



Live from Newport Beach. It's Smart Beta!

August 24, 2017

by Feifei Li, John West
of Research Affiliates

Key Points

- The outperformance observed before a typical smart beta index is launched virtually disappears once it's live, yet most investors are making decisions on backtest results.
- Two traits common to backtests—overfitting (or data-snooping bias) and ignoring transaction costs—bias investors' live return expectations higher than may be realistic.
- By expecting lower performance than backtest results show, questioning how those results were achieved, and selecting a strategy built on sound economic theory, smart beta investors can frame more realistic performance expectations.

Our headquarters in Newport Beach is only 50 miles from the Hollywood studios, although the drive can take up to two hours in rush-hour traffic. But far more than traffic separates the studios' world from ours. Film directors and actors are allowed multiple "takes" so each scene we see on the big screen is perfect. Occasionally, however, some productions are taped live, and the unplanned cannot be edited out. Who can forget the 2017 Oscars when *La La Land* was mistakenly announced as the winner of Best Picture?

Like the producers of the Oscars and other live shows, such as *Saturday Night Live*, investors don't have the luxury of re-takes—investing committed capital is "live." Portfolio results can and will go wildly off script, but there are no do-overs. With smart beta, investors often make decisions using simulations, or backtests. Backtests, like big movie productions, can be subject to editing.

Our research shows that approximately two-thirds of smart-beta-index track records are backtests and that most live track records extend no longer than a decade, which implies that much of the investment outcomes reported by smart beta providers are from simulations. In addition, much of the live history that exists is developed without having substantial assets invested in the strategy. Our collective 22 years on the front lines of smart beta research and investor engagement has shown us that investors nearly always base their decisions on these backward-looking, frictionless results.¹ We have no problem with that, if it's done with eyes wide open. Let's look at ways investors can set better expectations to maximize the benefits of these strategies.

Backtest vs. Live

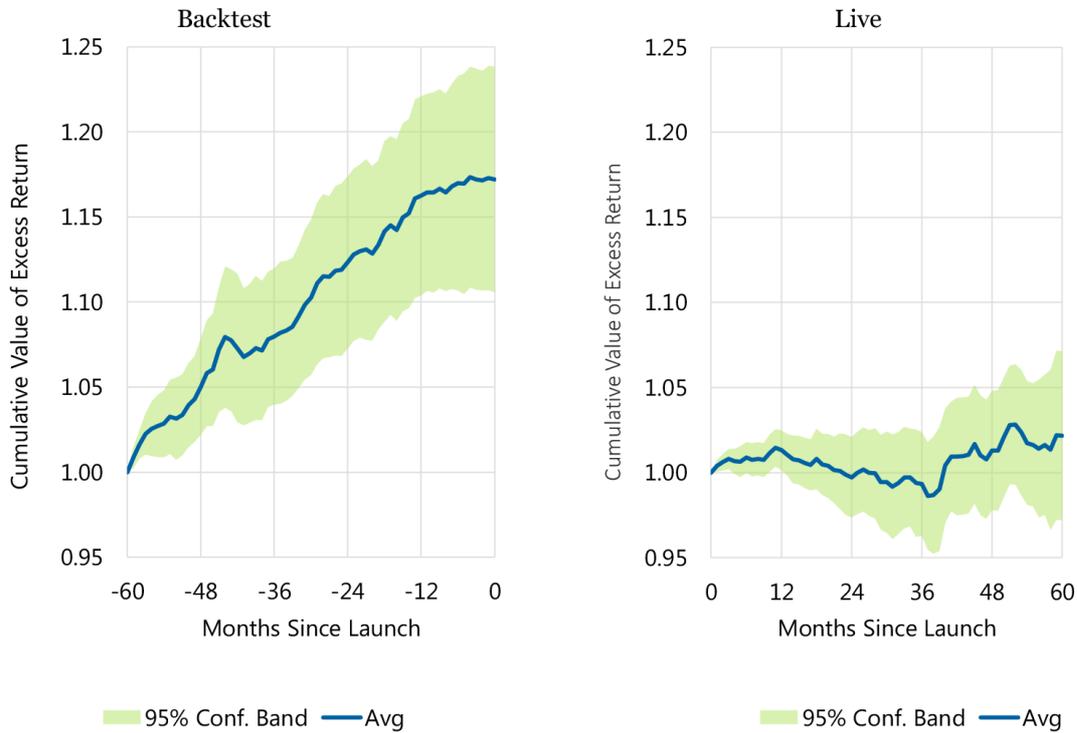
A backtest, a frequently used tool to frame forward-looking return expectations, is conveniently easy to calculate² and can be extremely helpful in proving a solid economic intuition with data—and, of course, with 100% hindsight. A backtest can also be useful in gaining a better understanding of the risks associated with an investment strategy—when the strategy is likely to underperform, and why.

Heavy reliance on backtest results can, however, be a harmful activity if investors are not fully aware of the limitations related to the simulated results. We examine the performance of 125 US equity smart beta indices on which exchanged traded funds (ETFs), characterized as strategic beta by Morningstar, are based. We exclude sector indices; indices for which we are not able to obtain the launch date; and indices with less than one year of backtest or live return data. If two or more ETFs track the same index, we include that index only once. The average live history of our universe of smart beta indices is 7 years, and the average total available history is about 21 years.

We find that prior to launch the indices tend to have superior performance relative to a market-capitalization-weighted benchmark, with outperformance peaking about six months ahead of the launch date. The outperformance seems to be extremely strong over the three-year period ahead of the launch. After the indices officially launch, however, their performance relative to the S&P 500 Index appears to hover around the base line, exhibiting virtually none of the outperformance demonstrated before they were live.

Live performance of smart beta indices exhibits virtually none of the outperformance shown in the backtest.

Cumulative Excess Return 60 Months Before and After Index Launch



Source: Research Affiliates, LLC, using data from Bloomberg. Note: Our sample consists of 125 US equity smart beta indices on which exchange-traded funds (ETFs), characterized as strategic beta by Morningstar, are based. We exclude sector indices; indices for which we are not able to obtain the launch date; and indices with less than one year of backtest or live return data. If two or more ETFs track the same index, we include that index only once. Excess returns are measured as of the launch date for the respective index.

Any use of the above content is subject to all important legal disclosures, disclaimers, and terms of use found at www.researchaffiliates.com, which are fully incorporated by reference as if set out herein at length.



In the backtest, the smart beta indices in our sample earned, on average, a 2.8% annualized excess return (t-stat = 8.72). The best-performing index was the S&P High Yield Dividend Aristocrats, which generated 14.5% outperformance above the return of the S&P 500 in the six-year backtest from January 2000 to November 2005. The average annualized live outperformance of our sample is 0.7% and 0.5% over a 5-year and a 10-year horizon, respectively; both outcomes are insignificantly different from zero, consistent with the data-snooping bias prevalent in backtests. Only 12 of 125 indices have significantly negative alpha in the backtest,³ whereas once live, the number almost triples.

Smart Beta Index Performance, Backtest and Live Results

	Annualized Excess Return					
	Backtest	Live				10 Year
		1 Year	3 Year	5 Year	5 Year	
Count	125	125	85	62	39	
Average	2.8%	1.4%	0.1%	0.7%	0.5%	
t-stat	8.72	2.46	0.27	1.53	1.38	
Minimum	-5.7%	-18.6%	-12.3%	-9.1%	-5.2%	
25th Percentile	0.7%	-1.8%	-1.2%	-0.5%	-0.9%	
Median	2.4%	1.2%	0.2%	0.8%	0.8%	
75th Percentile	5.0%	3.9%	1.6%	2.6%	1.8%	
Maximum	14.5%	29.4%	14.7%	9.0%	3.8%	

Source: Research Affiliates, LLC, using data from Bloomberg. Note: Our sample consists of 125 US equity smart beta indices on which exchanged traded funds (ETFs), characterized as strategic beta by Morningstar, are based. We exclude sector indices; indices for which we are not able to obtain the launch date; and indices with less than one year of backtest or live return data. If two or more ETFs track the same index, we include that index only once. Excess returns are measured as of the launch date for the respective index.

Any use of the above content is subject to all important legal disclosures, disclaimers, and terms of use found at www.researchaffiliates.com, which are fully incorporated by reference as if set out herein at length.



Five Best-Performing and Five Worst-Performing Smart Beta Indices Based on Backtest Results

Index	Backtest Start Date	Index Launch Date	Backtest	Annualized Excess Return			
				Live			
				1 Year	3 Year	5 Year	10 Year
Five Best-Performing							
S&P High Yield Dividend Aristocrats	Jan 2000	Nov 2005	14.5%	1.2%	4.1%	1.7%	0.4%
Dow Jones US Contrarian Opportunities	Jan 1992	Nov 2008	11.2%	29.4%	14.7%	9.0%	
WisdomTree US SmallCap Earnings	Mar 2002	Feb 2007	10.7%	-15.6%	2.0%	2.4%	1.0%
TWM Deep Value	Oct 2003	Aug 2014	10.4%	-9.0%			
GS Risk-Adjusted Return US Large Cap	Jan 2005	Dec 2011	9.4%	1.3%	-0.6%		
Five Worst-Performing							
Morningstar US Mid Growth	Jul 1997	Jul 2002	-5.1%	8.9%	6.9%	5.7%	3.2%
NASDAQ AlphaDEX(R) Mega Cap	Apr 2011	Jan 2016	-5.2%	-2.5%			
NASDAQ AlphaDEX(R) Small Cap Value	Apr 2011	Jan 2016	-5.5%	13.8%			
FTSE High Dividend Yield	Jan 2004	May 2009	-5.6%	-0.9%	1.5%	0.6%	
NASDAQ AlphaDEX(R) Mid Cap Value	Apr 2011	Jan 2016	-5.7%	12.9%			

Source: Research Affiliates, LLC, using data from Bloomberg. Note: Our sample consists of 125 US equity smart beta indices on which exchanged traded funds (ETFs), characterized as strategic beta by Morningstar, are based. We exclude sector indices; indices for which we are not able to obtain the launch date; and indices with less than one year of backtest or live return data. If two or more ETFs track the same index, we include that index only once. Excess returns are measured as of the launch date for the respective index.

Any use of the above content is subject to all important legal disclosures, disclaimers, and terms of use found at www.researchaffiliates.com, which are fully incorporated by reference as if set out herein at length.

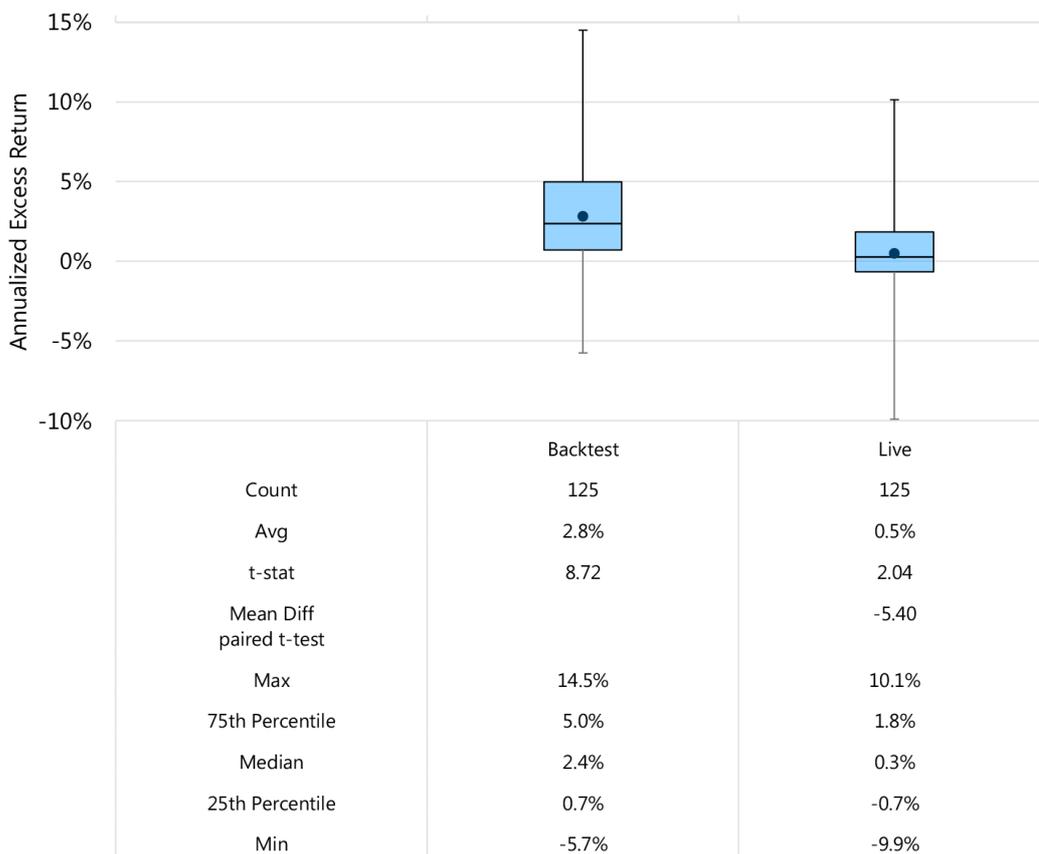


We perform a paired t-test and find that the average outperformance over the cap-weighted benchmark is significantly higher in the in-sample backtest compared to the out-of-sample live record. The strong alpha that dominates the backtest results does not survive over the indices' live histories. The performance pattern we observe around the launch date is very similar to the pattern observed by Brightman, Li, and Liu (2015), who find that the performance of the index underlying an

ETF is higher before the launch of the ETF than after. They conclude ETF providers appear to trend-chase index performance in creating their ETF products. We can say the same thing about index providers, who appear to trend-chase backtest results.

The average outperformance over the S&P 500 Index of the 125 smart beta indices we study is significantly higher in the in-sample backtest than the out-of-sample live record.

Distribution of Excess Returns, Backtest vs. Live⁴



Source: Research Affiliates, LLC, using data from Bloomberg. Note: Our sample consists of 125 US equity smart beta indices on which exchange-traded funds (ETFs), characterized as strategic beta by Morningstar, are based. We exclude sector indices; indices for which we are not able to obtain the launch date; and indices with less than one year of backtest or live return data. If two or more ETFs track the same index, we include that index only once. Excess returns are measured as of the launch date for the respective index.

Any use of the above content is subject to all important legal disclosures, disclaimers, and terms of use found at www.researchaffiliates.com, which are fully incorporated by reference as if set out herein at length.



The big gap between simulated and live performance can be largely explained by two common forces dominant in backtests—overfitting (or data-snooping bias) and ignoring transaction costs—both of which effectively bias investors’ return expectations higher than may be realistic.

Data-snooping risk. We all know there are no ugly backtests! More precisely, ugly-looking backtest results are rarely published in journals or client-facing materials. In the academic world, publication bias is well recognized, meaning that statistically significant results are three times more likely to be published than insignificant ones.

In our industry, quantitative managers are data mining every day in an attempt to identify signals that can accurately forecast a stock’s future return, and thus help improve a strategy’s performance. Smart beta strategies—model-driven strategies that involve the systematic selecting, weighting, and rebalancing of portfolio holdings based on factors or characteristics—are not exempt from this common practice. Importantly, this process should have proper guard rails to control data-snooping risk.

Even though a rich academic literature points out this problem and offers various solutions to mitigate it (McLean and

Pontiff [2016], Novy-Marx [2016], and Harvey and Liu [2017]), little has changed in practice. Investment managers still share their beautiful backtest results with investors, making few adjustments to the standard statistics. After all, who wants to make their results look worse?

We suggest a straightforward way for investors to establish realistic future return expectations. Backtests should be based on economically sound ideas that address the underlying relationship between signals and future performance. In analyzing a strategy, investors should consider who is on the other side of the trade, and why they would willingly choose to forgo the excess return the strategy is claiming to capture. Once the theory behind the excess return is established, the portfolio construction rules can be evaluated to assess their ability to best capture that excess return, after costs.

Take, for example, the Research Affiliates Fundamental Index™ (RAFI™), which is based on Research Affiliates' central investment belief of long-horizon mean reversion. We believe investors have a bias toward owning more of what has very good 3- to 5-year returns and an aversion to owning securities that have fared poorly. In keeping with this theory, a disciplined rebalancing strategy will sell recent winners and buy recent losers to produce an excess return. Those on the other side of these trades will be doing the opposite, taking the more inherently comfortable path of favoring recent winners and shunning recent losers.

The Fundamental Index uses accounting metrics to provide a stable anchor for contra-trading. When the market overestimates the future prospects of a stock and thus prices it too high, the fundamental-based weighting methodology helps investors pull back their investment in the stock. When the price mean reverts to the level justified by its discounted future cash flows, RAFI delivers alpha over a market-capitalization-weighted index by avoiding an overallocation to the stock, which would otherwise arise from price inefficiency.

The choice of the accounting metrics used in the weighting methodology does not really matter. The goal of the metrics is simply to capture the economic footprint of a company, independent of market perception, as a means of offering large capacity to investors by directing greater allocations to companies with higher liquidity. When a backtest deviates from solid economic intuition and theoretical support, the data-mining exercise loses a lot of credibility, and the results are useless, at best.

We strongly advocate for simplicity in smart beta methodologies to address data-snooping risk. The higher degrees of freedom in data mining, which are associated with a more complex methodology, give users more “knobs” to turn, potentially leading to stronger upward biases in in-sample outcomes. For example, an optimization-based approach, by its own definition, leads to the best in-sample return, volatility, or other targeted portfolio characteristic. While optimization and other complex methods of portfolio construction are very useful in obtaining certain objectives, adopting them simply due to their attractive in-sample performance is a dangerous practice.

Transaction costs. The other important factor that can explain disappointing live performance is implementation cost. Implementation costs are contributing to an ever larger portion of the gap between the expected performance of a smart beta index and its live record as the total amount of assets managed by these strategies rapidly grows.

The costs associated with executing a strategy are both explicit and implicit. The explicit costs, such as brokerage commissions and settlement/clearing charges, are directly observable, and explain a significant part of performance slippage, or the amount a fund's return underperforms the index it is tracking. The implicit costs, referred to as market impact costs, are the changes in a stock's price around index rebalancing dates, especially when the strategy's assets under management are large; that is, the prices of stocks being purchased are temporarily inflated, and those being sold are temporarily depressed. As prices revert in the days following the rebalancing, the strategy loses money. This outcome is not easily observable in smart beta strategies because the impact is embedded in the return of the underlying index, whose value is calculated on the basis of closing prices.

For strategy implementers, whose primary goal is reducing tracking error, a rational response is to lump all trading around the market close so the portfolio can perfectly track the index. These clustered trades also happen to be the most costly because they are reducing, within a very short time span, already-limited liquidity. Chow et al. (2017), after studying various portfolio characteristics related to implementation, recommend spreading trades over several days around the rebalancing, if possible.

Another way to lower market impact costs is to avoid smart beta strategies that invest in stocks with low liquidity. Screening out micro-cap and thinly traded companies' stocks is an important step in ensuring a strategy is “tradable,” even before considering the market impact of trades. Alphas produced “on paper” cannot successfully be reproduced when, for example, \$10 million in buying power is attempting to take advantage of a mispricing opportunity in a stock of a company with a total market capitalization of \$5 million. When they conduct simulations, thoughtful researchers will consider the trading volume of a stock, as well as set up proper constraints on the trades required by the strategy.

Investors who allocate to strategies, such as high dividend-growth, that typically require holding a relatively illiquid subset of the universe (Chow et al.) can apply a haircut to backtest results when setting their forward-looking return expectations. Illiquid stocks do offer more mispricing, and thus profit, opportunities, on average, because the price discovery process for these stocks is generally slower. Being attentive to the potential that the paper alphas of these strategies will likely be lower when they are live can shield investors from unpleasant surprises.

Strategies with high turnover rates, or when turnover occurs only for a few stocks rather than across the entire portfolio, also tend to experience high implementation costs. If this product feature is necessary to deliver the outcome investors seek, and no product design changes can address it effectively, sophisticated implementers can use algorithms to tactically take advantage of available liquidity. A momentum strategy falls into this category. For this reason, incorporating momentum in a passive smart-beta index strategy is very challenging.

As the popularity of smart beta strategies grows, the dollar volume of trades in the underlying securities—all competing for liquidity on rebalancing dates—likewise grows. This leads to higher market impact costs. Whereas the explicit costs of trading are decreasing over time as technology improves, we expect the implicit market impact costs associated with trading to increase. To help smart beta investors assess the market impact costs related to different strategies, we offer cost estimates based on Aked and Moroz (2015) on the Smart Beta Interactive tool on our website.

Conclusion

Saturday Night Live is the longest-running live television show in the United States. Viewers who tune in on Saturday nights know it's live and it won't be perfectly scripted. Likewise, investors who choose smart beta shouldn't expect the perfect alpha production promised by a simulated backtest. After all, backtests don't produce a single dollar, euro, or pound of investor benefit.

To improve the chance that the live results of smart beta strategies will produce the benefits investors expect, we suggest investors do three things:

Expect lower returns than the backtest produced. Backtest results can be an overly optimistic estimate of investors' experience going forward because of data-snooping risk and the omission of transaction costs.

Dig deeper. In order to achieve the superior investment outcomes promised by smart beta strategies, investors need to make decisions cautiously and request asset managers provide out-of-sample test results as well as return estimates that incorporate implementation costs.

Use theory. Most importantly, we recommend that investors select strategies built on strong underlying economic theory and that have a simple, transparent, and intuitive methodology.

Endnotes

1. In many cases, the investor doesn't have a choice.
2. With the proper machinery in place, building a not-super-sophisticated backtest from scratch that uses all listed US stocks over the past 50 years takes a well-trained researcher only a couple of days. Switching one backtest setup to another by altering certain parameters takes literally seconds.
3. We speculate these indices may be launched without backtests, and the history prior to launch is backfilled later so that negative value add is observed. We also speculate some of the negatively performing index launches are part of a more comprehensive family; if the suite covers a variety of investing styles over shorter time periods, some will naturally show underperformance.
4. The box plot shows the distribution of the annualized excess returns of the 125 indices using their entire available backtest return histories, which varies from 1 year to more than 30 years. Similarly, the time span of live returns varies from 1 year to more than 10 years.

References

Aked, Michael, and Max Moroz. 2015. "The Market Impact of Passive Trading." *Journal of Trading*, vol. 10, no. 3 (Summer):5–12.

Brightman, Chris, Feifei Li, and Xi Liu. 2015. "Chasing Performance with ETFs." *Research Affiliates* (November).

Chow, Tzee, Feifei Li, Alex Pickard, and Yadwinder Garg. 2017. "Cost and Capacity: Comparing Smart Beta Strategies." *Research Affiliates* (July).

Harvey, Campbell R. and Liu, Yan. 2017. "Lucky Factors." Available at SSRN, <https://ssrn.com/abstract=2528780>.

McLean, R. David, and Jeffrey Pontiff. 2016. "Does Academic Research Destroy Stock Return Predictability?" *Journal of Finance*, vol. 71, no. 1 (February):5–32.

Novy-Marx, Robert, 2016. "Testing Strategies Based on Multiple Signals" Simon Graduate School of Business, University of Rochester; available at <http://nm.simon.rochester.edu/research/MSES.pdf>.

© Research Affiliates