect of modern equity portfolio management is the forecasting of alpha, or relative return among stocks. It is also the most difficult aspect. Relative return forecasts may be generated directly from a model, or based on a statistical mapping of a characteristic variable onto the expected distribution of relative returns. In either case, information coefficients should be used to adjust alpha forecasts commensurately with the reliability of the forecast. Care should be taken that the level of portfolio turnover implicit in the model building process is in keeping with realistic consideration of trading costs.

References

Buckley, Oliver. "Information Coefficients and Linear Factor Models," *Proceedings of the Northfield Information Services Annual Investment Conference* (White Sulphur Springs, 1996)

diBartolomeo, Dan. "Risk of Equity Securities and Portfolios," *AIMR Equity Specialization Program Readings*, 1997, v2, 3-15.

Fama, Eugene and Kenneth French, "The Cross-Section of Expected Stock Returns," *Journal of Finance*, 1992, v47(2), 427-466.

Farrell, James. "The Dividend Discount Model: A Primer," Financial Analyst Journal, 1985, v41(6), 16-19, 22-25.

Gibbons, Jean. *Nonparametric Statistical Inference* (New York, McGraw-Hill, 1971). 226-240.

Grinold, Richard. "The Fundamental Law of Active Management," *Journal of Portfolio Management*, 1989, v15(3), 30-37.

Grinold, Richard and Ronald Kahn. *Active Portfolio Management* (New York, McGraw-Hill, 1995). 221-224.

Grinold, Richard and Mark Stuckelman. "The Value-Added / Turnover Frontier," *Journal of Portfolio Management*, 1993, v19(3), 8-17.

Jensen, Michael. "Risk, The Pricing of Capital Assets, and the Evaluation of Investment Portfolios," *Journal of Business*, 1969, v42(2), 167-247.

Knez, Peter and Mark Ready. "On the Robustness of Size and Book-to-Market in Cross-Sectional Regressions," *Journal of Finance*, 1997. V52(4), 1355-1382.

Levy, H. and Harry M. Markowitz. "Approximating Expected Utility By a Function of the Mean and Variance," *American Economic Review*, 1979, v69(3), 308-317.

Mosteller, F., Rourke R. and G. Thomas. *Probability with Statistical Applications*, Second Edition (Reading, MA, Addison Wesley, 1970). 158-160.

Richardson, Matthew and Tom Smith. "A Test for Multivariate Normality in Stock Returns," *Journal of Business*, 1993, v66(2), 295-321