NORTHFIELD | modeling financial markets worldwide

Evaluating Active Manager Skill

Dan diBartolomeo

Webinar, October 2014

Outline for Today

- We will discuss three different methods for evaluating active manager skill
- The first is the PWER method
  - Bolster, diBartolomeo and Warrick (2006)
  - Northfield evaluates all global mutual funds and a large hedge fund universe on a monthly basis, over 100,000 financial products in all
- The second approach will be EIC analysis
  - diBartolomeo (2008)
  - Uses position level data to get rapid statistical significance
- The third method will be to evaluate the skew in the vector product of active bets and active returns

Motivation for New Methods

- No value added
  - Despite internal staff and consultants, many plan sponsors found their average manager performance was average
  - Plan sponsors are not really compelled that their managers were skillful rather than lucky in the past
  - Potential skilled but unconventional managers were excluded because they were hard to categorize or benchmark
- The lack of discipline in the process makes allocating capital between active and passive, and among active managers extremely subjective

Assumptions

- To hire external active managers we must believe at least one of three things:
  - The average professional investment manager outperforms passive index funds because individual investors have below index performance
  - **Manager active returns have persistent patterns. We can predict with reasonable reliability which managers are going to outperform in the future, even if the average manager is just average**
  - We are doing a societal good because if all investors were passive, there would be no functional mechanism to ration capital in the economy. Our economy would break down over time

What We're After

- There are numerous performance metrics used as proxies for manager skill such as alpha, and information ratio
  - Most of these rarely have statistically significant values because you need a long time series of data, over which time conditions are presumed but not guaranteed to be stable
  - We would like a measure that uses more information so we can get statistically meaningful results over a shorter window
- Manager's occasionally experience very bad return outcomes for a period of time
  - We need a means to discriminate the manager being bad from a truly random event

What We Probably Don't Care About

- There is an enormous literature in finance regarding whether asset managers collectively exhibit skill
  - Obvious implications for concepts of market efficiency
  - Most of this work is based on the concept of "performance persistence": those that perform consistently well must be skillful
  - Stewart, Scott D. "Is Consistency Of Performance A Good Measure Of Manager Skill?," Journal of Portfolio Management, 1998, v24(3,Spring), 22-32.
  - Brown, Stephen J. and William N. Goetzmann. "Performance Persistence," Journal of Finance, 1995, v50(2), 679-698.
  - Elton, Edwin J., Martin J. Gruber and Christopher R. Blake. "The Persistence Of Risk-Adjusted Mutual Fund Performance," Journal of Business, 1996, v69(2,Apr), 133-157.
- But we want to evaluate only one manager

Usual Methods

- There is lots of literature on using traditional return performance metrics such as alpha and information ratio as proxies for manager skill:
  - Kritzman, Mark. "How To Detect Skill In Management Performance," Journal of Portfolio Management, 1986, v12(2), 16-20.
  - Lee, Cheng F. and Shafiqur Rahman. "New Evidence On Timing And Security Selection Skill Of Mutual Fund Managers," Journal of Portfolio Management, 1991, v17(2), 80-83.
  - Marcus, Alan J. "The Magellan Fund And Market Efficiency," Journal of Portfolio Management, 1990, v17(1),85-88.
- You need very long time series of return observations to have enough data to get anything statistically significant by which time conditions may change
- Just going to daily data doesn't help
  - diBartolomeo, Dan. "Just Because We Can Doesn't Mean We Should: Why Daily Attribution is Not Better", Journal of Performance Measurement, Spring 2003

Problems in the Evaluation of Managers

- Much manager evaluation occurs relative to benchmarks that are often not suitable for the manager's investment approach
- Evaluation of past performance is based on standardized periods (i.e. 5 years) rather than periods that are relevant to the manager in question
- Many evaluation measures such as Sharpe ratio or information ratio correspond to meaningful investor utility for only a small fraction of investors
- The statistical significance of ex-post performance is measured in a simple time series fashion.
  - It does not include the context of whether the manager exists among a tightly bunched set of peers or a widely dispersed set
  - This is critical in examining the "luck versus skill" issue

The Persistence Literature

- If markets are very efficient, there should be no persistence patterns in active management returns. While there are there are innumerable studies showing markets are relatively efficient, many fund studies show that some persistence does exist
- Hendricks, Patel and Zeckahuaser (1993)
  - Find positive persistence only over time horizons less than a year
  - Stronger persistence among worst managers who stay worst
- Elton, Gruber, and Blake (1996)
  - Persistence of "risk adjusted" returns over one to three year time horizons
  - Appears to be correlated with investor capital flows
- Goetzmann and Ibbotson (1994)
  - Persistence over a one to two year horizon
  - Effect is stronger for more volatile funds

The Persistence Literature is Persistent

- Carhart (1997)
  - Some persistence over a one year horizon but not longer
  - Investment style and expense ratios explain most persistence effects. No evidence of "stock picking" skill
- Stewart (1998)
  - Funds that have consistently outperformed the &P 500 over a screening period also outperform during subsequent periods
  - Consistent performers hold more diversified portfolios
- Brown & Goetzmann (1995)
  - Find performance persistence in mutual funds using several methods
  - Superior performance is correlated across managers (style herding)
- Detzel and Weigand (1998)
  - Some persistence in mutual fund returns, but after adjusting for manager investment style, all persistence in returns is explained.

A Simple Prescription for Success: PWER

- Classification
  - Make sure each fund is being measured against the right benchmark and the right peers
  - We use an augmented method of returns-based style analysis
- Process Control
  - Evaluate each manager over the evaluation period that is the best for that particular manager
  - We use a "Sequential Probability Ratio Test" called CUSUM to find the optimal evaluation period
- Evaluate Past Performance
  - Use a return measure such as alpha, not the Sharpe Ratio or Information Ratio as your measure
  - Use a Bayesian framework adjustment to ex-post alpha to reflect contemporaneous dispersion across managers
  - Does the CUSUM analysis show improving or declining efficiency?

Manager Classification Issues

- Many plan sponsors want to break out of simplistic classifications
  - No more "large-small", "growth-value", "international"
  - Strong desire to allow unconventional strategies that might add value
- Allowing unconventional strategies creates the increased risk of "gamed" peer comparisons
  - "The best way to win a contest for the largest tomato is to paint a cantaloupe red and hope the judges don't notice"
- Forming manager peer groups:
  - For conventional managers, using iterated returns based style analysis from diBartolomeo and Witkowski (FAJ, 1997)
  - For unconventional managers, we synthesize a peer group using a form of Monte Carlo simulation from Surz (JOI, 1994)

Time Horizons for Evaluating Track Records

- Practitioner tradition in the investment industry is to evaluate active manager track records over a long period
  - **At least** 3 to 5 years
  - Some will argue a full "market cycle" is needed
- As we've seen, all the academic studies refute this
  - No evidence that long-term past performance is predictive of future performance
  - If there is any meaning to past performance at all, its short-lived, perhaps the last year
  - Porter and Trifts (2014) finds an inverse relationship between PM tenure and cumulative performance

The Key Question

- What time portion of a track record do we really need to evaluate as part of our monitoring of manager "quality control"
- What we need is a procedure to draw the line between getting enough meaningful data within a manager's record and older, stale data that should be ignored
- Enter CUSUM

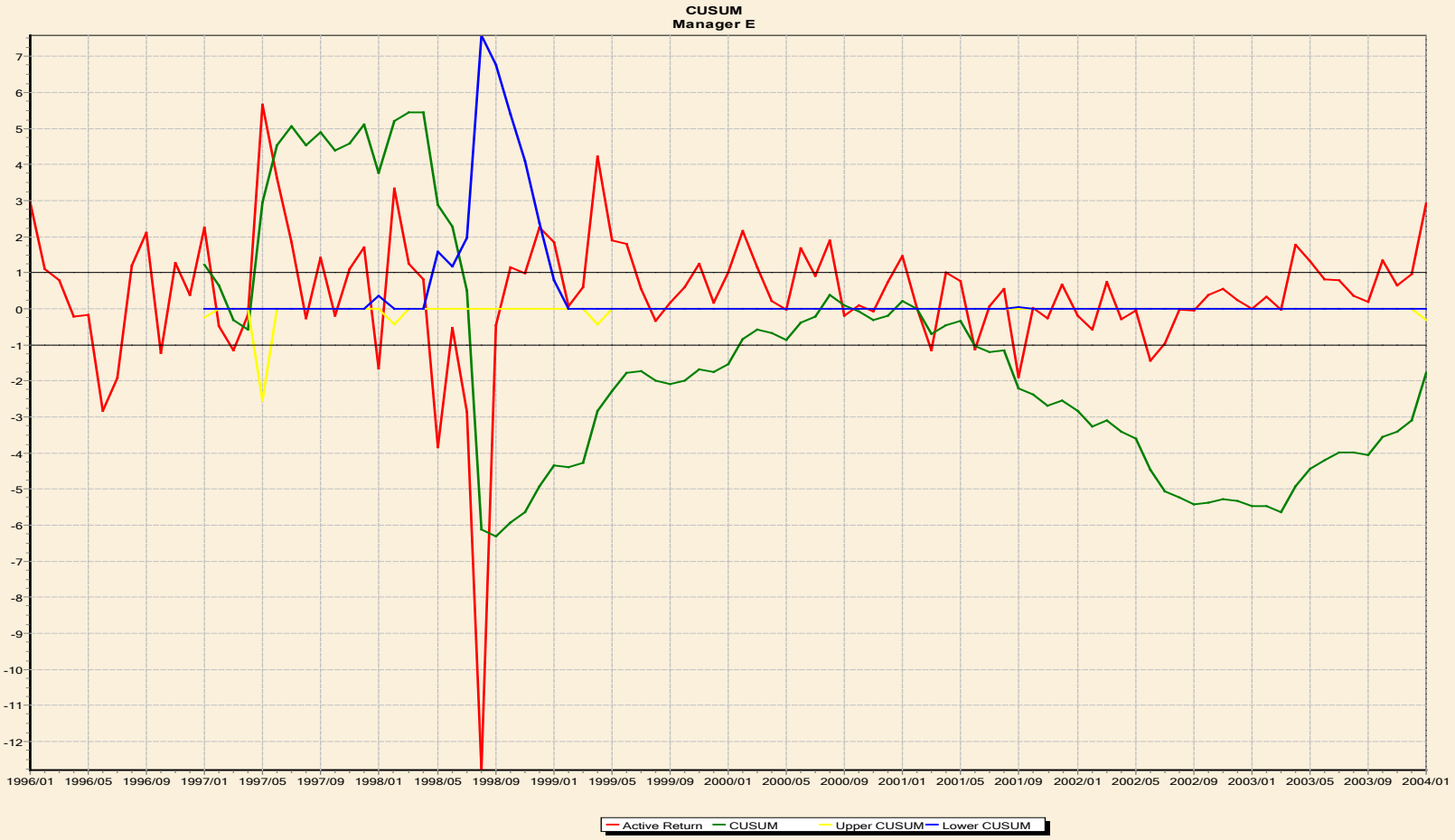
The CUSUM Technique

- Backward looking sequential probability ratio test
- Created by E.S. Page in 1954
  - Reliably detects small process shifts
  - Insensitive to probability distribution
  - Provably optimal: detects process shifts faster than any other method.
  - Robust, good under almost any definition of optimality
  - Much better than exponentially weighted moving average.
- Mathematically very tractable: its literally adding up a series of numbers
- Easily analyzed algebraically or graphically

A Robust Method of Manager Monitoring

- CUSUM analysis defines key turning points in the active return time series, and defines statistical significance of results subsequent to the turning point
  - Use of CUSUM to monitor active managers started with the IBM pension fund
  - Philips, Stein and Yashchin (2003). The PSY CUSUM method classifies managers into three categories: Good, We Don't Know, and Bad. Managers are reviewed whenever a class boundary is crossed, but is not an automatic "hire/fire" signal
- Our use of CUSUM is different
  - Focus on whether performance is improving or declining since the last regime change
  - If the effectiveness is improving, the CUSUM will plot as an upward sloping line
  - If the effectiveness is declining, the CUSUM will plot as a downward sloping line

CUSUM (Green) Plot Shows Regimes Shifts



We Have the When, Lets Deal with the What

- Many performance measures are not congruent to adding value for investors
  - deGroot and Plantinga (2001)
  - Consider a manager that adds exactly one basis point of return in every time period. The information ratio is infinite, but very little investor wealth is added
  - Availability of infinite leverage solves the problem
- We chose to measure excess return above a carefully chosen style benchmark that should reflect both risk and investing approach
  - This directly measures added value for investors
  - Our CUSUM analysis is already a variation on information ratios
  - Weight observations in a Kalman filter like approach to maintain goodness of fit

Separating Luck from Skill

- To maximally exploit our information about manager performance we need to separate skillful managers from the merely lucky
- We need to adjust for the fact that if manager returns are widely dispersed within a peer group, its easier to have a high excess return. If the dispersion of returns is low, its harder
- We adopt a method a Bayesian framework of a "precision weighted" estimate that incorporates information about the dispersion of peer fund returns during the evaluation period for each fund
  - Similar to Shanken and Jones (2004) without the Monte-Carlo simulations

Precision Weighted Excess Return Estimated

- Lets assume Manager X has an excess return (A) of 4% per year with a standard deviation (S) of 4%
- Over the same time period, the average peer manager had an annual excess return of .25% (Mean), and the dispersion (CSD) of the excess returns across the peer group is 1.5%

$$E = (A/S^2 + Mean/CSD^2) / (1/S^2 + 1/CSD^2)$$

$$A = 4,\ S = 4,\ MEAN = .25,\ CSD = 1.5$$

$$E\ (precision\ weighted) = (0.361) / (0.5069) = 0.712$$

- We assume the manager has skill sufficient to add 71 basis points per year over the benchmark

Empirical Analysis BDW (2006)

- Large scale tests on three data sets
  - Domestic mutual funds, International mutual funds, Hedge funds
- The hypothesis that past returns can be used to predict future returns is supported to a degree of virtual statistical certainty
  - Using raw excess returns, the expected excess returns are about 20% of the observed past returns
  - Using precision weighted excess returns, the expected values are over 40% of the past values
- Given the observed dispersion among manager returns, large and economically significant excess returns should be available to investors
  - Most recent tests on a smaller data set produces comparable results

IR as Skill

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

- $IR = IC * Breadth^{.5}$

  IR = alpha / tracking error
  IC = correlation of your return forecasts and outcomes
  Breadth = number of independent "bets" taken per unit time

- But IC and breadth are not transparent to investors, allocators or fundamental managers
- We only observe IR externally and its tough to work with
  - Jobson, J. D. and Bob M. Korkie. "Performance Hypothesis Testing With The Sharpe And Treynor Measures," Journal of Finance, 1981, v36(4), 889-908.

Fundamental Law Makes Big Assumptions

- There are no constraints at all on portfolio construction
  - Positions can be long or short and of any size
- We measure only "independent" bets
  - Buying 20 different stocks for 20 different reasons is 20 different bets. Buying 20 stocks because they all have a low PE is one bet, not 20!
  - Not transparent to clients and fundamental managers
- Transaction costs are zero, so bets in one time period are independent of bets in other periods
  - This is the property that casinos depend on. Once we have the odds in our favor, we want to make lots of bets
- Research resources are limitless so our forecasting effectiveness (IC) is constant as we increase the number of eligible assets

Enter the Transfer Coefficient

- Clarke, Roger, Harindra de Silva and Steven Thorley. "Portfolio Constraints And The Fundamental Law Of Active Management," Financial Analyst Journal, 2002, v58(5,Sep/Oct), 48-66.

- $IR = IC * TC * Breadth^{.5}$

  IR = alpha / tracking error
  IC = correlation of your return forecasts and outcomes
  TC = the efficiency of your portfolio construction (TC < 1)
  Breadth = number of independent "bets" taken per unit time

- We define:

  $EIC = (IC * TC)$

What Drives the Transfer Coefficient?

- Imagine a manager with a diverse team of analysts that are great at forecasting monthly stock returns on a large universe of stocks, but whose portfolio is allowed to have only 1% per year turnover
  - Good monthly forecasts, diverse reasons and a large universe imply high IC and high breadth
  - But if we can never act on the forecasts because of the turnover constraint TC can be zero or even negative
- If we can't short a stock that we correctly believe is going down, or take a big position in a stock that we correctly believe is going up, TC declines
  - The more binding constraints we have on our portfolio construction, the more return we fail to capture when our forecasts are good
  - For bad forecasters, a low TC is good. You hurt yourself less when you constrain your level of activity

A Quant Way to Think About It

- Every portfolio manager must believe that the portfolio they hold is optimal for their investors
  - If they didn't they would hold a different portfolio
- If we describe investor goals as maximizing risk adjusted returns, we know that the marginal risks associated with every active position must be exactly offset by the expected active returns
  - Guaranteed by the Kuhn Tucker conditions for finding the maximum of a polynomial function
  - For every portfolio, there exists a set of alpha (active return) expectations that would make the portfolio optimal. We call these the implied alphas
  - Sharpe, William F. "Imputing Expected Security Returns From Portfolio Composition," Journal of Financial and Quantitative Analysis, 1974, v9(3), 462-472.
  - We all just have to agree on the risk model

EIC is Externally Observable

- We define the EIC as the skill measure

- EIC is the pooled average correlation of the implied alphas and the realized returns at the security level but we don't even need to do the calculation
  - If our forecasting skill is good (high IC) and our portfolio construction skill is good (high TC) then EIC will be high
  - If either IC or TC is low, EIC will be low
  - As this measurement involves every active position during each time period, the sample is large and statistical significance is obtained quickly

Breadth is Observable Too

- Since we've agreed on a risk model, breadth is now observable

$$\text{Breadth} = (M/P)^2 * T/100$$

M = average volatility of individual securities in the portfolio
P = *portfolio* volatility
T = annual portfolio percentage turnover

- Since IR is observable too if we estimate breadth we can solve for EIC without actually doing the pooled correlation test

A Bonus Alternative View

- Active managers can add value in two ways:
  - Being right more often than they are wrong about which securities will outperform the market. Sort of like a batting average in baseball
  - Getting bigger magnitude returns on gainers than on losers. You can have a batting average below 50% and still make money if you hit a decent number of "home runs"
- Peter Lynch used to refer to "ten baggers"
  - Stocks that go up ten fold in value while you hold them
  - Just a couple can have a huge effect on portfolio returns
  - Basis of venture capital investing
- Batting average concept first formalized in:
  - Sorensen, Eric H., Keith L. Miller and Vele Samak. "Allocating Between Active And Passive Management," Financial Analyst Journal, 1998, v54(5,Sep/Oct), 18-31.

Batting Average and Active Return Skew

- A formalization was proposed by hedge fund manager Andrei Pokrovsky (formerly Northfield staff) in 2006
- We can easily measure batting average externally
  - It is the percentage of cases in which active returns and active weights are of the same sign
  - High numbers are good
- Take the vector product of active weights and active returns. Measure the skew statistic of the distribution
  - Positive skew in active returns is a measure of portfolio construction efficiency

Style Dependency

- Value managers will tend to have high batting average and low skew
- Growth/momentum managers will tend to have lower batting average but higher skew
- Trend following behavior creates the skew
  - Wilcox, Jarrod W. "Better Risk Management," Journal of Portfolio Management, 2000, v26(4,Summer), 53-64.
  - diBartolomeo (2003) http://www.northinfo.com/documents/132.pdf
  - Bondarenko, Oleg. "The Market Price of Variance Risk and the Performance of Hedge Funds", University of Illinois-Chicago Working Paper, 2004.

Playing with the Big Kids: Capacity Analysis

- As some plan sponsors are very large, we wanted to incorporate capacity limits into capital allocations
  - For background, see Vangelisti (2006)
- We adjust manager alpha estimates as a function of capital allocated
  - Market impact of trading will rise decreasing net alpha (k values from Northfield trading cost model)

$$\alpha = \Sigma_{i = 1 \text{ to } N} \alpha_i w_i (1 - k_i w_i)$$

- We frame the capacity upper bound in the form of a liquidation policy:
  - We want limit the amount of capital to the manager such that liquidation of P percent of the manager's portfolio could be accomplished in N days at cost of no more than C %, to begin within T days of the decision

Conclusions

- The PWER methodology involves several component steps that combine to provide demonstrated effectiveness in predicting active performance. I
- We propose the Effective Information Coefficient as a second metric of skill
  - It is estimated both over time and across assets so sample sizes get large quickly
  - It incorporates both key aspects of investment skill, forecasting returns and forming efficient portfolios
  - If we use a risk model to estimate breadth, we can reduce the computational requirements
- We present a third alternative representation of skill as "batting average" and "payoff skew" that is observable to allocators