Equity Efficacy Testing of Aapryl Forecasting Methodology

Abstract

Identifying active managers that have a high probability to outperform their peers and relevant market benchmarks has traditionally been challenging for asset owners. As we have written in the past, while commonly used quantitative methods (such as information ratio and alpha rankings) are extremely useful for characterizing a manager's past performance, because of either the cyclicality of style effects or the possible outsized influence of pure luck, they have limited ability to predict future relative performance.,

The objective of this paper is to show that using a combination of the portfolio replication techniques and skill metrics developed by FIS and Aapryl, one could identify products that are more likely to outperform their style, out of sample.   We tested the efficacy of 6 skill-based metrics to predict forward looking skill of an investment manager on multiple investment manager peer groups. Only one peer group tested, US Small Cap Core, failed to increase Alpha or the likelihood of outperformance, while the others added value.

Key definitions
- *Manager Product*: An investment portfolio or strategy managed by a portfolio manager, team, process and investment infrastructure.
- *Static Clone*: Long term passive replication portfolio consisting of 100% allocation to various market indices including indices focused on: style, region, fundamental or risk factors, market cap, and economic sensitivity. Weights are set by FIS' proprietary algorithm using the historical returns of product and indices.
- *Dynamic Clone*: Short term passive replication portfolio consisting of 100% allocation to various market indices including indices focused on: style, fundamental or risk factors, region, market cap, economic sensitivity. Weights are set by FIS proprietary algorithm using the trailing 36 months returns of product and indices.
- *Manager Skill*: Product performance minus Static Clone Performance normalized for market environment and peer group.

Methodology Overview

Manager performance varies over time, driven by factors that they directly control, as well as factors that are outside of their control. Many manager processes begin with a belief system or investment philosophy which dictates which factors from a broad universe of stocks are filtered for consideration in the portfolios that they manage. The aggregation of the group of factors filtered by the manager can be broadly described as representative of the manager's investment style. From there, managers typically perform additional fundamental or technical analysis in order to determine the final sub-set of stocks that will comprise their portfolio. However, this initial filtering of factors leads to either explicit or implicit factor exposures that can have a meaningful impact on the manager performance that is separate from their stock or factor selection skill.

Because the performance of factors is subject to market cycles, manager peer group rankings are also subject to such cycles. This cyclicality of the return of different market factors combined with investors' conflation of the performance impact of a manager's factor exposures (i.e., investment style) with their selection alpha, often leads to disappointing results (i.e., manager rankings that are not persistent and high ranking first-quartile managers falling below their peer medians over the market cycle period after they have been selected). Our portfolio replication techniques allow investors to identify key factor exposures and to disaggregate the effects of factors from skill. Additionally, the process of identifying manager factor exposures yields a beta profile which can be assessed cross-sectionally relative to managers that demonstrate a similar profile, or one can determine the relative exposure to different factors across a multi-manager portfolio.

In order to identify factors that are significant to each manager product, we implemented a regularized regression technique that performs both factor selection and beta regularization to enhance our portfolio replication model's prediction accuracy and interpretability. We further applied a constrained optimization technique in order to obtain realistic constraints to the regression outcomes that require all factor exposures to sum to 100% and that the respective exposure for each factor to be within a range of 0% to 100%. This helps the regression avoid factor exposures that are unrealistic to replicate; rendering each manager clone as a transparent and investable long-only factor replication portfolio that is straightforward for investors to implement.

As shown in the analysis of all products in the Evestment universe that were not pure index funds with at least three years of continuous performance, the product clone portfolios have a higher R_2 and lower standard deviation than their respective market benchmarks.[1] Since we did not exclude enhanced index products, it is logical to infer that an analysis which excluded enhanced index products would show an even lower R_2 to their respective market benchmarks. In other words, the manager product clones more accurately identify performance that is solely due to style effects. This is why the manager products' excess return relative to their product clone portfolios are generally lower (though not always so on an individual manager product basis) than their excess return relative to the market benchmark.

---

[1] Product vs Benchmark is the R-squared of the product vs the relevant peer group benchmark. MSCI EM, MSCI EAFE, MSCI EAFE SC, Russell 1000, Russell 1000 Growth, Russell 1000 Value, Russell 2000, Russell 2000 Growth, Russell 2000 Value. Product vs Static clone uses the methodology discussed in the paper.

Emerging Markets

R-Squared
Products vs Benchmark

R-Squared
Products vs Static Clone

89.6%

94.1%
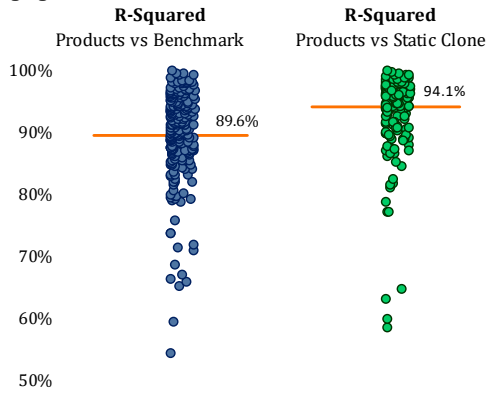
	R-Squared Vs MSCI EM	R-Squared Vs Static Clone
Average	90%	94%
Median	91%	96%

International Small Cap

R-Squared
Products vs Benchmark

R-Squared
Products vs Static Clone

88.5%

91.3%
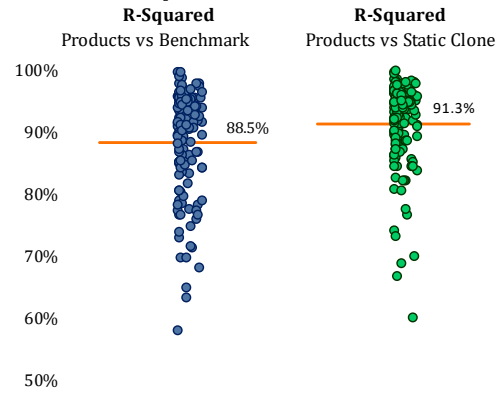
	R-Squared Vs MSCI EM	R-Squared Vs Static Clone
Average	88%	91%
Median	91%	94%

US Large Cap

R-Squared
Products vs Benchmark

R-Squared
Products vs Static Clone

91.1%

93.2%
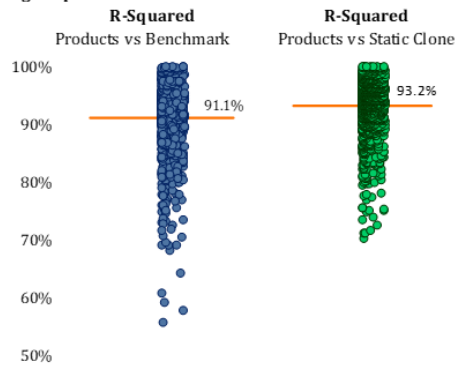
	R-Squared Vs MSCI EM	R-Squared Vs Static Clone
Average	91%	93%
Median	92%	94%

US Small Cap

R-Squared
Products vs Benchmark

R-Squared
Products vs Static Clone

89.8%

90.4%
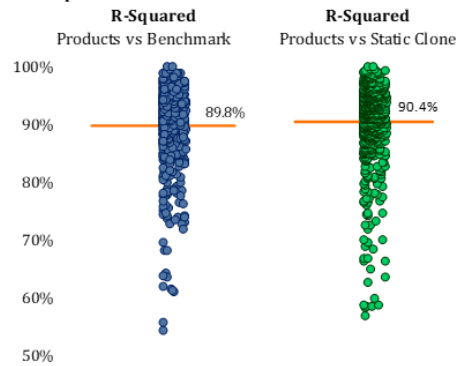
	R-Squared Vs MSCI EM	R-Squared Vs Static Clone
Average	90%	90%
Median	91%	92%

EAFE

R-Squared
Products vs Benchmark

R-Squared
Products vs Static Clone

93.4%

95.2%
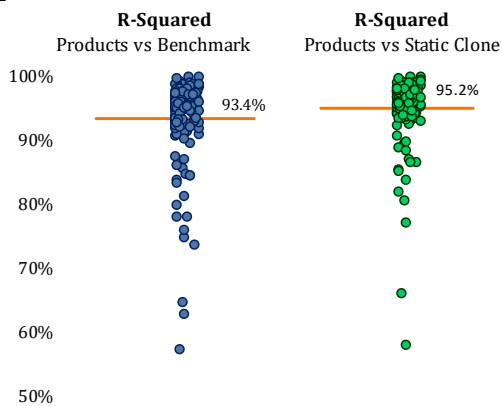
	R-Squared Vs MSCI EM	R-Squared Vs Static Clone
Average	93%	95%
Median	95%	97%

Development and Analysis of Key Skill Metrics

The skill metrics are derived by dissecting the non-style portion of the return using 3 types of excess return, each with 2 skill measurements.

Excess Returns

- *Total Excess Return:* Manager Skill = Product performance minus Static Clone Performance
- *Stock Selection Return*: Product performance minus Dynamic Clone Performance
- *Style Timing Return:* Dynamic Clone Performance minus Static Clone Performance

Skill Measurements (Calculated for each of the 3 excess return series)

- Consistency: Measure of the frequency of positive excess return. The measure begins with a standard batting average calculation and then adjusts for the length of the track record, market conditions and peer group performance. The result is a Z-score, representing the manager's relative position among similar investment products, with 0 being average.
- Edge: Measure of the relative magnitude of positive vs negative excess returns. The measure begins with an Omega ratio[2](0% threshold) calculation, then adjusts for the length of the track record, market conditions and peer group performance. The result is a Z-score, representing the manager's relative position among similar investment products, with 0 being average.

Methodology Testing

Product Universe -We examined long only equity strategies in the Evestment Expanded Monthly Database that had at least 36 months of consecutive returns between 7/31/2001 and 9/30/2018 and reported gross returns in US dollars.

In addition, a product was not evaluated unless it had the requisite 36 months of history needed to build the clone portfolios and calculate the performance metrics. The final skill metrics we evaluated were from the period ending 9/30/2015 to allow us to test a 3 year forward looking return.   To minimize survivorship bias, we included dead products in our analysis until their final month in the database and replaced any missing time periods with the returns from the Static Clone of the product. This is equivalent to replacing dead products with the most similar index fund.

---

[2] Omega Ratio can be summarized as the probability weighted **ratio** of gains versus losses for a threshold return target. It is unique in that it does not assume that returns will follow a set distribution, and large "outliers" have a much stronger impact on the Omega ratio that many other return statistics. These characteristics are what make it an ideal starting point for our Edge measure. The measure was designed to identify a manager who has relatively few very large wins and more frequent very small losses

We categorized the testing universe into peer groups based on investment style, market cap focus and region. The total number of products used in the test segregated by categorization is shown below in Table 1

Table 1

Developed International		Products	US Large Cap			Products
	Core	419			Core	335
	Value	163			Value	338
	Growth	181			Growth	265
			US Small Cap			
International Small Cap		106			Core	183
					Value	167
Emerging Markets					Growth	224
	Core	147				
	Value	49				
	Growth	57				

Testing Procedure

Using the universes above, we performed both an In Sample and Out of Sample Optimization to test the efficacy of the methodology. The tests were performed as follows:

In Sample Optimization (*Time Horizon*-9/30/2001 - 12/31/2009)

Objective- to identify a group of managers with that maximizes the forward-looking Manager Skill relative to their peer average using independent variables as screening criteria.

Independent Variables- We looked at all quarter end observations for Total Consistency, Total Edge, Stock Consistency, Stock Edge, Style Edge, and Style Consistency

Metrics-The optimization set a single fixed threshold for each variable. A product was identified as "Skilled" if the scores for ALL variables were above the thresholds set.

Out of Sample Testing (*Time Horizon*-3/31/2010 - 9/30/2015)

Method- We used the fixed thresholds from the optimization to test each product by quarter. We categorized manager products as "Skilled" if they exceed the threshold for all variables. We also examined whether the results were directionally in line with the in sample results and tested them for statistical significance where applicable.

The results are shown on the next page.

Summary of Results

| | **In Sample Optimization** | | | | | | **Out of Sample Test** | | | | |
	% Identified as Skilled	% Increase in likelihood Top 2 Quintiles	Excess Alpha of Skilled vs Peer Avg	T-Stat	Signif		% Identified as Skilled	% Increase in likelihood Top 2 Quintiles	Excess Alpha of Skilled vs Peer Avg	T-Stat	Signif
Developed International											
Core	10%	39%	0.15%	0.51	30%		44%	19%	0.65%	2.79	0%
Value	11%	16%	0.75%	4.90	0%		59%	51%	1.41%	11.61	0%
Growth	5%	13%	0.78%	3.95	0%		43%	18%	0.68%	2.72	0%
Emerging Markets											
Core	6%	68%	1.62%	4.73	0%		51%	57%	0.68%	2.81	0%
Value	12%	129%	0.84%	4.81	0%		39%	40%	0.07%	0.33	37%
Growth	5%	68%	1.81%	5.26	0%		59%	-2%	0.69%	2.32	1%
International Small Cap	6%	25%	0.65%	4.86	0%		39%	-7%	0.17%	-0.21	58%
US Large Cap											
Core	6%	31%	0.54%	2.84	0%		54%	25%	1.01%	9.56	0%
Value	6%	14%	0.21%	0.90	18%		50%	4%	0.02%	0.33	37%
Growth	5%	22%	1.04%	3.05	0%		39%	-6%	0.04%	0.17	43%
US Small Cap											
Core	36%	11%	0.35%	2.47	1%		45%	-2%	-0.06%	-0.08	53%
Value	5%	10%	0.57%	0.52	30%		49%	1%	0.09%	0.75	23%
Growth	5%	9%	0.14%	0.89	19%		44%	-1%	0.05%	0.52	30%

Increase in probability of finishing in top 2 quintiles
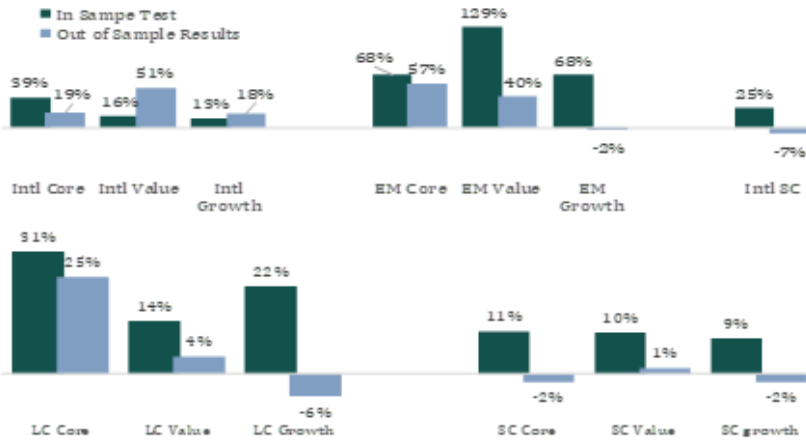
Forward 36m excess return vs Peer group average.
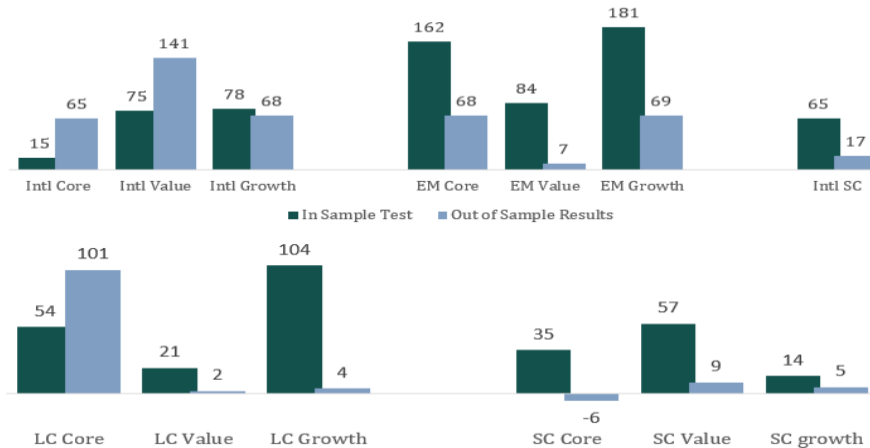
Conclusions and Next Steps

It is our belief that if there is true signal in a metric, it should not take a complex model to illustrate the value. Therefore, this test was designed to be simple and straight forward. The analysis had one primary goal. To show that using the portfolio replication and skill metrics developed by FIS and Aapryl, one could identify products that would outperform their style, out of sample. The results clearly showed this to be true. All peer groups but one (US Small Cap Core) increased Alpha or the likelihood of outperformance out of sample.

While our results are strong, we continue our research methods to improve our model and are currently developing and testing additional metrics and methodologies. Some of the improvements which we are working on include the following:

- Opportunity Score (a proprietary measure expanding on the concept of Active Share)
- Incorporating expenses, liquidity, and turnover
- Machine Learning to identify predictors of negative skill and refine positive skill predictors.

As we gain confidence in these enhancements and incorporate them into Aapryl, we will provide updated efficacy tests so that you can share our confidence.