A Risk-Modeling Approach to Equity Manager Oversight

Active skill is not simply a function of performance relative to a benchmark – anyone can invest in a benchmark with leverage, outperforming it most years and over the long term. Active skill is instead outperformance in excess of what could have been generated with a passive portfolio of index funds, ETFs, or futures matching the fund's historical risks. To show evidence of skill a fund must perform better than its passive replicating portfolio.

Beta differences from a benchmark are a byproduct, often unintentional, of any stock-selection process. Once identified, though, beta-differences can be freely obtained or offset. Asset owners no longer need to pay active fees for passive exposures.

Further, since these beta-exposures tend to be random or mean-reverting, isolating stock-selection return from their impacts removes much of the noise that obscures the stock-selection skill worth paying for.

This paper introduces a powerful technique for Isolating active from passive contributions which reveals active risk, persistent security-selection skill that merits compensation, and unintentional bets that may endanger performance. We start by examining existing attempts to separate active from passive risk and return, then compare equity risk modeling methodologies and capabilities, and conclude with new insights from an equity risk model built for manager oversight.

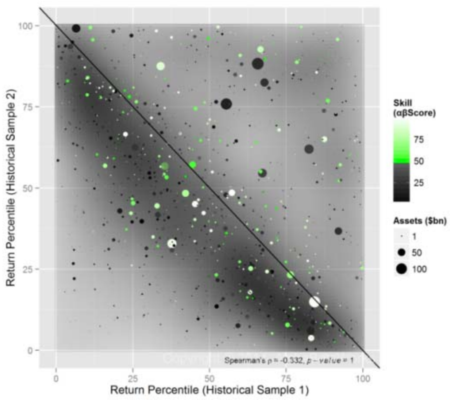
Asset owners devote significant resources to implementing active equity investment strategies, yet the uncomfortable truth is that traditional performance metrics do not reliably detect manager skill and predict future outperformance.

The problem with assessing skill using nominal performance is that the effects of luck or market fluctuations, potentially random or mean-reverting, overwhelm any effects of manager skill.

For example, a typical mutual fund's volatility attributable to security-selection accounts for approximately 5% of the total. Market noise accounts for approximately 95%.

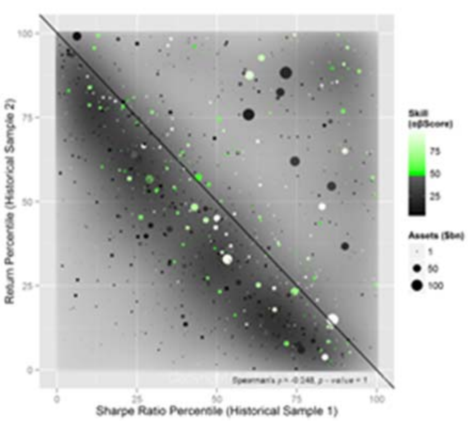
Due to this high level of market noise, traditional techniques relying on nominal returns require decades of data to detect statistically significant evidence of skill.

Nominally top-performing managers over a short 3-5 year period are likely to be top-performing because they happened to take favorable systematic risks in a particular market environment. Unfortunately, market regimes vary. Consequently, as the regimes change, the top quartile managers in one period are more likely to be in the bottom quartile in a subsequent period than to remain in the top one.



Among Medium Turnover U.S. Mutual Funds, the relative ranking of a fund's returns in one sample of history is *negatively correlated* to its relative ranking in the other. When "skill" is evaluated naively, "the best" funds in one period tend to become "the "worst" in another and vice versa:

The Sharpe Ratio and similar metrics simply re-process the same return data, presenting it in a different form. They too suffer from the same problems as simple return ranking:



The above limitation of nominal performance metrics is among the reasons consultants' recommended funds tend to underperform.

Academic evidence has consistently shown that managers are better at telling stories than consultants are at detecting which stories are predictive of future performance. Recent research and regulatory pressure suggest that consultant recommended funds tend to under-perform non-recommended ones and active fees further reduce portfolio performance (see, for example: Investment consultants fail to select funds that outperform Financial Times 3/2018; Picking Winners? Investment Consultants' Recommendations of Fund Managers *Journal of Finance*, 9/2014; Money Doctors *Journal of Finance,* 2/2015).

Existing Approaches to Isolating Active Risk and Return

In order to select managers likely to outperform in the future, and to know when managers should be replaced, the challenge is to look beneath the surface to determine whether the true source of returns is investment skill (stock picking, market timing, etc.) or some combination of luck, high beta, and out-sized risk.

Two attribution approaches attempt to separate active and passive contributions: holdings-based and returns-based analyses.

Holdings-Based Approaches Without a Risk Model

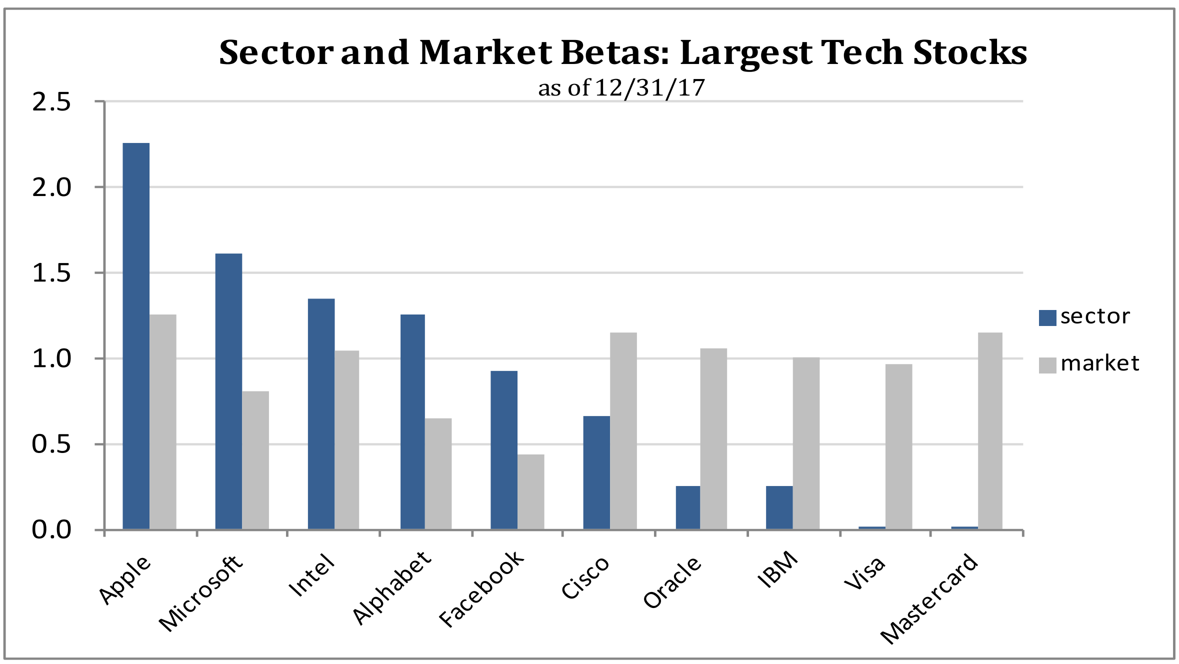
Holdings-based approaches to performance attribution typically rely on the principles discussed in Brinson and Fachler, *Journal of Portfolio Management,* 1985 and in Brinson, Hood, and Beebower *Financial Analyst Journal*, 1986. The Brinson model is widely used by the investment management community to attribute portfolio returns by following a simple analysis of sector allocation. The approach attributes relative returns to sector allocation relative to the benchmark and to stock selection, defined as outperformance of a stock's sector.

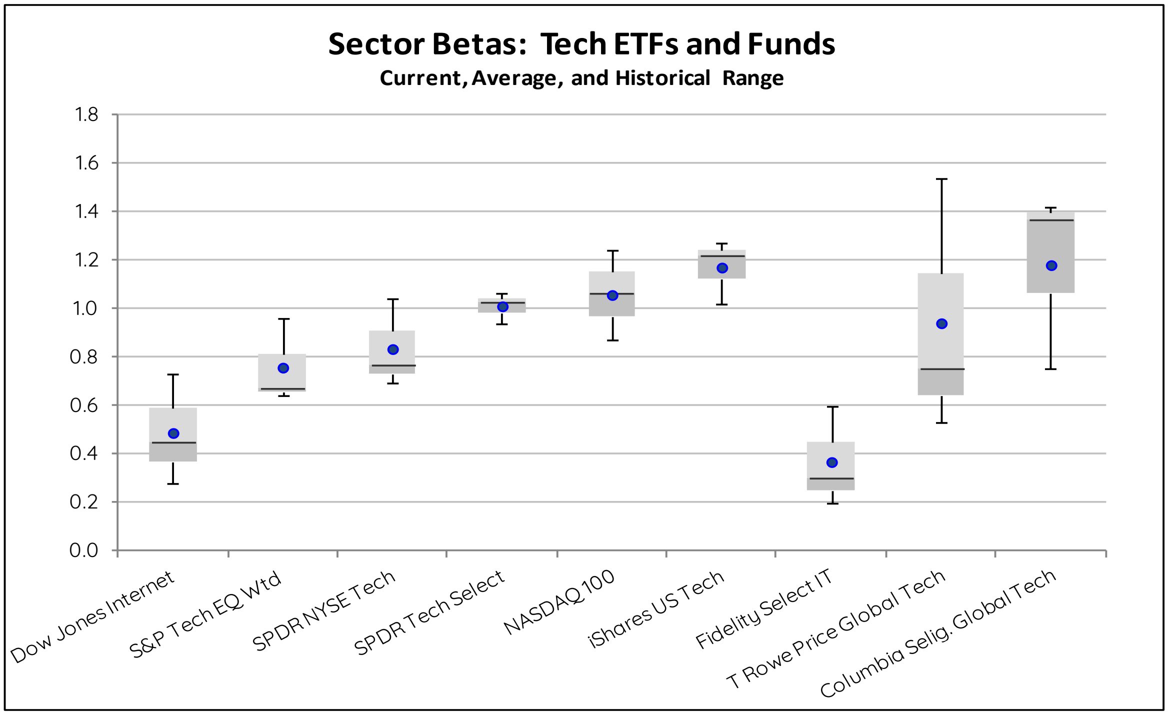
Providers, including Morningstar, FactSet, Novus, and others, offer attribution analysis using the Brinson approach.

Active Share, introduced in 2006 by Yale professors Martijn Cremers and Antti Petajisto, is a related approach that estimates an equity portfolio's level of activity relative to a passive benchmark. Active Share is calculated as one-half the sum of the absolute value of the differences of the weight of each holding in the manager's portfolio and the weight of each holding in the benchmark index.

Both the Brinson and the Active Share approaches to evaluating activity and performance assume that all stocks have identical market risk and all stocks within a sector have identical sector risk (e. g., a dollar in Apple and a dollar in Facebook have the same market, sector, and other systematic exposures).

Unfortunately, sector and market betas vary widely across securities, as do sector betas even among the largest sector indexes and funds.





Fortunately, any performance attribution system that claims to identify active return is easily tested.

Since security-selection return is a residual, free from systematic risk, it is by definition uncorrelated with passive benchmarks. To the extent security-selection return and passive return calculated by a given system are correlated, the system has failed.

Additional Tests of Holdings-Based Analysis

In Three Holdings-based Style Analysis Tests we demonstrate that a system that does not distinguish between low and high market and sector exposures (betas) will be fatally flawed. It will:

- Underestimate the risk of high-risk portfolios
- Overestimate the risk of low-risk portfolios
- During bull markets, deem as skilled poor stock pickers passively taking high risk
- During bull markets, deem as unskilled excellent stock pickers passively taking low risk

Regrettably, some of the most commonly used commercial products exhibit these flaws. Even the highest-priced offerings are not immune. Asset Owners that are unaware of these flaws will be doubly-blindsided in periods of market turmoil. Their realized risk may be higher than estimated. Their "best" managers may turn out to be their worst.

Returns-Based Style Analysis

Returns-based style analysis and returns-based performance attribution techniques perform regressions to compute portfolio betas (exposures to systematic risk factors) and alphas (residual returns unexplained by systematic risk factors).

The simplicity of the returns-based approach has made it popular. It is often the only practical method for evaluating multi-asset-class portfolios that span commodities, public securities, derivatives, and private investments. However, this simplicity comes at a heavy cost.

The Assumption of Stable Exposures

A key assumption of most returns-based analyses is the constancy of factor exposures. This assumption breaks down for active managers. In Flaws of Returns-based Style Analysis we show that:

- Returns-based analysis *can* be effective – but only when a passive manager does not significantly vary exposures to market, sector, and macroeconomic factors.
- When an active manager varies bets, a returns-based analysis typically yields flawed estimates of portfolio risk.
- When a manager varies bets, a returns-based analysis may not even accurately estimate average portfolio risk.
- A returns-based analysis will be the least predictive for active managers. In fact, errors will be most pronounced for the most active funds:
    - Estimates of a manager's historical and current systematic risks may be flawed.
    - Skilled funds may be deemed unskilled.
    - Unskilled funds may be deemed skilled.

More subtle, but no less dangerous, issues with investment risk and skill evaluation using returns-based performance attribution include overfitting and collinearity.

Equity Risk Models

Multi-factor equity risk models measure portfolio risk by calculating individual security factor exposures and can distinguish between systematic risk (due to endogenous factors that affect multiple securities) and idiosyncratic risk (specific to an individual security).

Equity risk models are classified as fundamental models, macroeconomic models, and statistical models.

The most popular type of model used in practice is the fundamental model. Many of the inputs used in a multi-factor risk model are those used in traditional fundamental analysis such as P/E, momentum, earnings growth, turnover, leverage, and yield. These models have become the standard approach to measuring portfolio risk and constructing portfolios in the investment management community and there are a number of commercially available fundamental multi-factor risk models and some investment management companies have developed proprietary models.

As these models are designed to estimate risk as precisely as possible for even the most narrow portfolios, popular models today often use well over 100 risk factors, most of which are not directly investable.

Bloomberg's Port function uses a fundamental risk model similar to but more rudimentary than Barra's and Blackrock's. Our tests found that Bloomberg's portfolio risk model often captured less than half of the relative systematic (factor) risk explained by the more robust models.

Bloomberg's, and other fundamental models, have two key limitations when used for portfolio oversight:

1) Fundamental models can use over 100 risk factors, most of which are not investable. The attribution provided, while useful for portfolio construction and optimization, is not meaningful for oversight. Knowing that you underperformed because you were over-exposed to the momentum, leverage or other factors that are not passively investable is not terribly useful for oversight.

2) Fundamental models assume that all exposures to country, currency, and sector factors are equal to one for all securities. There is no distinction in sector risk between the least risky stock in a sector and the most risky one. Since individual securities' sector risks vary widely (Apple, for example, recently had Technology sector exposure of 2.2 while IBM's has one of 0.2) Bloomberg's and other providers' exposures can provide a very misleading picture of sector, country, and currency risk.

Macroeconomic factor models are the simplest and most intuitive type. They use observable economic time series as measures of the prevalent factors in security returns. The return of each security is assumed to respond linearly to the macroeconomic shocks. As in all factor models, each security also has an asset specific return unrelated to the factors. A security's linear sensitivities to the factors are called the factor betas of the security. A drawback to macroeconomic factor models is that they require identification and measurement of all the systematic shocks affecting security returns. A small number of underlying sources of risk may exist, but without knowing exactly what they are, or lacking data to measure them, these sources are of little use in explaining returns. (see: Three Types of Factor Models, Financial Analyst Journal May-June 1995)

Statistical factor models use various maximum likelihood and principal-components analysis procedures on either cross-sectional or time-series security return samples to identify the significant underlying drivers of returns, or factors. Statistical factor models rely on fewer assumptions and use robust statistical processes to estimate factor betas.

Overview of the Three Types of Factor Models

Factor Model Types	Inputs	Estimation Technique	Outputs
Macroeconomic	Security Returns and Macroeconomic Variables	Time-series regression	Security factor betas
Statistical	Security returns	Iterated time-series/ cross sectional regression	Statistical factors and security factor betas
Fundamental	Security returns and security characteristics	Cross sectional regression	Fundamental factors

Some of the most popular statistical equity risk models are provided by Northfield.
From an oversight perspective, one of the greatest limitations of NF's models is that factors are weighted by the square root of market capitalization (see: Northfield U.S. fundamental equity risk model, page 6 and page 3 footnote 3).
Factors weighted by the square root of market capitalization split the difference between weighting by market capitalization and equal-weighting which can offer advantages in portfolio optimization and risk forecasting scenarios – the primary use of NF's models by investment management firms.

Unfortunately, since passive index funds are necessarily cap-weighted, rather than square-root-of-cap-weighted, NF's models cannot attribute performance to the passive exposures (as available with broadly-used index funds and ETFs), factor timing, and security selection components. This leads to an oddity: even if the attribution claims that a portfolio delivered alpha relative to the model it does not mean that it has outperformed a passive portfolio with the same risk.

As a result, NF's models do not provide meaningful attribution for manager and portfolio oversight. This shortcoming is due to an explicit design decision to target portfolio optimization and risk forecasting, rather than performance attribution relative to cheaply investable passive portfolios.

A Statistical Approach to Equity Oversight

A statistical equity risk model built for oversight, using a limited number of factors that map to common passive portfolios such as index funds and ETFs, explains risk as well as the most robust fundamental models in most cases, but can also distinguish skill from random market fluctuations.

For most portfolios, differences from the benchmark in exposures to Market, Sector, Regional, and Style factors, all available via passive investments, explain two-thirds of relative return.

These passive differences in systematic risk from the benchmark, whether intended or not, are not part of a manager's active return. Or, in any case, do not deserve compensation by allocators with a duty to deliver the best performance, since such exposures are available through cheap index funds or ETFs if the exposures are intentional and desirable. In many cases, exposures are instead unintended and undesirable risks to the investment objectives, and ought to be offset with index funds or ETFs, or through a careful multi-manager allocation.

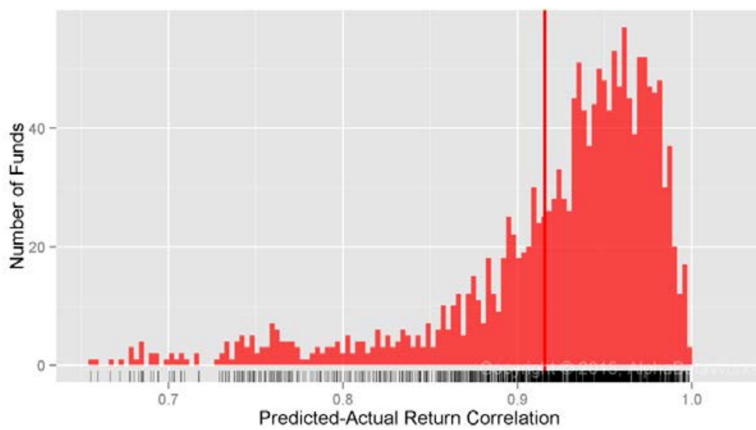
Isolating active risk from these passive exposures and separating their respective contributions to return reveals managers' real active risk, security-selection performance, evidence of active management skill, and any unintentional portfolio risk.

Findings

Passive factors, which explain the majority of absolute and relative return, vary significantly across individual funds and aggregate portfolios

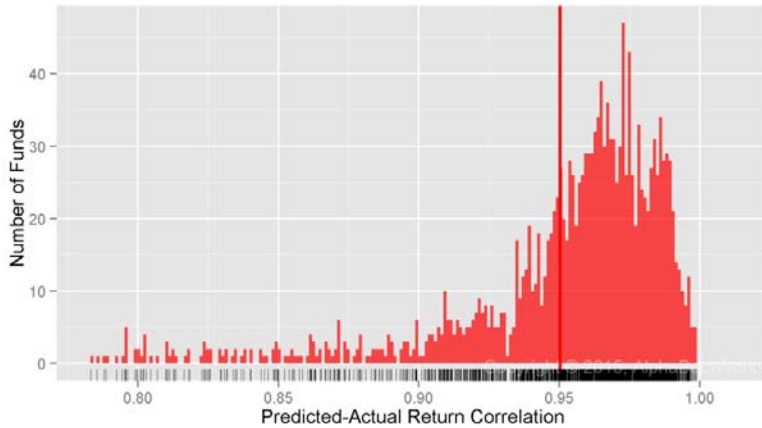
We used the ABW Peer Analytics Statistical Equity Risk Models to identify sources of risk for approximately 3000 non-index equity mutual funds and the one hundred largest U.S. insurer equity portfolios where holding data is publicly available via statutory filings. We summarize our findings below.

Since Market Beta is the dominant factor behind portfolio performance, even a very simple model built with robust methods delivers 0.92 mean and 0.94 median correlation between predicted and actual monthly returns:



Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.136	0.901	0.940	0.916	0.965	0.998

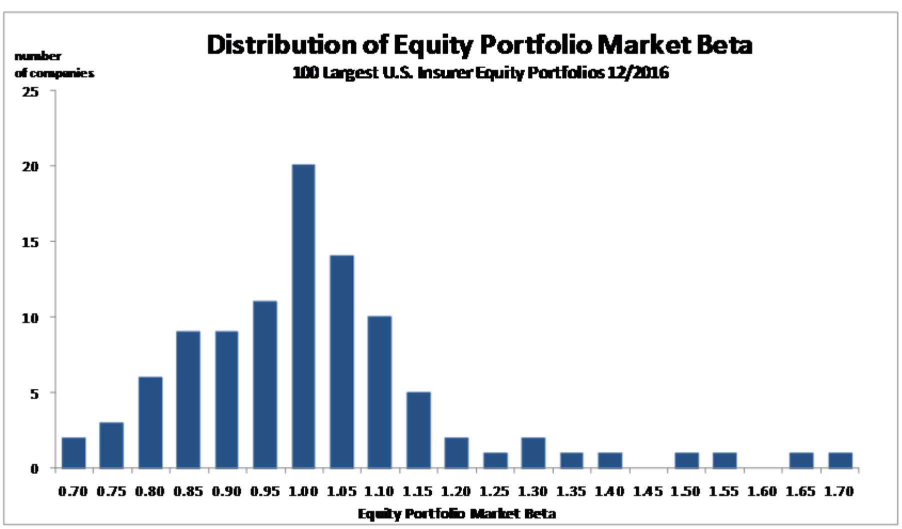
Extending the model with Industry Sector and Style Factors (Value/Growth and Size) as well as Macroeconomic Factors (Bonds, Oil, Currency, etc.), we arrive at the ABW Peer Analytics U.S. Equity Statistical Risk Model. It delivers 0.95 mean and 0.96 median correlation between predicted and actual monthly returns for U.S. Equity Mutual Funds:



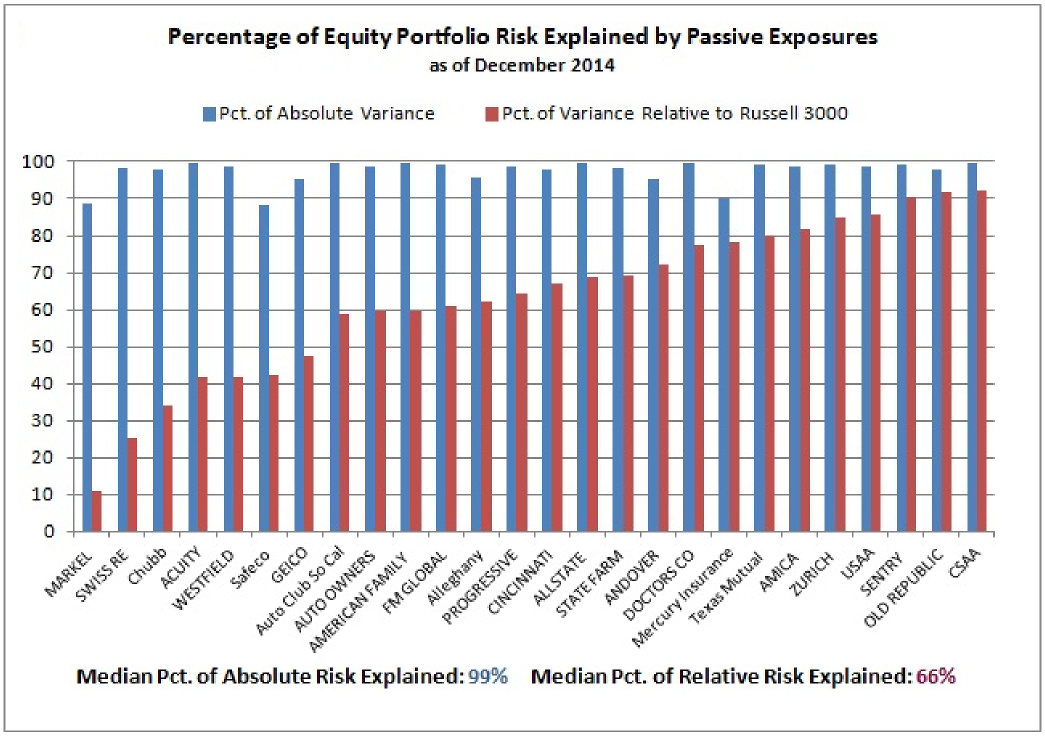
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.666	0.942	0.9623	0.950	0.977	0.999

Complex equity risk models with non-intuitive factors may offer no better predictions than robust models with a few intuitive factors. For a typical U.S. mutual fund, a statistical equity risk model with intuitive and investable factors delivers over 0.96 correlation between predicted and actual monthly returns.

For an idea of how much market risk varies across aggregate equity portfolios, we compare market beta for each of the 100 largest U.S. insurance company total equity portfolios.



Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.70	0.85	0.97	0.99	1.12	1.70



We further compare the percentage of variance explained by passive exposures to common risk factors (i.e., market, sector, size, etc.) for thirty of the largest insurance company equity portfolios. Both percentage of absolute variance and percentage of variance relative to the market due to these factors are shown for individual companies. At the median, 99% of absolute variance and 66% of relative variance are explained by passive factor exposures.

Market and other passive factor exposures vary significantly across companies and through time and are the principal drivers of performance. These exposures should be the focus of equity portfolio oversight.

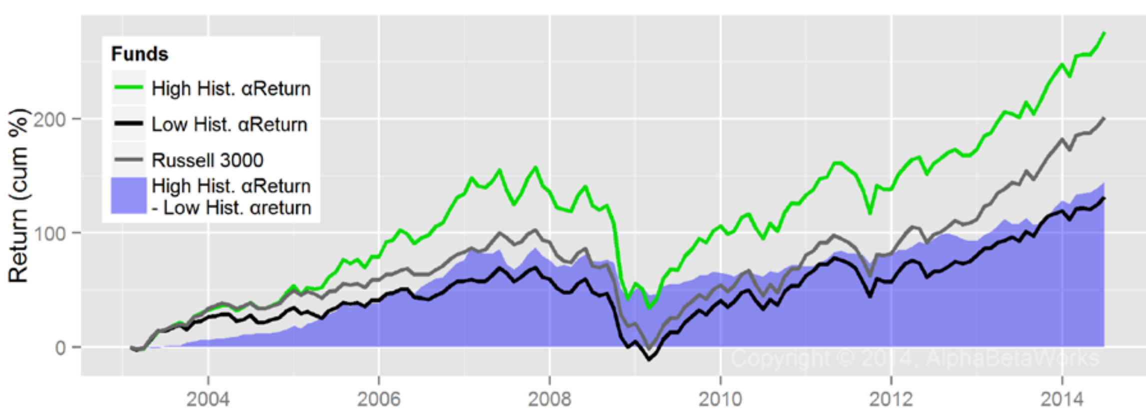
Security selection skill persists

By isolating returns due to security selection and market timing from those due to passive differences in factor (systematic) risk from the benchmark, the ABW Peer Analytics risk models mitigate the impact of noise and reveal skill.

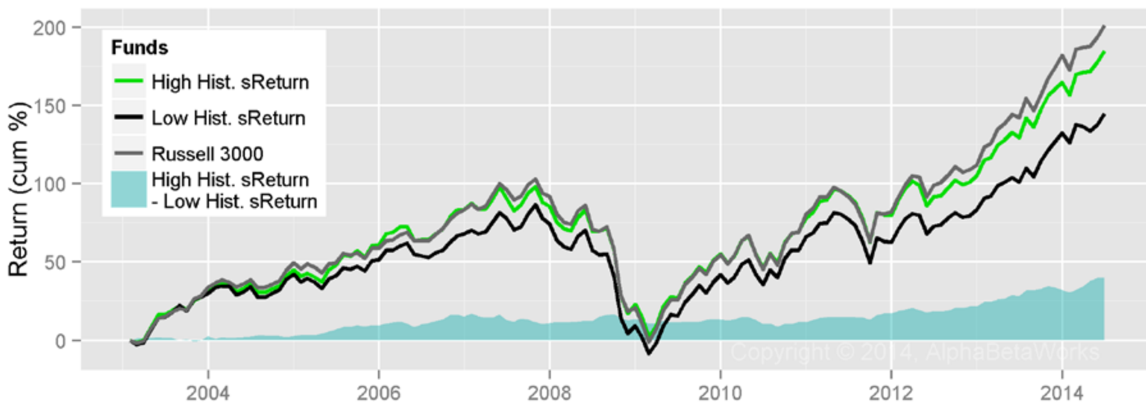
This isolation of returns attributable to skill improves the signal/noise ratio and makes the performance data insightful and predictive. As shown earlier, the correlation between past and future nominal outperformance is -0.3 — significantly counterproductive in assessing skill. But stock-selection skill, when properly isolated from the effects of systematic differences from the benchmark, results in a positive 0.3 correlation between past and future performance.

Top-decile stock-pickers are twice as likely to generate positive stock-picking returns as negative stock-picking returns in the subsequent few years. Bottom decile managers are more than twice as likely to generate negative stock-picking returns as positive ones in subsequent years (see: performance persistence within style boxes and performance persistence within international style boxes).

A fund that produced positive stock-selection return, αReturn, in the past will likely generate positive returns even in flat and moderately declining markets. In a recent study spanning 12 years, high αReturn funds outperformed the market by 75%; low αReturn funds underperformed by 70%:



Returns of mutual funds with highest and lowest trailing 36-month αReturns – equal risk sReturn is a measure of a fund's position sizing skill, liquidity, and scalability – the estimated annual percentage return a fund generated by varying position sizes. A fund that generated a negative sReturn would have done better if it had sized all positions equally; the manager either lacks position sizing skill, or the fund's AUM is too high. Negative sReturn is a strong indicator that a fund has grown too large. In a recent study spanning 12 years, low sReturn funds have underperformed the market by 56%:



Returns of mutual funds with highest and lowest trailing 36-month sReturns – equal risk:

When past performance is free from systematic noise, it can indeed be a predictive statistical indicator of future performance.

Closet indexing is prevalent

Mutual fund closet indexing is the practice of charging active fees for passive management. Over a third of active mutual funds and half of active mutual fund capital appear to be investing passively: Funds tend to become less active as they accumulate assets. Skilled managers who were active in the past may be closet indexing today. These active managers take too little active risk to compensate for an average fee, even assuming a top-decile information ratio.

Absent a risk model, asset owners can't know how active their managers are; with a risk model, asset owners can avoid paying active fees for passive exposures.

Simple analysis of holdings, Active Share for example, fails to detect closet indexing because it assumes that all stocks within the market have identical market risk and all stocks within a sector have identical sector risk. Active Share, R^2 and similar metrics are thus deeply flawed.

Simply by identifying closet indexers with a robust methodology, investors can eliminate half of their active management fees, increase allocation to skilled active managers, and improve performance.

Equity portfolio risk varies significantly

Aggregate equity portfolio market exposures vary substantially across companies and through time. Equity market risk is a necessary and important consideration in the asset allocation decision and, as the proposed new BCAR scores for insurers suggest, a critical component of portfolio oversight.

Risk models measure point-in-time market risk and other exposures based on current portfolio holdings and any changes in portfolio risk are known immediately.

True drivers of manager return are revealed

Equity risk models quantify the drivers of manager return as well as current "bets" relative to either a benchmark market index or a peer company composite index. Clients can ensure managers' relative risk exposures are consistent with expressed strategy, and manager discussions can focus on those decisions that significantly impact portfolio risk and return rather than stories that distract.

Allocation among managers can be improved

Asset owners can avoid unintentionally reinforcing or offsetting active bets among individual managers, control unintended risks, avoid closet indexing by the overall portfolio, and better assess how individual managers contribute to the aggregate portfolio's exposures.

Risk Model Validation

Equity risk models can be mathematically complex and hard to compare. Fortunately, these models are easily tested.

To evaluate the accuracy of an equity risk model, we compare returns predicted by past factor exposures to subsequent portfolio performance: We measure factor exposures using end-of-month holdings and predict the following month's return as a function of index returns.

The correlation between predicted and actual return measures a model's accuracy. The higher the correlation, the more effective a model is at hedging, stress testing and scenario analysis, as well as evaluating investment *risk* and *skill*.

The ABW Peer Analytics risk models are highly predictive and deliver over 0.96 median correlation between predicted ex-ante and reported ex-post portfolio returns for U.S. and Global Equity mutual funds (see: testing predictions of equity risk models and testing global equity risk models).

Prospective clients need not rely on our out-of-sample tests, we're happy to provide passive ETF replicating portfolios for any of your managers and you can then evaluate the models' accuracy independently. A few weeks of observations can provide dozens of observations and establish a high statistical confidence in the models' predictive accuracy.